CHRISTIE NAPA SCOLLON, CHU KIM-PRIETO and ED DIENER

EXPERIENCE SAMPLING: PROMISES AND PITFALLS, STRENGTHS AND WEAKNESSES

Since its inception in the late 1970s, experience sampling methodology (ESM) has enjoyed an explosion of popularity in psychological research. A literature search for ESM and related terms, such as ecological momentary assessment, on PsychINFO yielded 343 articles and dissertations, most of which have been undertaken in the past dozen years. Much of its popularity can be attributed to its ability to delve beyond single-time self-report measurement to answer complex questions about lives, such as the role of situations in individual functioning, as well as its ability to provide solutions to nagging methodological problems, such as memory biases.

Investigators have long recognized the need for an assessment tool that is more true to life experiences than laboratory assessments, global questionnaires, or observer ratings. Brunswik (1949) and Cattell (1957) addressed the importance of understanding how various psychological variables manifest themselves in different situations in order to understand the full constellation of behaviors and conditions that elicit them. Later, the call for studies of on-line experience was again taken up by Fiske (1971), who wrote that the assessment of on-line experiences should be one of the essential tools in assessing personality. More recently, Funder (2001) brought attention to the necessity of studying personality in a wide variety of settings, and the utility of ESM in meeting that need.

WHAT IS ESM? A BRIEF HISTORY

ESM refers to a method of data collection in which participants respond to repeated assessments at moments over the course of time while functioning within their natural settings. Although no single person or research program can be credited with inventing ESM *per se*, the precursor to today's ESM can be seen as early as Flügel's (1925) 30-day study of mood. However, the methodology that most resembles its current form is usually credited to Csikszentmihalyi et al.'s (1977) investigation

Journal of Happiness Studies **4:** 5–34, 2003.

of adolescents in their natural environments or Brandstaetter's (1983) study of mood across situations. ESM was also used extensively by Diener and his colleagues in the early 1980s for measuring mood across situations (e.g., Diener and Larsen, 1984; Diener and Emmons, 1985). What distinguishes these current forms from their earlier precursors is the introduction of random signaling and attempts to study intra-psychic phenomena.

The use of random sampling of behaviors used in the past by industrial researchers and behavior therapists (e.g., Ayllon and Haughton, 1964; Case Institute of Technology, 1958) were usually observer reports, and hence limited to the study of overt behaviors in institutional settings. The use of self-reports allowed for the sampling of a greater range of situations and investigation of more intra-psychic phenomena; however, participants were often required to keep track of the sampling times themselves, by either keeping a schedule of the times when one would be required to record one's behavior, or else setting alarm watches for the next assessment period (e.g., Barnes, 1956; Brandstaetter, 1983; Diener and Larsen, 1984; Heiland and Richardson, 1957; Hinricks, 1964; Case Institute of Technology, 1958). The disadvantage to this approach is that participants might anticipate alarms (although Diener and Larsen noted that most participants said they usually forgot about the alarms after they programmed the watch).

Fortunately for today's ESM researcher, technological advances have made many former problems moot. The tools of ESM have evolved to allow greater ease of data collection for the researchers as well as the participant. At its nascence, participants carried pagers or alarm watches, along with a stack of paper on which they recorded their responses when signaled. While the paging devices were usually small enough to be carried around comfortably, investigators still had to call the paging device. Other researchers telephoned the participants at home at random times. This approach, too, had its limitations; the participants could only be studied at certain times and while at home (see Stone et al., 1991 for a discussion of the advantages). Alternatives included pre-programmed alarm devices (Hormuth, 1986) that were relatively large.

Today, hand-held computers (a.k.a. personal digital assistants (PDAs) or palmtop computers) can be pre-programmed to signal participants at random moments. Participants can respond directly on the

computer, negating the need to carry around a separate stack of response sheets. Besides the time-saving advantage and the convenience for the participant, the decrease in the cost of electronic goods makes hand-held computers cost-effective as well. The palmtop computers allow data to be directly transferred to statistical software packages or other programs for immediate analysis, and with no data entry, mistakes are minimized or eliminated altogether. Furthermore, participants cannot as easily fake their responses as with the paper–pencil measures (see below).

Types of Experience Sampling

Under the broad umbrella of experience sampling methods, three distinct types of experience sampling exist (Reis and Gable, 2000; Wheeler and Reis, 1991). Interval-contingent sampling refers to the type of data collection in which participants complete self-reports after a designated interval for a pre-set amount of time (e.g., hourly reports, Nowlis and Cohen, 1968; daily reports, Wessman and Ricks, 1966). Event-contingent sampling occurs when participants complete self-reports when a pre-designated event occurs (e.g., reporting after every social interaction, Cote and Moskowitz, 1998). Last, signal-contingent sampling requires participants to complete their self-reports when prompted by a randomly-timed signal. This last form of sampling is what is typically labeled ESM. For the most part, the present paper focuses on signal-contingent sampling. Signal-contingent sampling is advantageous in that it allows for the sampling of a representative schedule of times, and avoids any expectancy effects that may come from having prior knowledge of the sampling period (Alliger and Williams, 1993).

