TAMLIN CONNER CHRISTENSEN, LISA FELDMAN BARRETT,
ELIZA BLISS-MOREAU, KIRSTEN LEBO and CYNTHIA KASCHUB

A PRACTICAL GUIDE TO EXPERIENCE-SAMPLING PROCEDURES

ABSTRACT. Experience-sampling is a powerful method for understanding a range of psychological phenomena as they occur in the daily lives of individuals. In this primer, we discuss the different techniques, equipment, and design options available to the experience-sampling researcher. We place special emphasis on computerized procedures and discuss the crucial social dynamic of the research team, which optimizes the success of experience-sampling procedures.

Launching an experience-sampling study provides a challenge to even the most seasoned researcher. The term "experience-sampling" refers to a set of empirical methods that are designed to allow respondents to document their thoughts, feelings, and actions outside the walls of a laboratory and within the context of everyday life. In addition to the standard array of issues that beset any empirical study, experience-sampling procedures bring with them unique design, implementation, and methodological issues. For example, experience-sampling designs vary in terms of when events are sampled (at fixed intervals, randomly throughout the day, or in response to some event; for reviews see Reis and Gable, 2000; Reis and Wheeler, 1991), how these designs are implemented (computerized vs. paper-and-pencil measures), and for how long they are implemented (from a few days to a month). In this article, we address these and other issues in an effort to provide a hands-on, step-by-step guide to designing and implementing an experience sampling study. We pay special attention to the use of electronic recording devices, like palmtop computers and personal data assistants (PDAs) (also see Feldman Barrett and Barrett, 2001; Shiffman, 2000). In this way, we hope to supplement the many detailed reviews of experience-sampling procedures that currently exist (see Bolger et al., 2003; Csikszentmihalyi and Larson, 1987; de Vries et al., 1990; Delespaul, 1992; Hormuth, 1986; Hurlburt, 1997; Reis and Gable, 2000; Shiffman, 2000; Stone et al., 1991; Stone and Shiffman, 1994; Wheeler and Reis, 1991).

We organize this article around seven steps integral for running a successful experience-sampling study (see Table I). Steps 1–3 discuss

Journal of Happiness Studies **4:** 53–78, 2003.

TABLE I
Checklist for designing and implementing an experience-sampling study

Step	✓
1 Decide whether you need experience-sampling	☐
2 Determine your resources	☐
Strong research team	☐
Remuneration	☐
Platform resources	☐
3 Set study parameters	☐
Type of protocol	☐
Sampling period	☐
4 Choose software and equipment	☐
Choose software	☐
Purchase equipment	☐
5 Implement security measures	☐
6 Implement the study	☐
Configure and pilot devices	☐
Create your documentation	☐
Anticipate participant issues	☐
Maintain your equipment	☐
7 Data Issues	☐
Prevent loss of data	☐
Data cleaning	☐

the preliminary stages of research design, beginning with whether experience-sampling is appropriate for your particular research question. Steps 4 and 5 focus on equipment and security, two issues unique to computerized experience-sampling. Steps 6 and 7 discuss various implementation and data issues. This step-by-step information represents condensed knowledge from our laboratory, interviews with expert colleagues, and published reports.

STEP 1: DECIDE WHETHER YOU NEED EXPERIENCE-SAMPLING

Experience-sampling is time- and resource-intense for researchers and participants. The first step in determining whether this method is appropriate for you is to consider whether your goal is to measure episodic or semantic representations of experience. Experience-sampling procedures capture the representation of experience as it occurs, or close to its occurrence, within the context of a person's everyday life (i.e., "How happy are you right now?"). As such, experience-sampling reports tend to measure representations that are *episodic* (or experiential) in nature (for a discussion of episodic memory, see Tulving, 1985). Episodic representations are different from those revealed through standard self-report measures, in which people report on their experiences "in general" (i.e., "Are you a happy person?") or after the fact, in retrospect (i.e., "Over the last month, how happy have you been?"). Such self-reports tend to measure representations that are *semantic* (or conceptual) in nature. They reflect people's generalized knowledge or theories about their experiences (in the case of global self-reports; e.g., Robinson and Clore, 2002) or they reflect representations of experiences that have been filtered and reconstructed through semantic knowledge structures (in the case of retrospective self-reports; e.g., Ross, 1989).

Episodic and semantic representations are probably related to one another, but in principle can be distinguished (Klein, 2001) and neither is more valid than the other. For example, moment-by-moment reports do not tell you whether a person organizes and retains representations of on-going knowledge once the events have passed. Rather, they describe the contents of representations made in situ.

In addition, it is important to realize that experience-sampling reports only yield information that a person is willing and able to represent in conscious awareness at the moment a report is made (Feldman Barrett and Barrett, 2001; Shiffman, 2000). Experience-sampling procedures are not a direct "pipeline" into consciousness; they cannot correct for or prevent the processes that transform conscious experience into information available for report (for a discussion of the distinction between conscious experience and self-reflective awareness, see Chalmers, 1996). Also, caution is in order when proposing that experience-sampling is a way to "get around" some kinds of motivated processing about the self. It is still not known to what extent momentary reports are affected by the same motivational processes that affect standard self-reports (i.e., socially desirable responding or psychological defense), although the reigning belief is that momentary reports are protected from these processes.

Next, consider how much control you need over the situation in which the reports occur. Experience-sampling procedures depend upon the natural incidence of particular events or experiences and do not permit controlled delivery of situational variables. If control or base-rates of target stimuli are a concern, then consider a laboratory procedure, or combining an experience-sampling protocol with a lab-based experiment, where, among other things, you can ask respondents about their experiences as they happen in response to a controlled stimulus. Of course, with this format, you give up another benefit of experience-sampling – a richly detailed profile of a person's phenomenological life, over time, in a naturalistic context.

STEP 2: DETERMINE YOUR RESOURCES

The feasibility of conducting an experience-sampling study depends on three main resources: (a) having a strong research team; (b) being able to remunerate participants; and (c) having financial support to implement your study on the desired experience-sampling platform (e.g., computerized vs. paper-and-pencil instruments).

