

6

Sampling Human Experience in Naturalistic Settings

Tamlin Conner and Eliza Bliss-Moreau

The topic of subjective experience—what people subjectively think, feel, and perceive—has interested philosophers, scientists, and laypeople alike. No matter what our role, as scientist or friend, we gain access to the subjective experience of others through the questions we ask and the answers people give. In the social sciences, this interest translates into the regular use of questionnaires, wherein people typically describe how they feel or behave “in general” (e.g., “How happy are you generally?”) or how they felt or behaved in the past (e.g., “How happy were you over *this past month*?”). While these global and retrospective assessments serve a purpose, they fail to adequately capture the rich contextualized representations of what people actually feel and think in daily life. Such limitations, coupled with advances in technology, have led to the emergence of a newer class of procedures, collectively called *experience sampling methods*. These methods allow people to report on their subjective experiences *in situ*, as they go about their daily lives.

Sampling Human Experience

Experience sampling methods refer to a collection of procedures that share three qualities—they allow people to report on experiences *in natural settings*,

in real time (not later from memory), and *on repeated time occasions*. For example, in a typical experience sampling method (ESM) study, people are asked questions about their experiences (e.g., “How happy do you feel right now?”) as they go about their daily activities. This process can be accomplished in several ways: (a) through the use of a computerized device (e.g., personal data assistant, called a “PDA”) that audibly signals participants throughout the day to answer questions displayed on the device, (b) through the use of paper booklets that participants complete at scheduled times or at random times (if pagers or programmable watches are used as signaling devices), or (c) through other emergent means such as logging onto websites or phoning in reports.

Brief History of Experience Sampling Methods

Naturalistic sampling methods have been around since the early years of social science (for a more detailed history, see Wheeler & Reis, 1991). Among the early pioneers of this methodology were sociologists Sorokin and Berger (1939), who gathered detailed accounts of how men and women spent their time in daily life, and psychologists Wessman and Ricks (1966), who investigated the links between daily mood fluctuations and depression in female college students. The 1970s and early 1980s saw the development of modern-day ESM, as paper-and-pencil diaries were combined with audible beepers to allow for the spontaneous sampling of subjective experience (Csikszentmihalyi, Larson, & Prescott, 1977; Hormuth, 1986; Hurlburt, 1979; Klinger, 1978; Larson & Csikszentmihalyi, 1978, 1983). With this technology, the term *experience sampling method* was coined (Larson & Csikszentmihalyi, 1983).

When first introduced, ESM referred to a particular technique involving the *random signaling* of participants during the day to answer questions about their experience (with a pager). Today, the term ESM is often used more broadly to refer to any procedure in which people report their experiences in naturalistic contexts over time, whether in response to a random signal, at fixed times during the day (e.g., at 10:00 a.m., 2:00 p.m., and 4:00 p.m.), or after specified events (e.g., a conflict, meal, or cigarette). Some people use the term ESM in the strict sense (to refer to random signaling), and others use it more in the general sense, as we do in this chapter. Note that ESM is also referred to as *diary methods* (Bolger, Davis, & Rafaeli, 2003), *event-sampling methods* (Reis & Gable, 2000), *thought sampling* (Hurlburt, 1997), and *ecological momentary assessment* (Stone & Shiffman, 1994). This last term, ecological momentary assessment, is used in the medical sciences to

refer to procedures that incorporate the ambulatory monitoring of physical states (e.g., blood pressure or heart rate) along with self-reports.

Uses of Experience Sampling Methods

ESM and related procedures have been used to investigate a broad range of phenomena across many domains of science. One of the earliest and most well developed applications of ESM has been for the study of emotional experience (e.g., Larson, Csikszentmihalyi, & Graef, 1980; Wessman & Ricks, 1966). By asking respondents to report on their emotions over an extended period of time, social and personality psychologists have gained considerable insight into the ebb and flow of daily emotional experiences. ESM research has revealed diurnal (Clark, Watson, & Leeka, 1989) and weekly patterns in mood (Larsen & Kasimatis, 1990; Stone, Neale, & Shiffman, 1993), as well as individual differences in mood variability (e.g., Eid & Diener, 1999; Fleeson, 2001; Penner, Shiffman, Paty, & Fritzsche, 1994) and in the entrainment of mood to normative cycles (Larsen & Kasimatis, 1990). These differences, in turn, correlate with important factors such as psychological vulnerability (i.e., greater mood variability is associated with higher neuroticism, Eid & Diener, 1999) and personality traits (i.e., extraverts are less entrained to weekly mood cycles; Larsen & Kasimatis, 1990). ESM research has also revealed that individuals differ in the degree of overlap between emotional states normally considered to be distinct (Feldman, 1995). Some people appear to distinguish between feelings of anxiety and feelings of depression in their daily lives, whereas others do not make such distinctions. This individual difference has implications for the comorbidity of anxiety and depression.

Indeed, ESM research has played an important role in clinical research. Researchers have used ESM to study how people cope with life's daily stressors (e.g., Stone, Neale, & Shiffman, 1993), the beneficial effects of being mindful (Brown & Ryan, 2003), and the characteristics of being in "flow" (Csikszentmihalyi & Csikszentmihalyi, 1988). ESM research has revealed the experiences of individuals suffering from depression (Swendsen, 1998; Wang et al., 2004) and other mental illnesses, including psychosis (Myin-Germeys, Delespaul, & van Os, 2003). For example, ESM research has shown that individuals vulnerable to psychosis are more emotionally reactive in daily life compared to nonvulnerable controls (Myin-Germeys, van Os, Schwartz, Stone, & Delespaul, 2001). Remarkably, ESM research has been used to study people diagnosed with schizophrenia (de Vries, 1992), showing how they feel better in smaller social groups and worse when alone (de Vries & Delespaul, 1989)—findings with clear, practical implications.