Besides the three broad categories that distinguish between the timing and method of gathering data, other researchers distinguish ESM from three other methods based on the type of data generated. First, thought-sampling, developed independently by Hurlburt (1997) and Klinger (1978–1979), differs from ESM in that it focuses on recordings of inner thoughts and mostly dispenses with external events. Hurlburt (1997) also distinguishes descriptive-experience sampling from ESM, in that the former is used only for gathering qualitative data. Second, Stone and his colleagues distinguish ESM from ecological momentary assessment (EMA). In EMA, measurement is concerned with not just the participant's momentary subjective experience, but also with the

elements of the environment related to momentary experience (Stone et al., 1999). It should be noted that many researchers do not make sharp distinctions between ESM and other random-sampling methods (Hurlburt, 1997; Reis and Gable, 2000).

PROMISES AND STRENGTHS

From the brief outline of ESM's historical roots and description, it is easy to see why ESM and its predecessors have enjoyed researchers' enduring interest. Even decades earlier, when technology made experience sampling difficult, time-consuming, and costly, researchers recognized the need for the type of data made possible through ESM. Five major strengths of experience sampling are discussed in detail below. First, ESM allows researchers to better understand the contingencies of behavior. Second, ESM takes psychology out of the laboratory and into real-life situations, thus increasing its ecological validity. Third, ESM allows for the investigation of within-person processes. Fourth, researchers can avoid some of the pitfalls associated with traditional self-reports, such as memory biases and the use of global heuristics. Fifth, ESM answers the call for the greater use of multiple methods to study psychological phenomena.

Contingencies can be Noted

Mischel (1968) highlighted the problem of understanding individual behavior across situations in his criticism of personality research, and suggested a need for "direct experience sampling" in order to understand the covariation in stimulus conditions and responses of human behavior. In fact, one of the strongest benefits of ESM is that it allows for the investigation of complex questions about the contingencies of behaviors. For example, research exploring the link between extraversion and happiness tracked whether extraverts' greater levels of positive emotion were due to greater time spent in social activities (e.g., Diener et al., 1984; Pavot et al., 1990). But also, ESM has been used to investigate the effects of momentary pleasant affect on social activity (Lucas, 2000). As Lucas (2000) demonstrated, ESM is better suited to this task than global reports because current levels of pleasant affect can predict future social activity, even after controlling for previous levels of social activity.

Through the use of ESM, other researchers have shown the intricacies of the interaction between persons and situations. Brandstatter (1983), for example, found that the types of situations and the social interactions encountered in daily lives were correlated with the personality of the individuals, and that the subsequent emotions were dependent on both situational and personality variables (see also Diener et al., 1984 for an example of situational selection using time-contingent sampling). Others have shown situation interactions with gender (Larson et al., 1994) and culture (Oishi et al., 2002), as well as personality variables (Flory et al., 2000). Indeed, these types of studies allow for a better understanding of the mutual role of the individual and the situation that gives rise to various psychological phenomena.

Ecological Validity

In addition to allowing the researcher to investigate the various situational contingencies of a psychological phenomenon, ESM also permits greater generalizability of the research findings. That is, psychologists can ecologically validate their theoretical concepts and empirical findings in real-life settings. For example, experience sampling studies have identified diurnal and weekly patterns of moods. That is, people tend to experience greater pleasant affect later in the day, versus in the morning, and on the weekends, versus on weekdays, (e.g., Egloff et al., 1995; Larsen and Kasimatis, 1990; Lucas, 2000). In another illustrative example, Lucas (2000) found that individuals who scored high on global measures of extraversion did not engage in more social behaviors on a moment-to-moment basis than individuals who scored low on the global measures. This finding demonstrates the usefulness of ESM in informing our theories; in this case, ESM data challenged the traditional and implicit beliefs about what it means to be extraverted.

Investigating Within-Person Processes

While psychologists over the years have emphasized the need for idiographic research (e.g., Lamiell, 1997; McAdams, 1995; Pelham, 1993), nomothetic investigations have largely formed the mainstay of research. However, with ESM, researchers are not limited to between-person investigations. This is important because within-person analyses can reveal interesting patterns that may be masked at mean levels. For example, researchers are often interested in what emotions occur together. However, this question can be framed at two different levels of analysis,

and can sometimes lead to different conclusions depending on the level of analysis. Correlations computed at the within-person level represent state conceptions of emotions, while between-person correlations computed from aggregated moment data reflect trait conceptions of emotion (see Zelenski and Larsen, 2000, pp. 180–181, for a detailed explanation). In other words, at the within-person level, we are primarily interested in what states go together at a given moment (e.g., Can a person feel both happy and guilty simultaneously?). On the other hand, at the between-person level, we can examine the "long-term structure" of affect (Diener and Emmons, 1985). For example, does knowing a person's mean level of happiness tell us anything about that person's mean level of guilt? Because the two levels are logically independent, positive and negative emotions can be negatively correlated at one level and yet independent or positively correlated at another level (e.g., Diener and Emmons, 1985).

Two studies illustrate the importance of distinguishing between the distinct levels of analysis. Zelenski and Larsen (2000) found that like-valenced emotions were only weakly correlated at the within-person or state level. On the other hand, like-valenced emotions were more strongly correlated at the between-person or trait level (see also Schimmack, this issue). Similarly, Scollon et al. (2002) found cross-cultural similarities in within-person correlations of pleasant and unpleasant affect, but cross-cultural differences in between-person correlations.

Yet another valuable aspect of ESM is its sensitivity to differences within individuals that emerge over time or across situations, in terms of the variability or intensity of behavior and feelings. For instance, in spite of cultural differences in mean levels of pleasant and unpleasant emotion, Oishi and colleagues (2002) found that across cultures, being with friends increased pleasant mood and decreased negative mood. But the benefit of being with others was qualified by culture and gender such that people in some cultures, as well as men, received a greater boost from being with friends.