A Strong Research Team

A strong research team is integral to the success of any study, but it is particularly important to one that includes experience-sampling procedures. In our experience, an optimal team consists of two team leaders (either a graduate assistant or a paid study coordinator) plus enough research assistants so that each is assigned between 5 and 10 participants. Participants are assigned to a specific research assistant for the entire study. This configuration gives participants and research assistants the opportunity to develop a relationship of mutual understanding, which in turn helps to maintain everyone's motivation. In our lab, each assistant undergoes extensive training in groups and individually by shadowing a more experienced peer. It also helps to have the same research assistants for the entire duration of the study. For example, if you are planning to run a study across two semesters, consider finding assistants who are willing to commit for that year. Maintaining continuity minimizes repeated training sessions and generally makes for a better study.

Remuneration

Remuneration is an important issue because experience-sampling studies are taxing to participants and it is important to be proactive in reducing attrition. We recommend a complex remuneration structure comprised of multiple incentives. First, participants are paid for their efforts. We typically pay our participants approximately $20 a week for their participation. Second, they receive smaller remunerations on a weekly basis (e.g., candy; movie passes). These weekly incentives help maintain participants' motivation throughout the duration of long studies. Third, we hold drawings each week for smaller prizes (e.g., $25 gift certificates to restaurants, university sweatshirts, tickets to university functions) and at the end of the study for a "grand prize" (for which participants are eligible to win a Personal Data Assistant). Participants "earn" weekly remunerations and raffle tickets each week by coming to a regularly scheduled lab meeting that lasts throughout the duration of the sampling period[1]. Participants may also be offered research credits (where such compensation is available).

Another form of incentive is to emphasize to participants that they are donating their time to science. This helps participants to feel (validly, we believe) that they are performing a much valued social function. For example, in our lab, we explain to participants that without their dedication and valuable contribution, psychology textbooks would be considerably shorter. We show them journal articles that have been published by our lab, using data collected from participants like themselves. We find it most effective to emphasize this incentive at the beginning of the study and to repeat it several times throughout the duration of the sampling period.

Platform Resources

Resources also affect which platform you use to implement an experience-sampling study. Experience-sampling studies can be implemented using resource-intense computerized methods, or using paper-and-pencil methods.

Computerized Methods

These include the use of palmtop computers or Personal Data Assistants "PDAs" (like PalmPilots) installed with specialized software that enables participants to report behaviors and experiences in response to an audible signal at various times during the day, or, self-initiated following a particular event. To date, computerized sampling has been used successfully to study a number of different phenomena, including coping (Schwartz et al., 1999), individual differences in the variability in mood (Penner et al., 1994), emotion-related process (Feldman Barrett, 2003), subjective well-being (Oishi, 2002), tobacco-related behavior (Shiffman et al., 1995) and cravings for alcohol (Litt et al., 1998).

Paper-and-Pencil Methods

These include the Rochester Interaction Record (Nezlek et al., 1983; Wheeler and Reis, 1991), or specialized rating forms (e.g., Feldman, 1995; Czikszentmihalyi and Larson, 1992), which can be combined with technology-augmented formats, in which participants complete their paper-and-pencil reports when signaled by devices such as electronic pagers (e.g., Czikszentmihalyi and Larson, 1987) or programmable watches (e.g., de Vries et al., 1990). More typically, however, respondents initiate reports at predetermined times (e.g., every evening) or under particular conditions (e.g., following social interactions).

How to Make the Choice

In our view, computerized methods are always preferable. Foremost is their ability to ensure that participants comply with interval and signal-contingent sampling procedures. Interval and signal-contingent procedures are "time-based" (Bolger et al., 2003) and rest on the assumption that respondents will complete their reports at fixed times (in the case of interval-contingent) or immediately in response to an audible signal (in the case of signal-contingent). Computerized methods control these timing elements to ensure that respondents completed their reports as instructed (note that computerized methods do not ensure compliance with event-contingent procedures, in which participants initiate their own reports in response to a particular event). Tabulating missed trials produces an objective index of compliance without having to rely on participants' retrospective accounts. Computerized procedures allow greater flexibility in the item presentation (e.g., items may be presented in fixed and/or random order minimizing the use of a response set), reduce human error associated with data management (e.g., devices transfer data directly to a master computer), and provide the ability to record ancillary information, like latencies to respond to each item.

Of course, computerized methods also have their downside. It can be expensive to purchase the units. Add to this the price of batteries, protective cases for the units, and the cost of insuring, repairing, or replacing the units, and costs can become prohibitive. There are also software considerations. Some software is available free of charge (e.g., The Experience Sampling Program, ESP; Barrett and Feldman Barrett, 2000); however, if you require special features not currently available in ESP, then you will need to hire a programmer or purchase a commercial software package, which may include consulting charges. Another disadvantage of computerized experience-sampling concerns the flexibility of item format. Current technology of these devices limits items to fixed formats (i.e., with set response options) and does not accommodate open-ended responses.

If you do not currently have the resources to purchase and maintain equipment, then you can use paper-and-pencil measures, but keep in mind their limitations. Foremost, you cannot ensure that participants will comply with time-based protocols. With paper-and-pencil instruments, participants can delay filling out their reports to more convenient times, thereby introducing sampling bias into the data (i.e., when the thoughts and feelings at moments of convenience do not represent the true distribution of thoughts and feelings present at other times), or, memory bias into the representations of experience. You can take steps to minimize breeches of protocol, however (e.g., collecting records on a daily basis; asking participants during the debriefing to estimate the number of records completed from memory and excluding participants with rates over some specified threshold). A second disadvantage of paper-and-pencil instruments is the inability to randomize item presentation, increasing the likelihood of response bias due to item context (e.g., Harrison and McLaughlin, 1993). A third disadvantage concerns data management. Entering data by hand is labor-intensive and subject to considerable human error. Some researchers have addressed this problem by converting their paper-and-pencil documents to scantron forms (e.g., JPL pers. comm.); however, this process can also be expensive and burdensome to participants who must use number two pencils and fill in bubbles completely – a task that can be difficult when completing multiple reports per day.