In health-related research, studies using ESM have given practitioners a unique window into the correlates and predictors of illness and wellness (for a review, see Stone, Shiffman, & de Vries, 1999). ESM research has been used to study the experiences of asthma sufferers (e.g., Smyth, Soefer, Hurewitz, Kliment, & Stone, 1999), individuals with chronic arthritic pain (e.g., Stone, Broderick, Porter, & Kaell, 1997), smokers who are trying to quit (e.g. Shiffman, Engberg, Paty, & Perz, 1997), and middle-income Americans and their drinking behaviors (Mohr et al., 2001), to name a select few. Recent technological advances in ambulatory physiological monitoring have also made it possible to track physiological activity along with subjective experience. Researchers have measured heart rate and blood pressure as people go about their daily activities (Goldstein, Jamner, & Shapiro, 1992; Jamner, Shapiro, Goldstein, & Hug, 1991; Kamarck, Peterman, & Raynor, 1998; Kamarck et al., 1998). They have also measured stress hormones via salivary cortisol to show that cortisol levels are higher for individuals who perceive and cope with stress less effectively (van Eck, Berkhof, Nicolson, & Sulon, 1996; see also Ockenfels, Porter, Smyth, & Kirschbaum, 1995).

ESM research is also used in sociology and business-related fields. Continuing the tradition of Sorokin and Berger, sociologists have employed ESM and related methodologies for investigating how time is budgeted (for an ESM-sociology discussion, see Juster, Hiromi, & Stafford, 2003). ESM research is also gaining ground in organizational research (Beal & Weiss, 2003), where it is used to track experiential factors that contribute to job satisfaction (Weiss, Nicholas, & Daus, 1999) and economic decision-making practices (Seo, 2004).

Rationale for Using Experience Sampling Methods

There are several reasons why ESM is emerging as a vital method in the social sciences.

First, ESM research enables measurement of subjective experiences in *naturalistic settings*, providing a window into people's experiences as they go about their daily lives. This setting contrasts markedly with many standard ways of assessing subjective experience, wherein people are asked to recall their experiences from memory or to fill out questionnaires in the lab, sometimes under experimental conditions designed to approximate real life. Consider the value of naturalistic reporting contexts for relationships research. Using ESM, researchers have gained considerable insight into how couples seek out and respond to each other emotionally. Researchers have asked couples to report on how they feel and what they do in response to conflict over a period of time. Analyses have revealed the systematic factors that promote

feelings of intimacy between partners (Laurenceau, Barrett, & Pietromonaco, 1998) and the processes by which partners affect each other emotionally under times of stress (Thompson & Bolger, 1999; for other examples see Gable, Reis, & Downey, 2003; and Larson & Almeida, 1999).

Second, ESM research captures experiences as they occur in real time, not recalled later from memory or reported in global terms. In the past, it was thought that real-time, retrospective, and global self-reports should all correspond—that a person who is happy in her daily life will also accurately recall these happy feelings and report being a happy person “in general.” But we now know that these self-reports reflect different types of knowledge and will not necessarily correspond (Robinson & Clore, 2002). Real-time reports capture immediate subjective states that fluctuate in response to changing events and conditions, and constitute a form of *episodic knowledge*. In contrast, global reports capture enduring beliefs about experiences and constitute a form of *semantic knowledge*. Retrospective reports fall somewhere in between, reflecting a blend of what actually happened with what we believe probably happened. In the past, the mistake has been to use global and retrospective reports as proxies for episodic experience.

In fact, there are several interesting types of memory biases that strongly call into question the use of retrospective self-reports for obtaining accurate reports of experiences as they originally occurred. First, we humans are not terribly good at accurately summarizing our past experiences over time. If we are asked to report how much pain we experienced *over the course of a week*, for example, our summaries tend to be overinfluenced by the most intense relevant experiences (“the peak”) and to a lesser extent by what happened most recently (“the end”), whereas underinfluenced by the duration of the experience (“duration neglect”; Fredrickson & Kahneman, 1993). Not only are we poor summarizers, but also we tend to fill in our memory gaps with our theories about ourselves. For example, memory for emotional experiences is often biased by people’s beliefs about their own emotionality (Barrett, 1997; Larsen, 1992), which can be influenced by gender stereotypes (Barrett, Robin, Pietromonaco, & Eyssell, 1998). As a result, when men and women recall how they felt over the past few weeks, women typically recall having more intense moods than do men, in line with gender stereotypes that women are the more “emotional sex.” Yet, remarkably, when men and women report their moods in the moment with ESM, they tend to show similarly intense moods (Barrett et al., 1998; for a review, see Robinson & Clore, 2002). Patterns like these also emerge in the retrospective reports of individuals from different cultures. For example, the cultural stereotype of the “happy American” has been shown to positively bias American college students’ retrospections of well-being compared to students from Asian

countries (Oishi, 2002). Again, few differences were found in actual well-being reported using ESM.

Last, ESM research yields an incredibly rich data set, giving researchers considerable *flexibility in data analysis*. Multiple observations are generated per person, which allow for intraindividual, or *idiographic*, analyses (i.e., the analysis of behavioral patterns for each person in a study). For example, with ESM, it is possible to examine how each person changes over time or how he or she reacts to changing conditions and contexts. Such analyses can reveal the rich contextual nature of human experience. If individuals vary in these patterns, investigators can test whether other person-level factors (demographic characteristics, personality traits, or any other relevant variables, typically measured separately from ESM) might account for the heterogeneity. In this way, ESM data are ideally suited to multilevel modeling procedures, where observations are the lower-level data “nested” within individuals as the upper-level data (for descriptions, see Nezlek, 2001; Reis & Gable, 2000). Data can also be analyzed using sophisticated time-series procedures to reveal systematic temporal patterns (for a description, see West & Hepworth, 1991).