At an idiographic level, ESM has helped understand the role of resources in personal goal strivings and subjective well-being (e.g., Diener and Fujita, 1995; Emmons, 1986; Zirkel and Cantor, 1990). Case studies also have been conducted using ESM, such as delle Fave and Massimini's (1992) charting of an agoraphobic patient's treatment progress and her experiences and moods during the nine week sampling period.

Reduction in Memory Bias

The problem of memory biases has been noted for years, but researchers were not always able to distill the biases from the self-reports themselves, or understand the processes that underlie the biases in self-reports. For example, problems of global self-reports include biases in retrospective recall (Cutler et al., 1996; Redelmeier and Kahneman, 1996; Ross, 1989), autobiographical memory (Han et al., 1998; Henry et al., 1994; Wang, 2001), and the use of heuristics in response patterns (Robinson and Clore, in press; Schwarz, 1994, 1999; Tversky and Kahneman, 1973).

In response to problems with global self-reports, researchers use ESM to separate on-line experiences from recall and global biases. This is possible because in ESM, not much room for recall bias exists between the signal and the response because of the shortness of the timelag between the signal and the response. Discrepancies between on-line and global self-report measures have been demonstrated in a variety of research areas, such as coping and emotion. For example, Ptacek et al. (1994) found that on-line reports of the usage of coping strategies correlated 0.58 with retrospective reports. Also, Stone and his colleagues (1998) found that in global reports, cognitive coping was underreported while behavioral coping was over-reported.

On the other hand, several studies have demonstrated convergent validity between aggregated experience sampling data and global or retrospective reports (Diener et al., 1995; Feldman Barrett, 1997). One reason for the mixed findings is that there are multiple ways of conceptualizing accuracy (see Thomas and Diener, 1990). For instance, it would be misleading to say that participants were inaccurate because they were not 100% accurate. In fact, the study by Ptacek et al. (1994) demonstrates both accuracy and inaccuracy. After all, a correlation of 0.58 between on-line and recalled reports, while not perfect, still indicates a good deal of overlap between participants' memories and reality.

One alternative is to examine the rank ordering of individuals or groups on both on-line and retrospective measures to determine if the measures converge. For example, Scollon et al. (2002) found that global, on-line, and retrospective measures of emotion converged in that the rank ordering of different groups was preserved across the various measures. Again, retrospective measures were not entirely unrelated to actual experience. However, retrospective measures overlapped considerably with individuals' self-beliefs, and this was consistent across various cultures. Nevertheless, greater cultural differences emerged

with some of the recall measures than with the on-line measures (albeit those displayed differences as well).

To further complicate matters, researchers can ask participants to recall the past in a variety of ways. For example, people are very inaccurate in estimating absolute frequencies (e.g., In the past hour, how many minutes did you feel happy?), but are very good at remembering which type of emotion is experienced relatively more frequently than others (e.g., happy more than sad; Schimmack, in press). Similarly, Scollon et al. (2002) showed that the recall of frequency was often colored by the intensity of on-line experiences presumably because intense emotions are more salient in memory.

Multiple Methods Assessment

ESM is most useful when applied in conjunction with other methods, for instance, traditional global reports. Since Campbell and Fiske's (1959) landmark paper explicating the need to assess multiple traits through multiple methods, psychologists have been acutely aware of the need for multi-method approaches. Mischel (1968) repeated this call in his caution against common method variance, and in recent years Bank et al. (1990) echoed similar warnings of what they called the problem of "glop" associated with common method variance. By allowing the researcher to study the differential effects of on-line versus other measures, ESM enables the researcher to adopt a multi-method strategy.

The study by Diener et al. (1995) provides a good example of the multimethod approach. In applying a multitrait-multimethod (MTMM) analysis to global self-reports, informant reports, and daily reports of emotion, Diener and his colleagues found that the different measures displayed substantial convergent validity. Like-valenced emotions were highly correlated even when measured with different methods, demonstrating that individuals who tended to experience one emotion (e.g., joy) also experienced similar emotions (e.g., love). Additionally, using structural equation modeling to identify the structure of affect, they found that despite the coherence among like-valenced emotions revealed in the MTMM, a model with only pleasant and unpleasant affect did not fit as well a model with specific emotions influencing pleasant and unpleasant affect. Thus, the authors cautioned against using the specific emotion terms interchangeably because the discrete

emotions were still necessary for describing individual differences in emotion.

Clearly, there are accuracies and inaccuracies in retrospective measures. Nevertheless, global recall measures can compliment ESM data because they avoid some of the pitfalls of ESM that are reviewed later. One major advantage of global reports is that they often cover a much longer time frame than can be covered in an experience sampling study, and this makes them better trait-indicators than aggregates over one week. Recently, global reports have been shown to be especially useful in predicting behavioral choices. For example, Wirtz et al. (in press) investigated the differential roles of on-line experiences, retrospective recall, and expectation for determining future behavior. In their study, vacationing students completed self-reports measuring expectations of pleasure, on-line reports of pleasure, and retrospective recall of pleasure. Results indicated that only recalled affect, not on-line experience or expections, directly and strongly predicted the desire to take a similar vacation in the future. Similarly, the strongest predictor of enduring romantic relationships was not partners' daily reports during times spent together, but rather their retrospective reports of their experiences together (Oishi, 2002). These studies suggest that it is absolutely essential to measure both aspects of experience.