STEP 3: SET STUDY PARAMETERS

Next, you must decide on the type of sampling protocol you will use (whether signal, interval, or event-contingent) and the length of the sampling period.

Type of Protocol

Decisions regarding protocol type are generally based on three central criteria: (a) the prevalence of the target behavior or events that are being studied (i.e., their base-rate); (b) the susceptibility of a phenomenon to retrospective memory bias; and (c) the burden to participants (these criteria and others are discussed in greater detail by Reis and Gable, 2000).

Interval-contingent protocols involve reporting on experience at fixed times throughout the day (e.g., at morning, afternoon and evening intervals; or at night daily). Respondents may be asked to report on their experiences at that particular time (e.g., rating self-esteem at 10 am and 10 pm daily; Kernis et al., 1993) or to report on their experiences that occurred during the time frame since the previous report (the prior "interval") (e.g., describing the positive and negative events that occurred during the day; Nezlek and Gable, 2001). This latter format requires some retrieval or reconstruction over a period of time, and so should not be used for experiences that are susceptible to retrospective memory bias (e.g., emotions, subjective well-being, or any experiences that are quick to decay). In general, interval-contingent protocols are well suited for studying relatively frequent experiences, because the experiences will likely occur at the time of the report or during each interval. Interval-contingent protocols also tend to be the least burdensome to participants. Reports are made at predictable times so participants can configure their schedules around reports. Predictability can be a liability, however. Knowing when to make a report gives respondents adequate time to prepare themselves cognitively or emotionally. Because interval-contingent procedures standardize the time span between reports, they are well suited to time series analyses. As such, the time units should be meaningful and not simply arbitrary.

Signal-contingent protocols involve reporting on experience in response to a signal at various times throughout the day. Signal-contingent protocols are well suited for studying target behaviors that are on-going and therefore are likely to be occurring at a given signal. Signal-contingent protocols are also appropriate for studying (a) states that are susceptible to retrospective memory bias if they were reported on later from memory (signal-contingent reports usually ask about experiences happening *at that moment* and not over the last 4 h or the entire day, as with some interval-contingent protocols); and for (b) states that are susceptible to cognitive or emotional regulation (signal-contingent reports come at unpredictable times, which may limit processes that restore self-esteem, for example). The main disadvantage of signal-contingent reports is their burden to participants, who are interrupted by the signal and must stop what they are doing to document their experiences.

Event-contingent protocols involve reporting experience immediately or closely following a particular event of interest. This protocol is appropriate for studying behaviors or events that are less common or relatively rare in daily life, which may not emerge at a particular signal or within a given interval (e.g., incidences of lying; DePaulo et al., 1996). Event-contingent procedures also minimize the reliance on memory, but only to the extent that reports are made close in time to the event. Event-contingent procedures can be a challenge to participants especially if the events are frequent (i.e., every social interaction last 10 min or longer; Pietromonaco and Feldman Barrett, 1997) or are too broadly defined. So, it is important to set clear and appropriately inclusive criteria.

Sampling Period

Decisions regarding the sampling period (how many days participants will report on their experiences) and number of samples per day is based on four considerations: (a) the number of observations needed for a stable estimate of a particular phenomenon within a person; (b) the naturalistic incidence of target events and states; (c) the burden to participants; and (d) anticipated compliance. The number of observations varies considerably within published studies, from as few as seven per person (once daily for seven days; e.g., Conner Christensen et al., 2003; Gable et al. 2000) to as many as 270 per person (3 times daily for 90 days; e.g., Feldman Barrett, 1998). Although normative estimates are hard to come by, the average number of observations for signal-contingent procedures is estimated to be between 56 and 168 (for studies that run for 1–2 weeks, averaging 8–12 signals per day; Reis and Gable, 2000). The ideal number of observations depends, in part, on whether you plan to estimate both within and between subject variance components (i.e., using multi-level modeling procedures; Byrk and Raudenbush, 1992; Nezlek, 2001). You should consider running a multi-level power analysis to determine the number of observations needed per person to estimate a within-subjects effect. This procedure is beyond the scope of the paper, but readers are referred to Snijders and Bosker (1999) and Snijders et al. (1999).

When deciding how to distribute the observation points over the sampling period, consider both the base rate of target events and the burden to participants. For interval- and signal-contingent procedures, observations should be frequent enough during each day to capture important fluctuations in experience, but not so frequent so to inconvenience participants without any incremental gain (Reis and Gable, 2000). For example, Delespaul (1992) advises against sampling more than 6 times per day over longer sampling periods (i.e., 3 weeks +) unless the reports are especially short (i.e., 2 min or less) and additional incentives are provided. For event-contingent procedures, the sampling period should be long enough to accommodate the targeted *N* numbers of observations per person.

Finally, you should take into account participants' estimated response rates when setting the length of sampling period. For example, if participants respond to only 75% of the trials, then you may want to lengthen sampling period in order to reach your target number of observations. Generally, 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. Response rates tend to be lowest (e.g., 70%, on average) for signal-contingent procedures employing computerized devices that signal multiple times per day (>8 times). Studies vary greatly in these estimates so you should always pilot test for your particular protocol. For example, in our lab, average response rates have ranged from as low as 50% (responding to 5 out of 10 signals per day across 2 weeks using PDA devices) to 70% (responding to 7 out of 10 signals per day across 4 weeks using palmtop computers).

STEP 4: CHOOSE SOFTWARE AND EQUIPMENT

Choosing software and equipment always takes longer than expected. So, it is essential to start early.

Choose Software

The choice of software depends upon the requirements of your study. For basic studies involving the presentation of a set number of items in an interval, signal or event-contingent protocol, you can use a pre-packaged, user-configurable program such as ESP (Barrett and

Feldman Barrett, 2000). ESP is an open-source software program that you 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 personal data assistants (running the Palm Operating System). The software comes with a step-by-step manual and the only prerequisite for using the software is working understanding of the chosen device and its operating system (e.g., see Pogue, 1999).