Disadvantages/Caveats in Using Experience Sampling Methods

Despite the many strengths of ESM, there are a number of disadvantages that should be highlighted. First, research using ESM is primarily *correlational in nature*. Experiences and situations are measured rather than manipulated, and so ESM can only provide insight into the co-relations between variables (e.g., feelings of greater control tend to correspond with feelings of well-being), not the causal relations between variables (e.g., feelings of control cause feelings of well-being). Furthermore, ESM research depends on the natural incidence of particular events or experiences. If researchers are studying the relation between conflict and coping, for example, then they will need to determine whether there will be enough situations of conflict that naturally occur in order to warrant sampling.

ESM studies are also extremely *resource intensive*. Implementing such a study has the potential of being financially costly, demanding of participants, and challenging for the research team. Furthermore, available resources will likely dictate the sampling platform (i.e., whether one uses computerized or paper-and-pencil instruments) as well as the sample size of the study (see the “Design and Implement ESM” section below).

Another issue to consider is *measurement reactivity*. ESM studies are unique in that they require people to actively attend to and verbalize their

experiences repeatedly over time (often as much as 10 times per day over several weeks). This raises the question about whether the repeated self-monitoring of experience influences the very experience being measured. In general, evidence for reactivity is weak and/or mixed, although, admittedly, understanding of reactivity is very limited at this point. For example, one study examined trends over time as participants reported on their alcohol consumption, smoking, and sexual activities, among other behaviors (Gillmore et al., 2001). Analysis of trends showed overall decreased reporting of drinking, smoking, and sex, suggesting that sampling may sometimes lead participants to begin limiting risky behavior (presumably because they become aware of how much they partake in risky behavior). These effects, however, are not consistent (at least for drinking; Hufford, Shields, Shiffman, Paty, & Balabanis, 2002; Litt, Cooney, & Morse, 1998) and it is unclear whether participants simply began to underreport their risky behaviors due to social desirability. Fewer studies have examined reactivity for more subjective experiences like emotion or pain, where we might expect that asking people to attend to and document these experiences might make them unnaturalistically aware. In all likelihood, reactivity depends on several factors, including how well an individual can verbalize the experience being probed. To the extent that one is used to attending and labeling particular experiences (e.g., one's emotional state), sampling may simply mimic normal conditions and should not be a problem.

Finally, it should be noted that ESM studies are *not necessarily the "gold standard"* of self-report measures. While it is often crucial to measure experiences in a situated fashion, these moment-by-moment reports of experience may not *always* be the best measurement choice—it depends on the research question. Experience sampling captures the representation of experience as it occurs, or close to its occurrence, within the context of a person's everyday life (episodic experience—i.e., “How happy are you *right now?*”). These procedures do not reveal how a person organizes and retains representations of experience once the events have passed—that knowledge, distortions and all, can be important, too. Recent research suggests that people make important decisions about their future based on how they remember their experiences, not necessarily what “actually” happened in the moment. For example, memories of pain experienced during a medical procedure, more so than actual pain felt during the procedure, have been shown to predict important decisions about whether to undergo follow-up colonoscopies (Redelmeier, Katz, & Kahneman, 2003), and retrospective reports of enjoyment during a vacation, more so than people's actual reported enjoyment during the vacation, have been shown to predict whether or not people would seek to go on a similar trip in the future (Wirtz, Kruger, Scollon, & Diener, 2003). These

same effects might hold for other types of investigations and so it is important to consider what types of representations of experience are important (momentary versus retrospective versus global).

How to Design and Implement ESM

This section presents guidelines for designing and implementing an ESM study. We also encourage readers to look at other useful primers (e.g., Bolger, Davis, & Rafaeli, 2003; Conner, Barrett, Bliss-Moreau, Lebo, & Kaschub, 2003; Csikszentmihalyi, Hektner, & Schmidt, in press; Reis & Gable, 2000; Stone, Shiffman, & de Vries, 1999; Wheeler & Reis, 1991). In general, the first step is to assess one's resources—how much money and human power can be dedicated to the project. We break these resources into three types: platform implementation, subject remuneration, and the strength of the research team.

Assess Resources

Resources for Platform Implementation. ESM studies can be implemented in two main ways (computerized vs. paper and pencil), which differ in cost. Computerized sampling is the most financially costly, but it can provide the greatest flexibility in experimental design as well as greater assurance that participants comply with the sampling procedure. Computerized sampling methods use palmtop computers, personal data assistants (PDAs), or mobile phones (Collins, Kashdan, & Gollnisch, 2003) outfitted with specialized software that enables participants to report on behaviors and experiences throughout the day (for a discussion of computerized ESM, see Barrett & Barrett, 2001; Conner et al., 2003; Shiffman, 2000). Participants carry around these devices and then report on their experiences when cued by the device or when they initiate the report themselves. The basic costs associated with computerized experience sampling are the price of the computerized units (roughly \$100 for a decent PDA \times 50 = \$5,000), batteries and protective cases (\$500–\$1,000), as well as a potential cost for sampling software (free to very costly; see the “Software and Equipment” section below).