PITFALLS AND WEAKNESSES

Clearly, the main strength of experience sampling lies in its ability to provide fine-grained, detailed pictures of human experience. However, investigators should be aware of potential pitfalls before investing both the time and money in an intensive ESM study. The bulk of problems associated with experience sampling studies can be divided into participant issues, situation issues, and measurement and data analytic issues. These are discussed in detail in the next section.

Who are we studying? Participant Issues

Self-Selection Bias and Attrition

Self-selection bias and attrition are potential problems for all studies, but the intensive nature of the data-gathering strategy creates special difficulties for the ESM researcher. A typical study lasts 1–2 weeks, during which participants respond to 2–12 signals per day (see Reis and

Gable, 2000). This is an onerous task for most people. Imagine oneself as a participant. (Or better yet, investigators should try carrying the palmtop device around just as a participant would.) The alarms disrupt one's activities, conversations, and work, and may not only annoy one self but surrounding others as well, such as in church, classrooms, or meetings. Furthermore, even with short forms that only take 1 min to complete, a participant answers 1000 or more questions over the course of an entire study, totaling well over an hour.

Who volunteers for such intrusive studies and who completes them? Who provides the most data? The difficult nature of these tasks might lead certain types of individuals to be over- or underrepresented in ESM studies. Some individuals will refuse to participate outright. The less motivated participants may drop out after a few days of being interrupted during their daily activities. The remaining participants may show greater motivation, conscientiousness, agreeableness, or other characteristics that may not make them a representative sample. After all, participants who forget the palmtop at home half of the time are likely to differ from participants who remember to take the palmtop with them everywhere and everyday; unfortunately, there will be a preponderance of data on the latter.

Motivation

Indeed research shows that motivation plays a significant role in determining whether a participant will successfully complete an ESM study. Wilson et al. (1992) found that poor volition and concentration made it difficult for depressed older people to complete experience sampling studies. Chronic illness also hindered participation (Wilson et al., 1992), although with high motivation, subjects were remarkably perseverant despite difficulties. The degree of motivation needed to complete an ESM study may also vary by groups. Because the signals are interruptive, those who have more time (e.g., unemployed people or college students) might require less motivation than those with little free time (e.g., young harried professionals with children).

Other Sample Limitations

ESM may not be suitable for studying some people or groups. For example, Csikszentmihalyi and Larson (1987) found that blue collar workers in the 1980s found the task too unusual and were less compliant than clerical workers – although we suspect this gap has

narrowed and will continue to do so due to the spread of technology. The elderly and children may be uncomfortable with the task, although Csikszentmihalyi and Larson (1987) report that their youngest participant was 10 and their oldest participant was 85 (although in that study they did not use palmtop computers). Beidel et al. (1991) noted that children in grades 3–6 could successfully complete ESM studies. Additionally, ESM has been used to study the naturally occurring moods of schizophrenics, showing that it is a viable method that can be used with special populations.

Beyond technological familiarity, however, people must be able to hear and respond to the signals. The elderly and people who spend a lot of time in loud, crowded settings (e.g., bartenders) might be excluded from experience sampling studies. Of course, these problems can be circumvented if researchers use devices that send vibrating signals or devices with a high volume setting. Similarly, Wilson et al. (1992) found that elderly participants had trouble reading digital displays. Even if a signal can be heard, the nature of some jobs makes it more difficult to participate (e.g., truck drivers). In fact, it would be dangerous for some people to try to complete reports during their work (e.g., air traffic controllers).

Possible Solutions and Recommendations

Naturally, one source of motivation for participants is money, and monetary incentives have been shown to significantly improve compliance (Lynn, 2001). However, incentives for completion might not be appropriate for some participants, such as anxious or perfectionistic subjects who are already apprehensive and concerned about completing the study. Also, caution must be exercised in deciding how much compensation to provide – a point to which we will return later. Interestingly, in one study reported by Stone et al. (1991), the monetary incentive of $250 resulted in overall poor quality of data (e.g., missing data, evidence of faking, etc.) because the high desirability of the reward attracted participants who were not intrinsically motivated to participate. In other words, when determining the appropriate amount of compensation, more may not be better.

A possibly more effective way of ensuring participant cooperation is to gain participant trust, or establish what Csikszentmihalyi and Larson (1987) called a "viable research alliance" (p. 529). In other words, participants need to understand the importance of the study (without

necessarily knowing the hypothesis), and they need to be thoroughly trained on the procedures and on the usage of the equipment. Furthermore, researchers should convey to participants the importance of continuing with the study even if they forget to carry the palmtop for one or two days (Stone et al., 1991).

Limiting the Number of Variables in the Design

By reducing the burdensome nature of the task, researchers are likely to narrow the gap between those who participate and those who do not or cannot. One way to do this is to signal participants less frequently and select as few items as possible per occasion. A general rule of thumb is: the more signals per day, the shorter the form should be. Unfortunately, the small number of questions one can ask in ESM studies is in itself a limitation of the methodology. On the other hand, multiple items are not necessary for establishing reliability because reliability can be computed from the aggregate of the single items over time (Csikszentmihalyi and Larson, 1987).