In its current downloadable format, ESP has the following features. First, it can be used to run any of the three types of sampling protocols (interval-, signal- or event-contingent). Devices can be configured so that participants initiate their own reports (for event-contingent studies) or complete reports in response to audible signals presented at set times throughout the day (say, 10 am, 1 pm and 4 pm) or at variable times within a particular time range (e.g., 10 times between the hours of 9 am and 11 pm). (In the latter case, ESP can divide the time range into *N* equal intervals and present one beep randomly within each interval.) Units can also be programmed with different time frames (say, 9 am–11 pm for Participant A, but 10 am–12 pm for Participant B) if you are tailoring the protocol to each individual's wake–sleep cycle. Second, ESP allows items to be randomized, but with limited flexibility (in the current ESP version, fixed trials must come before the randomized trials due to software code). Third, ESP controls certain response-timing elements. For example, you can specify the amount of time a participant has to respond to the initial prompt (typically 90–150 s) and the amount of time a participant has to respond to each item (each item can have its own response window, if needed). For Windows CE platforms only, devices can be configured to signal again 5 min later if a respondent misses an initial prompt. This option is not currently available for PalmOS platforms, however. For a complete listing of features, see the on-line manual (*http://www2.bc.edu/~barretli/esp*).

ESP may not have all the features you need. For example, ESP does not currently allow either branching capabilities, in which items vary according to how a participant responds on the previous item (e.g., Schwartz et al., 1999) or complex randomization of items within more than one block. ESP also "takes over" the machine, disallowing participants to access other programs on the device, and, it does not permit participants to turn off or on their devices or to delay their responding to more convenient times, say if they are in class or church (e.g., Schwartz et al., 1999). Other programs have allowed participants to delay the response beep for as much as 20 min (Shiffman et al., 1995), which can help boost response rates.

If advanced features are needed, then you will want to investigate other software options. One option is to hire a programmer to modify the ESP source code. ESP is an open-source software package, freely distributable under the GNU Public License agreement, so you are free to change the code to suit your needs (C++ code is available through the ESP website). A second option is to use a commercial development tool like that from Pendragon Software (*http://www.pendragon-software.com*), which specializes in formatting pen-and-pencil questionnaires onto PDAs at a set cost per device. A third option is to hire an independent programmer or a specialized consulting company (e.g., InvivoData Corporation) that will design and implement specialized software for you. Consulting services are expensive, however, as many are geared toward supporting all stages of a research study (e.g., design; implementation, data preparation). But they do provide considerable customer support, unlike ESP, which is provided with no warranty and no support. That means that researchers are independently responsible for implementing the program.

Purchase Equipment

After figuring out which software to use, then you are ready to purchase your equipment. In our experience, there are four important factors in deciding which equipment to purchase: (1) compatibility between the device's operating system and your desired software; (2) cost of the devices; (3) human factors decisions, including the size of the screen and sound of the audible signal; and (3) warranties and customer service provided by the company.

Software Compatibility

Be sure to choose a device with an operating system that (a) is compatible with your intended software and (b) has enough memory capacity to run the software. The two major operating systems are Windows CE, which runs on most palmtop computers, and PalmOS, which runs on most PDAs. Operating systems also come in different versions so check that the version you purchase is compatible with your software. Most devices have enough memory to run experience-sampling software. For example, PalmPilots have a minimum of 2MB of RAM, which is sufficient to run the ESP program and to store data up to 100 days without filling up the memory (of course, you should upload data regularly and change batteries). If you are using tailored software, you may need to pay greater attention to memory requirements. It is strongly recommended that you purchase one device and pilot test it with your software prior to purchasing your entire fleet.

Cost

At the time this article was written, palmtop computers cost anywhere from $400 and $650, which is an average of three times more expensive than the PDAs, which can be purchased for anywhere between $80 and $400 each. Palmtop computers also require more battery usage, which increases their overall operating costs. For these reasons, many researchers are using the lower-cost PDAs. Whichever you choose, we recommend contacting companies directly to ask for bulk discounts or special educational offers.

Human Factors

The screen on your chosen device should be large enough to accommodate the longest item and its response selections. Screen size is less of a factor for palmtop computers, but PDAs do vary in their screen sizes, and unfortunately, it usually happens that the least expensive PDAs have very small screens. Also consider the brightness or contrast of the display. Most devices allow some adjustment of contrast, but those with lower baseline levels of contrast are much more difficult to read than those with high contrast. We refer the reader to Consumer Reports Magazine (*http://www.consumerreports.org*), which often has articles on current models with their features and prices.

Palmtop computers and PDAs have different responding options for you to consider. Palmtop computers have mini-keyboards, so that participants can respond by using the keyboard or by tapping response options on the screen (i.e., "soft buttons" with arbitrary labels). We have found that when using the keyboard, participants are more reliable in their responses than when they use soft buttons. The downside to keyboards is that the labels are fixed and participants tend to have to "hunt and peck" to locate the correct key. The PDA lacks a keyboard, so participants respond by tapping the soft buttons on the screen. With soft buttons, it is easier to find the desired response, but tapping the screen can sometimes be unreliable. For example, a single tap on the screen, if held too long, can be registered as two taps (as if responding to Items 1 and 2), which produces error. These kinds of errors can be identified and removed prior to analyses, however (see Step 7: *Data Issues* below).

Finally, researchers should consider the quality and volume of the device's audible tone. Many devices offer different types of tones (e.g., "alarm," "wave," etc.) at various intensities (soft, medium, loud). It has been our experience that the PDAs tend to have quieter audible tones even at their loudest intensities compared to some palmtop computers, but there is probably variability across brands. Participants in one of our studies using PDAs said they would have been able to answer more signals if the devices were louder.

Warranty and Customer Service

Remember to investigate warrantees, support, and return/repair policies. Also, be sure to ask how long a model will be in production and to what extent the company will continue to provide support once the model is taken out of production.[2]

STEP 5: SECURITY

Equipment is crucial for electronic experience sampling research, so it important to take steps to protect the devices from loss or damage. One of the first issues to consider is whether to insure the devices or not. Start by checking with your University or organization to determine the annual costs for policy coverage. In our experience, we have found these costs to be prohibitive and have opted to "self-insure," meaning, our lab (and, by extension, the University) absorbs the cost to replace any machines that are reported lost or stolen. Fortunately, we have had few losses. From a fleet of 50 palmtop computers that have been in constant use for 4 years, none have been lost or stolen to date, although 2 are damaged each semester of data collection, on average. From a fleet of 36 PDAs that were used constantly for one semester (running 93 participants), two PDAs were lost and never recovered.