For all the costs associated with computerized ESM, there are considerable benefits. Many types of sampling procedures are “time-based” (Bolger, Davis, & Rafaeli, 2003) and rest on the assumption that respondents will complete their reports at certain times or immediately in response to a random audible signal. Computerized methods control these timing elements to ensure that respondents complete their reports at the appropriate times and not later

from memory. Computerized procedures also allow for flexibility in how the questions are presented. Questions may be presented sequentially or randomly to minimize response sets and order effects. These procedures also reduce human error associated with data management because data can be transferred directly from the PDA to a master computer without being entered by hand. And, finally, computerized procedures also provide the ability to record additional information, like latencies to respond to each question, which are not attainable with paper-and-pencil reports. In research, it can be useful to know how quickly a person was able to answer a question about his or her subjective experience.

Paper-and-pencil measures are a cost-effective alternative when resources are tight or when computerized procedures are ill suited for the study population (i.e., the elderly, subjects in higher crime areas where a PDA would be a safety risk). Questionnaire sheets, booklets, or any type of rating form (e.g., the Rochester Interaction Record; Nezelek, Wheeler, & Reis, 1983) can be designed and given to participants to fill out during the day (for additional guidelines on using paper-and-pencil measures, see Bolger, Davis, & Rafaeli, 2003; and also Green, Rafaeli, Bolger, Shrout, & Reis, in press). These paper-and-pencil measures are often used in combination with electronic pagers (e.g., Czikszenmihalyi & Larson, 1987) or programmable watches (e.g., de Vries, Dijkman-Caes, & Delespaul, 1990), which act as random signaling devices. The advantages of paper-and-pencil methods include reduced cost and lower overhead in terms of equipment. Paper-and-pencil methods also allow for open-ended responses, which makes them good for qualitative analyses. The disadvantages include the inability to randomize item content and greater risk of "non-compliant" responding (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2002). Noncompliance can occur if people forget to fill out their reports at designated times; if they fill them out at the wrong times; or worse, if they complete multiple reports later from memory. Fortunately, compliance can be improved dramatically by establishing good working relationships with participants (see Green et al., in press). So, although rapport with participants is crucial for all ESM studies, it is especially critical for studies using a paper-and-pencil platform.

Resources for Participant Remuneration. Another consideration is the availability of resources to remunerate participants for their time. Compared with other empirical methods, ESM studies are relatively taxing to participants because they require a significant time and energy commitment. Adequate remuneration entices respondents to sign up for a study and deters attrition. Depending on the population, a complex remuneration structure comprising multiple incentives can work well, starting with participant payment. At the

time of publication, the average rate of pay for college students was approximately \$20 a week. To maintain participants' motivation throughout the duration of long studies, researchers also provide smaller remunerations on a weekly basis (e.g., movie passes) and hold drawings for smaller prizes (e.g., \$25 gift certificates to restaurants, university sweatshirts, tickets to events). At the end of the study, researchers also sometimes hold drawings for a "grand prize" (e.g., for a PDA). Participants earn weekly remunerations and raffle tickets by coming to their regularly scheduled meeting with their research assistant. If the study is being run in a psychology department with a research requirement, research credits can also be used. Another form of incentive is to emphasize to participants their contributions to science. This helps participants to understand (validly, we believe) how important their participation is and that they are performing a much-valued social function.

Resources of the Research Team. A strong and potentially large research team is particularly important for successful ESM studies. Research assistants work closely with participants over the course of the study and strive to develop a relationship of mutual understanding. Research assistants must undergo extensive training (particularly in the case of computerized ESM) so that they are able to independently troubleshoot problems when problems arise. Research assistants should also pilot the procedure (e.g., carry around the device for a week) so they can understand what it is like to be a participant. Also, because ESM research generates large volumes of data, it is important that research assistants be well trained and comfortable with managing such data sets.

Determine Sampling Protocol and Parameters

Another important design consideration is the type of *sampling protocol* and the *length of the sampling period*. There are three types of sampling protocols (see Reis & Gable, 2000, for an in-depth discussion of these protocols). The first type of sampling protocol is signal-contingent sampling, in which participants report on their experiences at random times following a signal. Signal-contingent protocols typically ask participants to answer questions about their experience at that particular moment (i.e., How are you feeling *right now?*) and so this protocol is well suited to investigating target behaviors that are ongoing and likely to be occurring at the time of a given signal (e.g., mood). Signal-contingent protocols typically use between 4 and 12 signals each day, and the signals are usually delivered "randomly within equal intervals." This means that for a study sampling eight times a day between the hours of 10:00 a.m. and 10:00 p.m., the first signal would come randomly

between 10:00 and 11:30 a.m., the second signal would come randomly between 11:30 a.m. and 1:00 p.m., and so on, because there are eight 1½ hour intervals between 10:00 a.m. and 10:00 p.m. This ensures that the unpredicted signals are distributed throughout the day. The unpredictable nature of the signals combined with the immediacy of the reporting is advantageous when studying experiences that are highly susceptible to memory bias or possibly affected by mental preparation; however, because of their randomness and frequency, signal-contingent procedures are the most burdensome for participants.

The second type of sampling protocol is interval-contingent sampling, in which participants report on their experiences at fixed times (i.e., specific hours during the day; morning, afternoon and evening; or daily at night). Interval-contingent protocols generate strong time-series data because time points are fixed for all participants, so comparisons of responses can be made within and between individuals across specific times. Interval-contingent sampling is also great for measuring experiences that are less susceptible to memory bias (i.e., concrete, countable events). Participants can be asked to report on experiences or behaviors that occurred since the last time they made their report (e.g., "How many cigarettes have you smoked today?"). Interval-contingent protocols are the least burdensome to participants because reports are made at standardized times and so participants can configure their schedules around these reports. This regularity can be a drawback for the measurement of certain experiences (e.g., stress) because it allows participants to mentally prepare to report on their experiences (e.g., calming themselves down prior to making a report).