Cautious Generalizations

The most compliant participants for experience sampling studies will be conscientious, agreeable, non-depressed, young people who are not too busy – essentially, college students. Thus, researchers need to consider what effects, if any, these subject characteristics may have on the results of the study and the ability to generalize to broad populations. Although generalization issues are endemic to any psychological research based on college samples, because ESM attempts to link basic psychological processes to situations in the daily lives and experiences of the participants, dangers are even greater. That is, the types of situations – classroom learning, living in dorms, meals in cafeterias – encountered by college students are necessarily different from the types of situations that working adults encounter in their daily lives. On the other hand, most of these situations are social and/or involve the pursuit of one's daily goal. In this sense, perhaps college students are not so different from the larger population. However, the generalizability of findings is one that must be explored.

When are we Studying People? Situation Issues

Quality of Data

Even if a variety of participants volunteer, the quality of data is another problem that is exacerbated by the nature of ESM. Stone et al. (1991)

wrote that declining quality of data reporting is estimated to occur after 2–4 weeks of data collection. With wrist watches or beepers and paper-pencil reports, participants can easily fake their responses by completing all their forms in one sitting. This problem can be alleviated by requiring participants to turn in their completed forms on a daily basis. Palmtop computers further reduce the problem of faked responses by recording the exact date and time of each report, making it painfully obvious if a participant completes all forms in one sitting, and allowing researchers to compute the time lag between the signal and response.

Unfortunately, even with the newer technology, people might simply use the same responses across time. Brandstaetter (1983) argued that participants' responses would become more accurate over time through increased self-awareness. However, it is not clear how accuracy and habitual responding can be separated. Individuals may not use the full range of responses over time (as indicated by decreased variability or increased stability in ratings), but this could be indicative of either habitual, repeated responding or greater self-awareness and accuracy (Hormuth, 1986). Stone and colleagues (1991) detail weighting procedures that can help correct for habitual responding, and Reis and Gable (2000) suggest pilot-testing and refining procedures to minimize the problem.

Select Situations

Random sampling from a person's everyday life invites a host of problems not associated with random sampling from a set of stimuli or variables, because ultimately it is the participant who decides whether to respond to an experience sampling signal. There may be some instances in which one is less likely to respond to a signal, or in which it is impossible to do so. Response rates tend to decreases slightly in the evenings (Alliger and Williams, 1993), in the home (Pawlik and Buse, 1982 as cited in Hormuth, 1986), and in places where the signaling device could not be carried, such as swimming pools (Hormuth, 1986) or where the signal could be disruptive (e.g., church). Of course, the ecological strength of the ESM depends on the degree to which a full range of participants' activities and situations are sampled. This requires that subjects respond to signals and complete forms, even when it is inconvenient and they do not feel like doing so. Again this underscores the need to provide participants with explicit instructions and to help them understand the importance of sampling all situations, not simply select ones.

The fact that some situations (e.g., swimming, being in church, operating machinery, driving) will not get sampled is a difficult problem. Although response rates increase if participants are allowed to respond to the signal afterwards, this practice raises its own concerns. Similarly, rare situations or emotions might not be sampled, such as being the victim of crime or intense fear. Even events that might seem frequent, for instance studying among students, might be recorded relatively infrequently in on-line reports (see Wong and Csikszentmihalyi, 1991). For infrequent events, Wheeler and Reis (1991) and Reis and Gable (2000) recommend event-contingent sampling because "the rarer the event in question, the less useful the signal-contingent method – there is not much chance that the signal and the event will coincide" (Wheeler and Reis, 1991, p. 347). Hormuth (1986) argues that while situation selectivity is a potential problem for ESM studies, other methods are even more vulnerable to threats of ecological validity.

Time Lag Between Signal and Response

The quality of ESM data is best if participants respond to signals immediately. The trade-off here is that more responses will be gained if participants are allowed to respond to signals at a more convenient, later time. However, with greater time lag, memory biases and the use of heuristics can contaminate reports, and thus defeat the purpose of experience sampling. Thus, when participants are allowed to make delayed responses, researchers typically restrict responses to no more than 30 min after the signal (e.g., Cerin et al., 2001; Diener and Larsen, 1984; 20 min in Csikszentmihalyi and Larson, 1987; Stone et al., 1998). Fortunately, much of the data on this issue are encouraging. Hormuth (1986) reported that half of his participants responded immediately to the signal, 70% within 3 min, and 80% within 5 min. Similar figures are reported in Csikszentmihalyi and Larson's (1984) study of adolescents: 64% immediately responded when signaled, and 87% responded within 10 min.

Admittedly, the 20 min- or the 30 min-response window is an arbitrary cut-off, although it provides some uniformity as to when to use the data. Future researchers should consider whether differences in the data exist between timely and tardy responses, in terms of types of participants, time periods, events, moods, or other variables of concern. We suspect more industrious participants will respond on time or closer to the time of the signal. Similar to compliance in timeliness is

compliance in response. Future research should aim to identify the factors that increase or decrease response rates. For example, what are the differences in the subject characteristics of those who might respond to 80% of the signals versus those who respond to only 40%? Are some people more willing to respond to signals when they are in a happy mood than when they are feeling irritated? Research showing that positive moods result in increasing helpfulness (Isen, 1970; Isen and Levin, 1972) suggests that this may be a possibility, but it has yet to be tested.