The second issue to consider is whether to hold participants responsible for the loss or damage of equipment. We know of no lab that requires participants to pay for the machines if they damage or lose them, and we are not sure that it is ethical to do so. Instead, we inform participants in both oral and written form that we are issuing them University property and that by signing the informed consent they agree to return the device when the study is completed. Before issuing the device, we take a photocopy of their driver's license (stored separately from all study materials) and inform participants (both orally and on the consent form) that if they miss a scheduled lab session and are not in contact with their experimenter for a period greater than 48 h after that time of the missed session, we will report the palmtop computer or PDA missing to our University Police. In our four years of running computerized sampling studies, we have filed a University Police report twice. It also helps to keep a record of the serial numbers for each device (i.e., PDA #33 with serial number AAAED49012319) in case you need to file a report. Serial numbers allow you to identify a machine even if identifying stickers have been removed.

To facilitate the return of misplaced equipment, it goes without saying to put some form of identification on the device. We label our equipment with numbered metal ID tags (available at such companies as EMEDCO; *http://www.emedco.com*) that bear the University's name and laboratory phone number. Some universities may be willing to provide university equipment identification tags. Equipment can also be embossed or engraved with such identification information (JPL, pers. comm., 2002).

STEP 6: IMPLEMENT THE STUDY

At this point, you have set your study parameters, pilot tested your software on at least one device, purchased your fleet of devices, and determined your security measures. Now, you are ready to implement your study. This process consists of four parts: (1) configuration and piloting (a second time); (2) documentation; (3) participant issues; and (4) maintenance.

Configure and Pilot Devices

Configuring your portable devices involves (a) installing the software on each device in your fleet; and (b) setting the program parameters (e.g., timing, type of tone, items with response choices, etc.) on each unit. Software is typically installed through a "hotsync" operation, which transfers the software files from a main computer to the portable device. This process can take some time, as each device must be connected one at a time to an installation computer. Parameters are set in different ways depending upon the software you are using. For example, with ESP, you can specify some parameters prior to installing the software (e.g., items and response choices can be listed in a special file on the installation computer, which are then transferred to the device at the same time you install the software); other parameters are specified after installing the software (e.g., timing elements are set by opening the software program on the device). Each software package is different, so be sure to check your manual for details.

Once you have configured your devices, it is helpful to have research assistants carry the devices for a minimum of one 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 by 30 s (to $2\frac{1}{2}$ min) based on their feedback. Pilot testing is also helpful for estimating response rates. By extrapolating these rates across the intended sampling period, you can anticipate whether your intended sampling period (e.g., 2 weeks) is long enough to yield enough observations per person. If motivated research assistants only responded to 60% of the prompts, then you might need to extend the sampling period by a few days.

Create Documentation

All experimentation requires some documentation, but we have found it crucial for experience-sampling procedures. Foremost, research assistants should have a *manual* that provides step-by-step instructions for how to run a participant through all of the study procedures. Manuals are necessary to ensure that all participants receive the exact same instructions, as well as to ensure that all tasks are completed correctly and in full. This is surprisingly difficult in experience-sampling studies, which include many details that are easily forgotten if not written down. Research assistants should be explicitly familiar with what happens in each laboratory session, how to upload/hotsync data, where to save data, etc. Included in this manual should be a system for managing files that are uploaded from the devices. This system should ensure that research assistants save the uploaded files in the correct directory, in a way that does not overwrite pre-existing files, and that all data are backed up regularly (see Step 7: *Data Issues* below). We recommend using a *saved data log* in which research assistants document the date and time of each upload, record the name of the file, and indicate whether the file has been saved to the correct directory. Data logs help to maintain the integrity of the system and to troubleshoot any data file issues later on. We have also found that *incident reports* are an essential source of documentation. Incident reports help to keep track of unusual conditions that may affect the outcome of data analysis. They are sheets of paper kept in the participant's file in which research assistants document anything unusual that occurred during the session, ranging from minor behavior (e.g., the participant appeared tired or drifted off to sleep during one of the tasks) to major technical difficulties with computers (e.g., cracked screen resulting in the loss of data).

In addition to the usual requirements (e.g., consent forms), documentation for participants should include *information sheets* that they can consult if they have questions about their devices during experience-sampling. Participants should also have *contact sheets*, which include their research assistant's phone number and e-mail address as well as another emergency contact person in case they should have a problem with the device. We also provide *instructor/employer letters* for each participant to give to his or her professors and employers. The letter explains that the person is taking part in an experience-sampling study and that the device may beep in their class and/or place of business at some point. The letter is signed by the Principal Investigator and includes contact information should the professor or employer have any questions or concerns. Across eight continuous semesters of running, only two professors have contacted the PI indicating that they would rather not have the palmtop/PDA in their class.

Participant Issues

One of the reasons that experience sampling is such a challenging method is that it necessitates high levels of commitment on the part of the participants. Not only must participants be willing to start a study, but also they must be committed to complete it. As experience sampling researchers, we are thus faced with challenges to (a) recruit participants; (b) maintain their motivation; and (c) ensure that they comply with study protocol throughout the duration of the sampling period.

Recruitment

Recruiting is a major challenge. In our lab, we use a combination of banners and flyers posted in high visibility areas (e.g., a central walking path on campus; dorms). These signs may include a variety of information, but we, like other researchers, find it helpful to peak 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 dollars in the process."

Maintaining Motivation

Studies vary greatly in length of time required by participants, both in number of hours or days that the complete study lasts, and in the amount of time required to answer the questions asked of the participants. It follows that maintaining motivation in participants is of crucial importance and one of the largest tasks facing the experience sampling researcher and research team. In our experience, we have found that there are three main components important to maintaining participant motivation. The first component we discussed in previous sections – having a complex remuneration system, with incentives beyond basic cash or research credit. The other two components are positive attitudes on the part of the research team and establishing good working relationships between research assistants and participants.