The third type of sampling protocol is event-contingent sampling, in which participants report on their experiences in response to specific events defined by the researcher (e.g., following an interpersonal conflict, a meal, etc.). Such protocols are well suited for assessing experiences and behaviors surrounding specific rare events, such as unprotected sex, which may not be occurring if one is sampled randomly or at set times. While event-contingent protocols are typically less demanding of participants than signal-contingent protocols, they require that participants be able to identify the contingent event, which can sometimes require training. Also note that compliance is an issue because it is impossible to ensure that participants respond immediately following the target event. Even computerized procedures cannot ensure compliance with event-contingent sampling.

The other major design considerations are the length of the sampling period and the number of sampling moments per day. These factors vary considerably within published studies. Although normative estimates are hard to come by, the average length of an ESM study (using a signal-contingent

procedure) is between 1 and 2 weeks, averaging 8 to 12 signals per day, yielding 56 to 168 observations (see Reis & Gable, 2000). There are several things to consider in making these decisions, the first being the naturalistic incidence of target events and states. For interval- and signal-contingent procedures, observations should be frequent enough during each day to capture important fluctuations in experience but not so frequent that they inconvenience participants without any incremental gain (Reis & Gable, 2000). Second, researchers should consider the total number of observations needed for a stable estimate of the phenomena. A multilevel power analysis should be run (Snijders & Bosker, 1999) to determine how many sampling moments are required for a stable estimate, assuming that the goal of the experiment is to model both within and between subject variance components. And, for event-contingent procedures, the sampling period should be long enough to accommodate the targeted numbers of observations per person. A fourth consideration is the burden to participants. Delespaul (1992) advises against sampling more than six times per day over longer sampling periods (i.e., 3 weeks plus) unless the reports are especially short (i.e., 2 minutes or less) and additional incentives are provided. Barrett (2004) has had success sampling 10 times daily over 4- and 6-week intervals by keeping the question time short (i.e., 2 minutes) and by adding weekly incentives (i.e., movie tickets, weekly lotteries, etc.). The last consideration is the anticipated response rate (the average number of signals/moments to which participants respond). Response rates tend to be highest (95% and above) for interval-contingent procedures using paper-and-pencil instruments that are completed either once or twice daily. On the other hand, response rates tend to be the lowest for signal-contingent procedures using computerized devices that signal multiple times per day (e.g., 50–70%; responding to 5–7 of 10 signals per day on average). If participants respond to only 50–75% of the trials, lengthening the sampling period will allow one to reach the target number of observations.

Choose Software and Equipment

If you decide to run a computerized ESM study, the next step is to choose the software and purchase equipment. A more detailed description of software and equipment considerations can be found in Conner et al. (2003). The choice of software depends on the requirements of your study, and the software will dictate the equipment that is needed. Researchers should conduct a Web search of sampling software, because new software is continually being developed and existing software changes rapidly. For basic studies involving the presentation of a set number of items in a signal-, interval-, or

event-contingent protocol, researchers can use a prepackaged, user-configurable program such as ESP (the Experience Sampling Program, Barrett & Barrett, 2001). ESP is a free, open-source software program that researchers can download from the Web, configure with little or no programming knowledge, and install onto either palmtop computers (running the Windows CE operating system) or PDAs (running the Palm operating system). The software comes with a step-by-step manual and is easy to use. For a complete listing of features, see the online manual (<http://www.experience-sampling.org>). Other existing freeware options include iESP (see <http://seattleweb.intel-research.net/projects/esm/iESP.html>), and the Purdue Momentary Assessment Tool (PMAT, <http://www.mfri.purdue.edu/pages/PMAT/>). If advanced features are needed, there are several other options. Researchers can use commercial software, hire an independent programmer, or employ a specialized consulting company that will design and implement specialized software. Consulting services are expensive, however, as many are geared toward supporting larger, government-funded studies.

After selecting the software, equipment comes next. In our experience, there are several important factors in deciding which portable sampling devices to purchase. First, the devices must be compatible with the software. Typically, this requires making sure the devices run on the same operating system for which the software was designed (e.g., Palm versus Windows). A second factor is the cost and quality of the devices. With each year, lower-cost PDAs are becoming increasingly available, but devices do vary in quality. It is important to research the different models to determine which brand and model best balances cost and quality. Also consider the size of the screen, the brightness and resolution, and the sound of the audible signal. Year-long warranties and excellent customer service help, too.

A Few Additional Tips

Prior to running any computerized ESM study, we strongly advise purchasing a single "test PDA" before purchasing the entire fleet. Install the software, configure it according to desired parameters, and see how it works. If all is well, then purchase the rest of your fleet. To cut down on costs, also consider purchasing a smaller fleet of devices and then running participants in waves. For example, a fleet of 40 would allow you to collect data for 120 participants in three waves of 40. Another useful tip is to initially purchase about 20% more units up front. This will save costs in the long run because after one or two studies, units will no longer be under warranty and they will start to break down. Researchers can replenish their fleet with units known to be compatible with the current software.

Once the equipment is purchased and the software installed and configured on the fleet, it is crucial to pilot test a sample of the devices. Our research assistants typically pilot a PDA for at least a week to provide feedback on their experiences. Pilot testers can tell you whether items need clarification, whether the audible tone was loud enough, and if they needed more time to answer the initial prompt. For example, in one study, we extended the time participants had to respond to the initial prompt to 2½ minutes based on their feedback. Pilot testing also helps estimate response rates. If motivated research assistants only respond to 60% of the prompts, then consider a louder tone or extending the sampling period by a few days to yield enough observations per person.