Reactivity

Psychologists must contend with their own version of Heisenberg's Uncertainty Principle – in this case reactivity or the potential for any phenomenon under study to change as a result of measurement or reporting (Wheeler and Reis, 1991). Although reactivity is a problem for many researchers interested in human behavior, it is especially problematic for ESM studies because the repeated assessments may lead people to pay unusual attention to their internal states and own behavior. For example, completing mood measures 7 times a day might alert someone to insights such as, "I am the kind of person who is sad a lot," or "I am happy when I am with my friends." Reflections of the latter sort, in particular, may lead to behavioral changes such as spending more time with one's friends which in turn may change the person's moods. With non-random sampling techniques (i.e., event-contingent or interval-contingent sampling), people might look for events or anticipate behaviors or situations (Hormuth, 1986). In fact, self-monitoring was believed to be so strongly associated with experience sampling that behaviorists used this method as a tool for behavior modification (Wheeler and Reis, 1991). Schimmack (pers. comm., 2002) recalls one incident in which a participant was advised by his therapist not to participate in an experience sampling study for fear that the constant self-awareness would trigger a relapse. Similarly, alcoholics who participated in an ESM study said that reporting on drinking made them more aware of their drinking, although greater awareness did not lead to behavioral changes (Litt et al., 1998).

In a provocative study on premenstrual symptoms, Ruble (1977) manipulated participants' beliefs about their menstrual cycle and showed its effects on attending to bodily symptoms. Some women were led to believe they were premenstrual while others were led to believe they were not, when in fact all of them were tested 6 days before

the onset of their next menses. The women who thought they were premenstrual reported greater pain, water retention, change in eating habits, and sexual arousal than women who were led to believe they were midcycle. Ruble's (1977) findings suggested that beliefs about premenstrual status alerted the women to look for symptoms that they normally might not have noticed. Arguably, because participants in Ruble's study were asked to report their symptoms for the "last day or two," the observed differences in symptoms between the groups may have been due to implicit theories of menstruation (see also McFarland et al., 1989; Ross, 1989), rather than any differences in attention.

Few studies have directly examined the reactivity effects of experience sampling. Cruise et al. (1996) had chronic arthritis sufferers complete pain dairies several times a day and observed that pain levels and affect ratings remained constant over time, suggesting no reactivity effects. Cerin et al. (2001) randomly assigned participants in a week-long study to either (a) an ESM condition, (b) a repeated measures condition that sampled emotions four times, roughly once every three days, or (c) a retrospective assessment condition that asked participants to recall their emotions on four previous days. The authors compared ratings on four days using (a) the average of the day's ratings for the ESM participants, (b) the day's rating made by the repeated measures group, and (c) the recalled ratings made by the retrospective group. Results revealed higher estimates of feeling energetic, enjoyment, and worry, and lower estimates of feeling anger and irritation, in retrospective reports than in on-line or daily reports. Arguably, these results are inconclusive because group differences may have been driven by memory reconstruction rather than reactivity *per se*. The pattern of emotions for the retrospective group appeared to conform to implicit beliefs of how a person should feel before competition.

Besides priming participants to pay attention to certain states, the intrusive nature of the ESM experiment may actually *make* participants more irritated. This would certainly have ethical as well as methodological implications. Fortunately, Cerin and colleagues (2001) did not find that use of the ESM increased negative mood due to its intrusiveness. In fact, ESM measures of cognitive intrusion were lower than repeated or retrospective measures.

Even if experience sampling does not have an effect on participants' current moods, some evidence suggests that repeated assessments might influence the recall of emotions. For instance, Thomas and Diener

(1990) found that daily and momentary reports correlated higher with retrospective estimates of emotion than with prospective estimates. Similarly, Schimmack (in press) found that absolute estimates of emotions after a daily diary study were more accurate than those before the study. On the other hand, if experience sampling influences accuracy of recall, then we would expect greater accuracy with more repeated assessments. However, Thomas and Diener (1990) found that momentary reporting did not have a greater effect on retrospective accuracy than daily reporting.

In short, we currently know surprisingly little about the effects of ESM on a person's subjective experience. Clearly, more studies are needed. Until reactivity is better understood, researchers cannot be certain that ESM is indeed tapping a phenomenon as it exists, or as it has been transformed by measurement. Future research on reactivity can lead to improvements in testing procedures to minimize reactivity. For example, if researchers find that responding to the same question five times a day results in greater self-monitoring of behavior and alterations in the variable in question, but responding to the same question twice a day does not, this information can guide the design of future research. Because such information is not available, the number of sampling occasions each day typically represents a compromise between maximizing the number of data points while minimizing participant drop-out.

Future research should also explore whether reactivity differs across groups and across variables. Many ESM studies of emotion have, thus far, been conducted in Western cultures where the norm is to think about one's own emotional states (cf. Ji et al., 2000). Cultures that are less emotion-focused may experience greater reactivity to the constant reporting of emotions. Even within cultures, however, differential reactivity may even occur among people of different classes or educational levels (e.g., Csikszentmihalyi and Larson, 1987), or may vary according to what is being measured. For example, attention to anxiety or worry may trigger cognitions or ruminations that may lead to greater anxiety and worry (e.g., Mathews and MacLeod, 1994; Nolen-Hoeksema, 1993).

The Limits of Self-Report

Although the on-line reporting of emotions eliminates retrospective bias to some extent, it is still subject to some of the same problems

as any self-report measure (see Schwarz, 1999). Social desirability, cognitive biases, and cultural norms might influence responses even at the momentary level of reporting. For example, if there is a cultural norm that feeling negative emotions is undesirable, there may be reluctance to reporting feelings such as sadness. Similarly, highly defensive people may "filter" their responses, and even the most honest person might find it difficult to report on some states, such as unconscious motives or feelings (Shedler et al., 1993).