Research assistants' attitudes are immensely important to a successful run. Both explicitly (i.e., in the form of direct communication with participants about the integrity of the study) and implicitly (i.e., through non-verbal communication and the general atmosphere between research assistants), research assistants set the tone for each study. Implicit attitudes have been shown to affect many aspects of social life, including interpersonal behavior and communication, affect and motivation (Bargh, 1994; Greenwald and Banaji, 1995) and we see these effects play out in the lab. In our lab, we have observed that participants will adhere to the policies and practices of the study, or disregard them, largely in response to a research assistant's attitude. As in other types of behavioral science research, a good experience-sampling research team is comprised of research assistants who are made aware of their contribution to the integrity of the study and the importance of their work.

Lastly, a good working relationship between participants and research assistant relationship serves to enhance the study in two ways. First, research assistants better understand what motivates their participants, thereby helping them to maintain focus throughout the study. Second, participants feel a greater sense of responsibility to the particular research assistant with whom they work. Participants who have a working relationship with their researcher are less likely to cancel appointments or show blatant disrespect for study protocols. A good working relationship can be developed by attentive contact (e.g., calling participants on the second day to "check in") and positive reinforcement (e.g., direct praise).

Increasing Compliance

The success of an experience-sampling study depends upon the conscientiousness (or "compliance") of participants to respond to as many trials as they can, as honestly as they can and not in a random fashion. Experience has taught us two ways to help boost participants' performance. The first is to ensure that participants are absolutely clear on the meaning of each item and/or events. It is crucial to walk each participant through the questions that will be asked at each experience sampling prompt. In our lab, research assistants initiate a trial and hand the device to the participant who goes through the questions at his or her own pace, giving each participant the opportunity to ask questions of the assistant. This practice trial (the first recorded in the data) can be subsequently deleted from the master data file. Other labs allow more time for practice, as much as the first two days (e.g., Shiffman et al., 1995).

The other way to boost performance is to provide clear, immediate feedback to participants regarding their response rates. In the case of electronic experience sampling, software is available to quickly calculate response rates from the previous sampling period so that research assistants can inform participants on their progress. For example, Barrett and Feldman Barrett (2000) have written a small program called ESPCount that extracts response rate information from data collected using the Experience Sampling Program. It has been our experience that feedback about response rates can dramatically increase the amount of usable data. If participants' response rates do not increase to the necessary level, extra days can be added to the participants' sampling time to help boost their overall response rates. We extend the sampling period only by a maximum of 25% (i.e., a maximum of 7 extra days for a 28-day period). Beyond that, the participants tend to differ too much in terms of their motivation and often cannot be treated as part of the same population.

This drive to maximize response rates must also be tempered by accuracy. In our studies, participants are told clearly both orally by the research assistant and in writing on the informed consent that they must comply with all study procedures in order to be remunerated in full. Participants are expected to answer the questions honestly, avoid random responding, not allow others to respond for them, and complete the study for the stated sampling period. Of course, participants are free not to answer specific questions as they see fit, but if they do respond to an item, it should be in good faith. And they are also free to withdraw from the study at any time (whereby they forfeit part of their remuneration but retain other parts).

Maintenance

Damage and wear of equipment is an inevitable part of running a computerized experience-sampling study. Damage includes cracked screens and hardware problems. Wear includes burned out screen bulbs, scratches on the display, broken/missing battery covers or cases, missing styluses. Fortunately, steps can be taken to minimize damage and wear. To protect the screens, you can ask participants to carry their devices in their protective cases (e.g., one researcher uses cases called the "Bumper" by Concept Kitchen to protect his PDAs.). Researchers have also experimented with a Teflon based car wax to prevent scratching or covered the screen with a protective plastic sheet to reduce damage from the stylus from tapping. Some labs refrain from using the styluses altogether and ask participants to tap the screen with their finger; however, finger-touch can be difficult for smaller screened devices and it also leads to a buildup of oil and dirt. If you do plan to use the styluses, we suggest requesting replacement styluses when ordering the devices as a precautionary measure. Companies usually provide extra styluses free with each device, but only when asked.

A second maintenance issue concerns power. Portable sampling devices run on batteries and do not come equipped with writeable a hard drive. If the device runs out of power before the data are retrieved (i.e., the back-up battery dies or batteries are removed from the device), then all of the data are lost. There are technologies like data chips that can prevent accidental erasing or destruction of data (e.g., Penner et al., 1994), however, these are not yet widely available. Palmtop computers have backup batteries in case the regular batteries run out, but PDAs do not have such backup. Thus, it is crucial to schedule regular lab visits to upload/hotsync data and change batteries.

Finally, researchers are strongly encouraged to buy top of the line batteries. One colleague used generic brand batteries that died before the end of a 10-day period (two recordings daily) and ended up corroding some of the PDA battery compartments. He switched to a top of the line battery and has not had any of those problems since. Of course, battery lifetime varies greatly on the devices depending on brand, model, and usage patterns. But good batteries tend to last 1 week with the palmtop computer (the back up battery lasts a month or more) and 2 weeks or more with PDAs.

STEP 7: DATA ISSUES

With computerized experience-sampling procedures, data are retrieved from the palmtop computers and the PDAs through a simple data "upload" (palmtop computers) or "hotsync" (PDAs) to a master computer. This process uses simple, out-of-the box software that comes with the devices, which eliminates error-prone manual data entry. Yet, computerized experience-sampling procedures do require some additional considerations regarding the data. These considerations include (a) preventing loss of data and (b) cleaning and checking the data.

Preventing Loss

There are three steps you can take to prevent data loss, in addition to maintaining continual battery power to the devices as described above. First, we recommend dedicating one computer as your hotsync/backup computer to use for that purpose and for data management only. During the course of a study, avoid installing additional software or downloading files from the Internet, which can interact in unpredictable ways with pre-existing software and may corrupt data files. Second, back up your data religiously on a regular schedule, if possible once a day. Some Universities offer an automatic back-up system that provides additional protection against lost data (i.e., all networked hard drives are backed up each week), but you should not rely solely on that.