Implement the Study

ESM studies present several challenges in terms of implementation. These include recruiting participants, maintaining their motivation, ensuring compliance with the study protocol, and maintaining the integrity of the equipment.

Recruitment. ESM research necessitates high levels of commitment on the part of the participants, so oftentimes extra steps must be taken to recruit interested participants. With college populations, researchers often use banners and flyers posted in high-visibility areas (e.g., a central walking path on campus, dorms) that try to pique students' interest with an interesting question. For example, Reis and Gable (2000) report recruiting with questions like, "How many hours a day do people spend socializing?" (p. 207). We also find it effective to include the amount of money we offer as remuneration. For example, "Need money? Learn more about yourself, and earn \$50 in the process." Ads can also be placed for nonstudent participants in local newspapers.

Maintaining Motivation. Not only must participants be willing to start a study, but also they must be committed to completing it. ESM studies can be fairly time intensive, and so maintaining participant interest and motivation is crucially important. Again, one of the most important factors in maintaining motivation is a positive working relationship between the researcher and the participant (Green et al., in press). Participants will adhere to the policies and practices of the study, or disregard them entirely, largely in response to a research assistant's attitude. Participants should feel important, that researchers are responsive to their needs (via attentive contact and positive reinforcement), and that they are doing something meaningful with their time. Participants who feel positive about the study and who feel a greater

sense of responsibility to their research assistant are less likely to cancel appointments and breach study protocols. Also, as discussed earlier, a complex remuneration system also helps, especially with multiple incentives (e.g., prizes, drawings, etc).

Increasing Compliance. The success of an ESM study also depends on the conscientiousness, or compliance, of participants—that they respond when they are asked to respond, as honestly as they can, and not in a random fashion. Again, a positive relationship between the researcher and the participant is extremely important. Also important is for participants to be very clear about the study procedures. Participants should be walked through the ESM procedure and given the opportunity to ask questions and get comfortable with the procedure. Another way to boost performance is to provide clear, immediate feedback to participants regarding their response rates. Feedback about response rates can dramatically increase the amount of usable data by making participants aware if they need to be more watchful in making their reports.

Maintaining Equipment. Damage and wear of equipment is an inevitable part of running a computerized ESM study, but some steps can be taken to minimize such damage. PDAs can be carried around in protective cases and their screens covered either with Teflon-based car wax or a protective plastic sheet to prevent scratching. And, using top-of-the-line batteries will extend the lifetime of the units because they are less likely to corrode. Top-of-the-line batteries will also prevent fewer data disasters (e.g., when batteries die and data stored is permanently erased).

Prepare, Clean, and Analyze Data

ESM research yields volumes of data that must be entered, organized, and readied prior to any statistical analysis. Consider an ESM study with 100 participants. If each participant answers 20 items at 100 observation points, this study could potentially yield 200,000 data points. With paper-and-pencil sampling procedures, all data must be entered manually by hand—a lengthy and error-prone process, but certainly achievable. Computerized ESM bypasses this step because data are retrieved directly from the portable devices through a simple “hotsync” procedure to a desktop computer. This process eliminates manual data entry; however, careful steps must be taken not to inadvertently override or erase files. Once the data are entered, they are typically compiled and cleaned prior to analysis. Data are compiled by moving the data from individual files into one large file, arranged in the way best suited to statistical analysis. We typically use a univariate set-up so that each

row contains the answers for one sampling moment for each person (suitable for the Hierarchical Linear Modeling program; Raudenbush, Bryk, & Congdon, 2005). Each participant has the same number of rows (some of which contain missing data), and participants' data are stacked on top of each other, starting with Participant 1. Once the data are compiled, they are then cleaned. Data derived from computerized ESM are checked for duplicate or missing records and the data entered from paper-and-pencil sampling are checked for data entry errors. Also, with computerized data, trials with extremely fast reaction times typically indicate participant error (i.e., inadvertently tapping the screen twice for the previous item) and should be excluded from analysis and documented in the write-up.

After being compiled and cleaned, the data are ready to be analyzed. There are innumerable ways to analyze such a rich data set. We recommend several excellent texts, each of which describes statistical procedures in detail, including multilevel modeling procedures (Bolger, Davis, & Rafaeli, 2003; Kenny, Bolger, & Kashy, 2002; Kenny, Kashy, & Bolger, 1998; Nezlek, 2001; Reis & Gable, 2000).

Conclusion

Emergent methods have the power to stimulate theory and research in the social sciences. Once introduced and refined, these methods reveal original directions in research, allowing investigators to ask and answer questions otherwise unconsidered. In the spirit of this idea, we have reviewed the theory and practice of experience sampling methods as useful tools in the social sciences. Although experience sampling should not be undertaken lightly without consideration of rationale and resources, these methods are becoming increasingly feasible with gains in technology and a growing body of methodological guidelines. We hope this chapter serves as a helpful resource.

References

- Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin*, 23(10), 1100–1110.
- Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of emotional experience. *Journal of Personality and Social Psychology*, 87, 266–281.
- Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, 19(2), 175–185.