Experience sampling also relies on the use of numerical scales that, although useful among college participants, might be inappropriate or awkward for some people. In fact, Wilson and colleagues (1992) found that Likert scales were too confusing for elderly depressed participants, whereas visual scales were better understood.

Perhaps the best remedy to these problems is to take a multi-method approach and supplement ESM with other non-self-report measures. Promising possibilities include informant reports, on-line physiological measures (e.g., ambulatory measures, Fahrenberg and Myrteck, 2001), or on-line ambient recording, a new methodology developed by Mehl et al. (in press) that captures random 30-s recordings of a participant's surrounding sounds. However, no other method besides self-reports is able to capture the hedonic tone of people's emotional experiences, an aspect of emotion phenomena that is arguably most relevant to subjective well-being research.

The Issue of Scaling

Schwarz (1999) demonstrated that the time frame of the question can influence participants' responses. For example, when asked to recall how often they felt angry in the past year, participants' estimates were lower in frequency but higher in intensity than when they recalled how often they felt angry in the past week (Winkielman et al., 1998). Presumably as the time frame increases, participants discount low levels of anger and retrieve instances of more salient, intense anger. But with repeated assessments every few hours or so, what frame of reference do respondents use? Some researchers instruct participants to respond according to how they were feeling at the moment, while others instruct participants to recall their experiences since the last report. In any case, the time of reference is much shorter than with global reports. The threshold for what is considered an angry state, for example, might be lower when one considers the past few hours as opposed to the

past week. Additionally, participants might rate their present state in reference to their previous states (e.g., Compared to my other reports, how happy am I right now?). Thus, the meaning of momentary reports might change compared to between-subject responses. These are empirical questions that, unfortunately, have not received much attention thus far.

Data Issues: Now What?

Once the data are successfully collected, the ESM researcher is often left staring at 10 000 to 50 000 data points, wondering how to proceed. After all, as Larson and Delespaul (1992) note, "ESM data have a complexity which defies textbook analysis" (p. 58). With the ability to capture more of the complexities of life comes the complexities of data analysis, and ESM researchers face the challenge of choosing a statistical strategy that addresses the unique nature of the data.

Aggregation Issues

One way to handle the massive amount of data that experience sampling studies generate is to aggregate momentary data, for example, computing a mean intensity rating of happiness for each participant over the entire week. This procedure has the main advantage of being straightforward and manageable. Furthermore, aggregation increases reliability and gives higher correlations (Epstein, 1980). But does the meaning of one's construct change as variables are aggregated over time? Diener and Larsen (1984) found that personal consistency and stable patterns emerged when responses were aggregated over several occasions, whereas for single moments, there was little consistency for most variables. Thus, the two levels of analysis are independent of one another, and aggregate level data cannot be used to make assertions about what occurs at the momentary level (Snijders and Bosker, 1999). Ultimately, the nature of one's research questions should specify the level of inquiry (see Larson and Delespaul, 1992), and one's theories will determine which level is most appropriate.

A second concern is that response sets might get amplified through aggregation. For example, factor analysis at the within-person level, which controls for response styles, reveals more negative correlations between positive and negative affect than between-person correlations. However, positive and negative affect are more independent (e.g., Diener and Emmons, 1985) or positively related (e.g., Diener

et al., 1985; Larsen and Diener, 1987; Schimmack and Diener, 1997) when aggregated states are used. This difference in correlations between momentary data and aggregated data may be partly an effect of response style, such as number use, which might cancel the inverse relation between different types of emotion. Although Schimmack and his colleagues (Schimmack, this issue; Schimmack et al., 2002) have documented the operation of response sets on aggregated data, their results show that the effect is negligible and unlikely to bias correlations in a strong way. Nevertheless, the problem of aggregation is central to subjective well-being research because of the increasing reliance on aggregate experience sampling data as measures of trait affect.

Challenges in Analyzing Data

ESM data can be analyzed at the event, subject, and group level (Larson and Delespaul, 1992). At the event level, it is typical to count frequencies of certain events. However, the researcher must take care that first, the event is common enough that the frequency count is meaningful, and second, that a few individuals who experience a given event frequently do not skew mean levels. For example, at a cursory glance, the researcher may find that attending church service may be associated with heightened feelings of joy and gratitude. However, a closer look may show that most of the participants did not respond to the signal while actually attending the service because of its disruptive nature. The few participants who did respond might therefore have a greater influence on the overall results. While some of these issues may be solved by conducting a subject-level analysis supplement to the event-level analysis, events that do not recur several times a day, such as church attendance, may be too rare for such analysis to be meaningful.

If the target variable can be recorded with requisite frequency, then the researcher can simply consider differences in the within-person variability in the frequency of the target event or behavior (Larson and Delespaul, 1992). Furthermore, the researcher can make use of multilevel modeling, such as hierarchical linear modeling (HLM), that allow the simultaneous estimation of within- and between-person effects, and the interactions between the within and the between variables (Reis and Gable, 2000). Multilevel modeling is particularly useful in analyzing ESM data because it allows the researcher to take full advantage of the fact that multiple data points were gathered from a single individual. Furthermore, multilevel analyses tend to handle missing data well

(Snijders and Bosker, 1999); thus, the missed signals by the respondents pose a smaller problem for the researcher.