Redundancy is the best. Third, be careful not to overwrite data when uploading/hotsyncing. Uploading and hotsyncing are directional processes. Files can be transferred from the device to the master computer or from the master computer to the device. When retrieving data, be sure that the files are transferred from the *device* to the *master computer* to avoid overwriting the data that are stored within the device (note that the direction is reversed when installing software). Also, be sure to develop a system for renaming recently transferred data files. The process of uploading/hotsyncing from the device to the master computer typically creates a data file in the folder associated with that device's unique identifier (e.g., PDA # 12). We recommend renaming this data file and copying it into another master directory to prevent the file from being overwritten from subsequent uploads/hotsyncs (which would use the same data file name).

Data Cleaning

Prior to any analysis, data need to be cleaned and checked for non-compliant responding. In our experience, trials with reaction times faster than 10–30 ms typically indicate participant error (i.e., inadvertently tapping the screen twice for the previous item) and should be excluded from analysis. Non-compliant responding is more difficult to detect and may include responding randomly (e.g., without regard to the item content) or with a set response (e.g., always answering with a "7"). In our lab, we screen for random responding by computing, for each participant, the correlation between two items that should be related if a person is responding honestly (e.g., reports of happy and pleased are usually strongly positively related; reports of happy and sad are rarely positively correlated). We then examine the consistency with which participants do not show normative patterns (we do not test patterns directly relevant to our hypothesis). If we do suspect random responding, then we review their raw data for evidence, and remove that person from the sample *only if* we have determined that the participant has responded randomly. In our experience, this situation rarely occurs, but when it does, we are careful to report it in scientific papers. Set responses can be easier to detect. Running descriptive analyses for each participant can reveal data with little or no variability (variance). In sum, it is each researcher's responsibility to set criteria for exclusion and to apply them fairly across all participants in their sample.

SUMMARY

The richness of data collected from experience sampling reflects the labor and forethought integral to a successful study. After deciding that experience-sampling is the right methodology for your research question (Step 1), and, following careful consideration of your resources (Step 2), you are ready to design your study (Step 3). Consideration of your study design will then allow for the successful selection of software and equipment (Step 4). After ensuring that your equipment is secure (Step 5), then you are ready to implement your study (Step 6). Once your data are collected, backed-up and cleaned (Step 7) you can then analyze and see the fruits of your labors. By approaching experience-sampling studies in a step-by-step fashion, and allowing ample time to think through the preliminary steps, researchers can unlock the power of this methodology.

ACKNOWLEDGEMENTS

This manuscript benefited from conversations with Shige Oishi and Jean-Philippe Laurenceau. Manuscript preparation was supported by NSF grant SBR-9727896 and a grant from the Fetzer Foundation to Lisa Feldman Barrett and Tamlin Conner Christensen.

NOTES

[1]Be aware that institutional review boards may express concern about raffles. So, choose prizes that are not large enough to constitute coercion and include in the informed consent a participant's estimated odds of winning (especially for the grand prize). In our lab, we do not make prizes contingent upon full completion of the study. For example, if participants withdraw from the study before the 5 weeks, they do not get paid, but they do keep any raffle tickets or weekly remunerations that they have accumulated.

[2]When purchasing equipment (e.g., PDAs), it is prudent to buy extra units and store them away for future use. When machines in your original fleet start to fail (and they will!), you can replace them with fresh machines known to be compatible with your current software. [Technology changes rapidly, so you cannot assume either that you will

be able to buy the same machines in the future or that new machines will run in exactly the same way as those in your current fleet.] We recommend purchasing between 10% and 25% (preferably) extra machines. So, if you plan to deploy a fleet of 40 PDAs, consider buying 4 to 10 extra units. We learned this lesson the hard way. After continual use for nearly four years, our fleet of 50 palmtop computers began failing and we had no way to refresh our diminishing stock. We implemented this replacement strategy when purchasing our next fleet of PDAs and so far it has worked very well.

REFERENCES

Bargh, J.A.: 1994, 'The four horsemen of automacticity', in R.S. Wyer and T.K. Srull (eds), Handbook of Social Cognition (Erlbaum, Hillside, NJ), pp. 1–40.

Barrett, D.J. and L. Feldman Barrett: 2000, The Experience-Sampling Program (ESP). (http://www2.bc.edu/∼barretli/esp/).

Bolger, N., A. Davis and E. Rafaeli: 2003, 'Diary methods: Capturing life as it is lived', Annual Review of Psychology 54, pp. 579–616.

Bryk, A.S. and S.W. Raudenbush: 1992, Hierarchical Linear Models: Applications and Data Analysis Methods (Sage Publications, Thousand Oaks, CA, US).

Chalmers, D.J.: 1996, The Conscious Mind: In Search of a Fundamental Theory, Philosophy of Mind Series, Oxford University Press, New York.

Conner Christensen, T., J.V. Wood and L. Feldman Barrett: (2003), 'Remembering everyday events through the prism of self-esteem', Personality and Social Psychology Bulletin. Volume 29(1), pp. 51–62.

Csikszentmihalyi, H. and R. Larson, 1987, 'Validity and reliability of the experience-sampling method', Mental disorders in their natural settings: The application of time allocation and experience-sampling techniques in psychiatry, Journal of Nervous and Mental Disease 175(9), pp. 526–536 (special issue).

Csikszentmihalyi, M. and R. Larson: 1992, 'Validity and reliability of the experience-sampling method', in M.W. deVries (eds), The Experience of Psychopathology: Investigating Mental Disorders in their Natural Settings (Cambridge University Press, Cambridge, UK), pp. 43–57.

de Vries, M., C. Dijkman-Caes and P. Delespaul: 1990, 'The sampling of experience: A method of measuring the co-occurrence of anxiety and depression in daily life', in J.D. Maser and C.R. Cloninger (eds), Comorbidity of Mood and Anxiety Disorders (American Psychiatric Press, Washington, DC), pp. 707–726.