- Barrett, L. F., Robin, L., Pietromonaco, P. R., & Eysell, K. M. (1998). Are women the "more emotional" sex? Evidence from emotional experiences in social context. *Cognition and Emotion*, 12(4), 555–578.
- Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organizational Research Methods*, 6(4), 440–464.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84(4), 822–848.
- Clark, L. A., Watson, D., & Leeka, J. (1989). Diurnal variation in the positive affects. *Motivation and Emotion*, 13(3), 205–234.
- Collins, R. L., Kashdan, T. B., & Gollnisch, G. (2003). The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption. *Experimental and Clinical Psychopharmacology*, 11(1), 73–78.
- Conner, T., Barrett, L. F., Bliss-Moreau, E., Lebo, K., & Kaschub, C. (2003). A practical guide to experience-sampling procedures. *Journal of Happiness Studies*, 4(1), 53–78.
- Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (1988). *Optimal experience: Psychological studies of flow in consciousness*. New York: Cambridge University Press.
- Csikszentmihalyi, M., Hektner, J., & Schmidt, J. (in press). *Measuring the quality of everyday life: The ESM handbook*. Thousand Oaks, CA: Sage.
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease* [Special issue: Mental disorders in their natural settings: The application of time allocation and experience-sampling techniques in psychiatry], 175(9), 526–536.
- Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent experience. *Journal of Youth and Adolescence*, 6, 281–294.
- de Vries, M.W. (Ed.). (1992). *The experience of psychopathology: Investigating mental disorders in their natural settings*. Cambridge, UK: Cambridge University Press.
- de Vries, M. W., & Delespaul, P. A. (1989). Time, context, and subjective experiences in schizophrenia. *Schizophrenia Bulletin*, 15(2), 233–244.
- de Vries, M., Dijkman-Caes, C., & Delespaul, P. (1990). The sampling of experience: A method of measuring the co-occurrence of anxiety and depression in daily life. In J. D. Maser & C. R. Cloninger (Eds.), *Comorbidity of mood and anxiety disorders* (pp. 707–726). Washington, DC: American Psychiatric Press.
- Delespaul, P. A. E. G. (1992). Technical note: Devices and time-sampling procedures. In M.W. de Vries (Ed.), *The experience of psychopathology: Investigating mental disorders in their natural settings* (pp. xvii, 363–373). New York: Cambridge University Press.
- Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, 76(4), 662–676.

126 Emergent Methods in Social Research

- Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of Personality and Social Psychology*, 69, 153–166.
- 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(6), 1011–1027.
- Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, 65(1), 45–55.
- Gable, S. L., Reis, H. T., & Downey, G. (2003). He said, she said: A quasi-signal detection analysis of daily interactions between close relationship partners. *Psychological Science*, 14(2), 100–105.
- Gillmore, M. R., Gaylord, J., Hartway, J., Hoppe, M. J., Morrison, D. M., & Leigh, B. C. (2001). Daily data collection of sexual and other health-related behaviors. *Journal of Sex Research*, 38(1), 35–42.
- Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (in press). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*.
- Goldstein, I. B., Jamner, L. D., & Shapiro, D. (1992). Ambulatory blood pressure and heart rate in healthy male paramedics during a workday and a nonworkday. *Health Psychology*, 11(1), 48–54.
- Hormuth, S. E. (1986). The sampling of experiences in situ. *Journal of Personality*, 54, 262–293.
- Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors*, 16(3), 205–211.
- Hurlburt, R. T. (1979). Random sampling of cognitions and behavior. *Journal of Research in Personality*, 13(1), 103–111.
- Hurlburt, R. T. (1997). Randomly sampling thinking in the natural environment. *Journal of Consulting and Clinical Psychology*, 65(6), 941–949.
- Jamner, L. D., Shapiro, D., Goldstein, I. B., & Hug, R. (1991). Ambulatory blood pressure and heart rate in paramedics: Effects of cynical hostility and defensiveness. *Psychosomatic Medicine*, 53(4), 393–406.
- Juster, F. T., Hiroimi, O., & Stafford, F. P. (2003). An assessment of alternative measures of time use. *Sociological Methodology*, 33(1), 19–54.
- Kamarck, T. W., Peterman, A. H., & Raynor, D. A. (1998). The effects of the social environment on stress-related cardiovascular activation: Current findings, prospects, and implications. *Annals of Behavioral Medicine* [Special issue: Social psychophysiology of cardiovascular response: Implications for the study of reactivity, risk, and disease], 20(4), 247–256.
- Kamarck, T. W., Shiffman, S. M., Smithline, L., Goodie, J. L., Paty, J. A., & Gnys, M. (1998). Effects of task strain, social conflict, and emotional activation on ambulatory cardiovascular activity: Daily life consequences of recurring stress in a multiethnic adult sample. *Health Psychology*, 17(1), 17–29.
- Kenny, D. A., Bolger, N., & Kashy, D. A. (2002). Traditional methods for estimating multilevel models. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling*