DePaulo, B.M., D.A. Kashy, S.E. Kirkendol, M.M. Wyer and J.A. Epstein: 1996, 'Lying in everyday life', Journal of Personality and Social Psychology 70(5), pp. 979–995.

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 (Cambridge University Press, New York), xvii, pp. 363–373.

Feldman, L.A.: 1995, 'Valence focus and arousal focus: Individual differences in the structure of affective experience', Journal of Personality and Social Psychology 69, pp. 153–166.

Feldman Barrett, L.: 1998, 'Discrete emotions or dimensions? The role of valence focus and arousal focus', Cognition and Emotion 12(4), pp. 579–599.

Feldman Barrett, L.: 2003, 'Feelings or words? Understanding the content in self-report ratings of emotional experience', Manuscript under review.

Feldman Barrett, L. and D.J. Barrett: 2001, 'An introduction to computerized experience sampling in psychology', Social Science Computer Review 19(2), pp. 175–185.

Gable, S.L., H.T. Reis and A.J. Elliot: 2000, 'Behavioral activation and inhibition in everyday life', Journal of Personality and Social Psychology 78(6), pp. 1135–1149.

Greenwald, A.G. and M.R. Banaji: 1995, 'Implicit social cognition: attitudes, self-esteem, and stereotypes', Psychological Review 102(1), pp. 4–27.

Harrison, D.A. and M.E. McLaughlin: 1993, 'Cognitive processes in self-report responses: Tests of item context effects in work attitude measures', Journal of Applied Social Psychology 78, pp. 129–140.

Hormuth, S.E.: 1986, 'The sampling of experiences in situ' , Journal of Personality 54, pp. 262–293.

Hurlburt, R.T.: 1997, 'Randomly sampling thinking in the natural environment', Journal of Consulting and Clinical Psychology 65, pp. 941–949.

Kernis, M.H., D.P. Cornell, C.R. Sun, A. Berry and T. Harlow: 1993, 'There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem', Journal of Personality and Social Psychology 65(6), pp. 1190–1204.

Klein, S.B.: 2001, 'A self to remember: A cognitive neuropsychological perspective on how self creates memory and memory creates self', in C. Sedikides and M. Brewer (eds), Individual Self, Relational Self, Collective Self (Psychology Press, Philadelphia), pp. 25–46.

Litt, M.D., N.L. Cooney and P. Morse: 1998, 'Ecological momentary assessment (EMA) with alcoholics: Methodological problems and potential solutions', Health Psychology 17, pp. 48–52.

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, pp. 771–785.

Nezlek, J.B. and S.L. Gable: 2001, 'Depression as a moderator of relationships between positive daily events and day-to-day psychological adjustment', Personality and Social Psychology Bulletin 27(12), pp. 1692–1704.

Nezlek, J.B., L. Wheeler and H.T. Reis: 1983, 'Studies of social participation', New Directions for Methodology of Social and Behavioral Science 15, pp. 57–73.

Oishi, S. 2002, 'The experiencing and remembering of well-being: A cross cultural analysis', Personality and Social Psychology Bulletin 28(10), pp. 1398–1406.

Penner, L.A., S. Shiffman, J.A. Paty and B.A. Fritzsche: 1994, 'Individual differences in intraperson variability in mood', Journal of Personality and Social Psychology 66(4), pp. 712–721.

Pietromonaco, P. and L. Feldman Barrett: 1997, 'Working models of attachment and daily social interactions', Journal of Personality and Social Psychology 73(6), pp. 1409–1423.

Pogue, D.: 1999, PalmPilot: The Ultimate Guide (2nd ed) (O'Reilly & Associates, Sebastopol, CA).

Reis, H.T. and S.L. Gable: 2000, 'Event sampling and other methods for studying daily experience', in H.T. Reis and C.M. Judd (eds), Handbook of Research Methods in Social and Personality Psychology (Cambridge University Press, New York), pp. 190–222.

Reis, H.T. and L. Wheeler: 1991, 'Studying social interaction with the Rochester Interaction Record', Advances in Experimental Social Psychology 24, pp. 269–318.

Robinson, M.D. and G.L. Clore: (2002), 'Belief and feeling: Evidence for an accessibility model of emotional self-report', Psychological Bulletin 128(6), pp. 934–960.

Ross, M.: 1989, 'Relation of implicit theories to the construction of personal histories', Psychological Review 96, pp. 341–357.

Schwartz, J.E., J. Neale, C. Marco. S.S. Shiffman and A.A. Stone: 1999, 'Does trait coping exist? A momentary assessment approach to the evaluation of traits', Journal of Personality and Social Psychology 77(4), pp. 360–369.

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 and V.S. Cain (eds), The Science of Self-report: Implications for Research and Practice (Erlbaum, Mahwah, NJ), pp. 277–296.

Shiffman, S., J.A. Paty, M. Gnys, J.D. Kassel and C. Elash: 1995, 'Nicotine withdrawal in chippers and regular smokers: Subjective and cognitive effects', Health Psychology 14(4), pp. 301–309.

Snijders, T. and R. Bosker: 1999, Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (Sage Publishers, London).

Snijders, T., R. Bosker and H. Guldemond: 1999, 'The Power analysis IN Two-level designs (PINT) software', Available at http://stat.gamma.rug.nl/multilevel.htm.

Stone, A., R. Kessler and J. Haythornthwaite: 1991, 'Measuring daily events and experiences: Decisions for the researcher', Journal of Personality 59, pp. 575–608.

Stone, A.A. and S. Shiffman: 1994, 'Ecological momentary assessment (EMA) in behavioral medicine', Annals of Behavioral Medicine 16, pp. 199–202.

Tulving, E.: 1985, 'How many memory systems are there?' American Psychologist 40(4), pp. 385–398.

Wheeler, L. and H. Reis: 1991, 'Self-recording of everyday life events: Origins, types, and uses', Journal of Personality 59, pp. 339–354.

Address for correspondence:
LISA FELDMAN BARRETT
Department of Psychology
McGuinn Hall 301
Boston College
Chestnut Hill MA, 02467

E-mail: barretli@bc.edu