- intraindividual variability with repeated measures data: Methods and applications* (Multivariate applications book series, pp. 1–24). Mahwah, NJ: Lawrence Erlbaum.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, et al. (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 233–265). New York: McGraw-Hill.
- Klinger, E. (1978). Dimensions of thought and imagery in normal waking states. *Journal of Altered States of Consciousness*, 4(2), 97–113.
- Larsen, R. J. (1992). Neuroticism and selective encoding and recall of symptoms: Evidence from a combined concurrent and retrospective study. *Journal of Personality and Social Psychology*, 62(3), 480–488.
- Larsen, R. J., & Kasimatis, M. (1990). Individual differences in entrainment of mood to the weekly calendar. *Journal of Personality and Social Psychology*, 58(1), 164–171.
- Larson, R. W., & Almeida, D. M. (1999). Emotional transmission in the daily lives of families: A new paradigm for studying family process. *Journal of Marriage and the Family*, 61(1), 5–20.
- Larson, R. W., & Csikszentmihalyi, M. (1978). Experiential correlates of time alone in adolescence. *Journal of Personality*, 46(4), 677–693.
- Larson, R. W., & Csikszentmihalyi, M. (1983). The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, 15, 41–56.
- Larson, R. W., Csikszentmihalyi, M., & Graef, R. (1980). Mood variability and the psychosocial adjustment of adolescents. *Journal of Youth and Adolescence*, 9(6), 469–490.
- Laurenceau, J. P., Barrett, L. F., & Pietromonaco, P. R. (1998). Intimacy as an interpersonal process: The importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of Personality and Social Psychology*, 74(5), 1238–1251.
- Litt, M. D., Cooney, N. L., & Morse, P. (1998). Ecological momentary assessment (EMA) with alcoholics: Methodological problems and potential solutions. *Health Psychology*, 17, 48–52.
- Mohr, C. D., Armeli, S., Tennen, H., Carney, M. A., Affleck, G., & Hromi, A. (2001). Daily interpersonal experiences, context, and alcohol consumption: Crying in your beer and toasting good times. *Journal of Personality and Social Psychology*, 80(3), 489–500.
- Myin-Germeys, I., Delespaul, P., & van Os, J. (2003). The experience sampling method in psychosis research. *Current Opinion in Psychiatry*, 16(12), 33–38.
- Myin-Germeys, I., van Os, J., Schwartz, J. E., Stone, A. A., & Delespaul, P. A. (2001). Emotional reactivity to daily life stress in psychosis. *Archives of General Psychiatry*, 58(12), 1137–1144.
- 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., Wheeler, L., & Reis, H. T. (1983). Studies of social participation. *New Directions for Methodology of Social and Behavioral Science*, 15, 57–73.

128 Emergent Methods in Social Research

- Ockenfels, M. C., Porter, L., Smyth, J., & Kirschbaum, C. (1995). Effect of chronic stress associated with unemployment on salivary cortisol: Overall cortisol levels, diurnal rhythm, and acute stress reactivity. *Psychosomatic Medicine*, 57(5), 460–467.
- Oishi, S. (2002). The experiencing and remembering of well-being: A cross-cultural analysis. *Personality and Social Psychology Bulletin*, 28(10), 1398–1406.
- Penner, L. A., Shiffman, S., Paty, J. A., & Fritzsche, B. A. (1994). Individual differences in intraperson variability in mood. *Journal of Personality and Social Psychology*, 66(4), 712–721.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2005). *HLM 6: Hierarchical linear and non-linear modeling*. Lincolnwood, IL: Scientific Software International.
- Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, 104(1), 187–194.
- Reis, H. T., & Gable, S. L. (2000). Event sampling and other methods for studying daily experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). New York: Cambridge University Press.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128(6), 934–960.
- Seo, M. G. (2004, August). *The effect of affective experience on the direction, intensity, and persistence of task behavior*. Paper presented at the meeting of the Academy of Management, New Orleans, LA.
- Shiffman, S. (2000). Real-time self-report of momentary states in the natural environment: Computerized ecological momentary assessment. In A. A. Stone, J. S. Turkkan, C. A. Cachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 277–296). Mahwah, NJ: Lawrence Erlbaum.
- Shiffman, S., Engberg, J. B., Paty, J. A., & Perz, W. G. (1997). A day at a time: Predicting smoking lapse from daily urge. *Journal of Abnormal Psychology*, 106(1), 104–116.
- Smyth, J. M., Soefer, M. H., Hurewitz, A., Kliment, A., & Stone, A. A. (1999). Daily psychosocial factors predict levels and diurnal cycles of asthma symptomatology and peak flow. *Journal of Behavioral Medicine*, 22(2), 179–193.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Sorokin, P. A., & Berger, C. F. (1939). *Time-budgets of human behavior*. Oxford, UK: Harvard University Press.
- Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, 16, 199–202.
- Stone, A. A., Broderick, J. E., Porter, L. S., & Kaell, A. T. (1997). The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis Care and Research*, 10(3), 185–193.
- Stone, A. A., Neale, J. M., & Shiffman, S. (1993). Daily assessments of stress and coping and their association with mood. *Annals of Behavioral Medicine*, 15(1), 8–16.

- Stone, A. A., Shiffman, S. S., & de Vries, M. W. (1999). Ecological momentary assessment. In D. Kahneman & E. Diener (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 26–39). New York: Russell Sage Foundation.
- Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *British Medical Journal*, *324*(7347), 1193–1194.
- Swendsen, J. D. (1998). The helplessness-hopelessness theory and daily mood experience: An idiographic and cross-situational perspective. *Journal of Personality and Social Psychology*, *74*(5), 1398–1408.
- Thompson, A., & Bolger, N. (1999). Emotional transmission in couples under stress. *Journal of Marriage and the Family*, *61*(1), 38–48.
- van Eck, M. M., Berkhof, H., Nicolson, N., & Sulon, J. (1996). The effects of perceived stress, traits, mood states and stressful daily events on salivary control. *Psychosomatic Medicine*, *58*(5), 447–458.
- Wang, P. S., Beck, A. L., Berglund, P., McKenas, D. K., Pronk, N. P., Simon, G. E., et al. (2004). Effects of major depression on moment-in-time work performance. *American Journal of Psychiatry*, *161*, 1885–1891.
- 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), 1–24.
- Wessman, A. E., & Ricks, D. F. (1966). *Mood and personality*. Oxford, UK: Holt, Rinehart, & Winston.
- West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, *59*, 609–662.
- Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality* [Special issue: Personality and daily experience], *59*(3), 339–354.
- Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, *14*(5), 520–524.