Another challenge in the analysis of ESM data is the existence of time dependencies in the data. That is, not only are the data nonindependent within individuals, various patterns that may be present in the data may be time-dependent as well (Reis and Gable, 2000; West and Hepworth, 1991). As mentioned earlier, various factors can influence mood, such as time of day (Rusting and Larsen, 1998), day of week (Egloff et al., 1995; Larsen and Kasimatis, 1990), or mood from the previous day or moment (Larsen, 1987). Researchers should be careful to control for such factors when modeling their data. Additionally, time series analysis may help clarify the time-dependencies in the data. For example, spectral analysis is one way in which the frequency of change over time can be analyzed, as opposed to simple examination of within-person standard deviations. In other words, fluctuation in the responses can be understood idiographically, but also can be used as an index through which between-person comparisons can be made (see Larsen, 1987, for more in-depth explanation and example of this analytic technique).

Because spectral analysis assumes a curvilinear relation, one needs to first make sure that the data support such a relationship. For example, Larsen and Kasimatis (1990) report that daily fluctuations in mood follow a 7-day cycle, such that the pattern assumes the shape of a sine curve. That is, a curvilinear relation between the day of the week and mood exists, such that the daily hedonic level increases from a low-point on Monday to a high-point on Saturday, then drops again to the Monday low. In contrast, a linear relationship would assume that a straight line would best capture the relationship between the days of the week and hedonic levels. That is, hedonic level would either continue to increase, decrease, or remain the same over the number of days.

While new analytic methods allow researchers to better understand the various nested relationships within the data, the robustness of these new methods remains largely untested (Reis and Gable, 2000). Furthermore, assumptions that underlie the analytic strategy must also be considered before these methods can be used. For example, while hierarchical linear models assume a linear relationship, spectral analytic models assume a curvilinear relationship. Unless these assumptions are met, such analyses may be inappropriate or misleading. New

statistical techniques are needed that take into account the time dependencies in ESM data, and which are compatible with the uneven time periods between assessments.

Special Ethical Issues

As mentioned earlier, participants must be provided with just compensation for their cooperation in the burdensome tasks of an experience sampling study. But as Tennen et al. (1991) note, "At what point does incentive shade into coercion?" (p. 320). For this reason, IRB committees may frown upon too much compensation for participation, leaving ESM researchers with the challenge of providing fair, yet non-coercive, incentives.

Also, through the use of ESM, are researchers intruding into people's private lives? Compared to other forms of behavioral assessment or experiments, the privacy of individuals is considerably protected in experience sampling studies. First and foremost, the participant retains a great deal of control over experience sampling information – for instance, when to respond to signals, and what to report. Of course, from the researcher's perspective, this is a limitation. For example, as Csikszentmihalyi and Larson (1987) pointed out, if employers used ESM to assess worker productivity, or if ESM were used to investigate sensitive, perhaps even illegal activities, then the veracity of such reports would be questionable. But still, from the participants' perspective, privacy would be preserved. However, technological advances may force researchers to reexamine the boundaries of privacy and science in the future, for example if global positioning satellite devices were to be implanted in palmtop computers.

FUTURE DIRECTION

Researchers looking for a panacea in ESM will be sorely disappointed. Like all other methods and measures, it has both strengths and weaknesses. Clearly, future research should turn towards testing reactivity, clarifying the meaning of scales, and validating data analytic procedures such as aggregation and modeling techniques. Until these issues are resolved, experience sampling measures, although highly valuable and informative, should not be accepted as the "gold-standard" of measures.

Nevertheless, ESM remains a powerful tool that can aid researchers in tackling new questions as well as investigating current questions in greater depth. Our lengthy discussion of the pitfalls associated with ESM should not discourage the potential and current ESM user. We have highlighted some areas of concern so that these concerns can be resolved and addressed. By drawing attention to these deficits, we hope to stimulate research that will address some of these concerns and bolster ESM's usage in psychological research.

CONCLUSIONS

In spite of ESM's relatively short history, it is enjoying increasing popularity among researchers in various fields. The growing affordability of the tools used in ESM, such as palmtop computers, should make ESM even more attractive to many researchers. Of course, ESM entails greater commitment in terms of time and monetary resources than a single time questionnaire, but it is likely to produce greater payoffs, in terms of deeper understanding and more detail. Plus, when the researcher is aware of the pitfalls and takes precautions to avoid potential hazards, ESM can be a boon to research. Ultimately, the nature of one's research questions should guide the decision of whether to use experience sampling.

When paired with other methods of assessment, ESM may be particularly beneficial to the study of happiness. First, given the definition of subjective well-being (SWB: a combination of pleasant affect, low unpleasant affect, and satisfaction with life, see Diener, 1984), ESM is most useful in assessing the affective components of SWB, particularly because these components are vulnerable to distortions in memory (Kahneman, 1999). Experience sampling of the cognitive component, however, may not be necessary because this component shows high consistency (see Conner et al., this issue; Diener and Larsen, 1984). On the other hand, momentary satisfaction judgments might still prove useful in understanding the different ways in which people incorporate satisfaction with specific domains or satisfaction on specific occasions with global life satisfaction (Diener et al., 2002). Second, experience sampling can enable SWB research to move beyond the understanding of the demographics of happiness towards identifying

the specific mechanisms and causal influences of SWB (Diener and Biswas-Diener, 2000).

The potential pitfalls outlined above underscore the need to pilot test materials before embarking on long studies with actual participants. Oneself and one's research group make excellent trial participants, and we recommend that researchers carry the signaling device for the expected duration of the study, or for at least several days, prior to actual data collection. This allows the researcher to experience first-hand the various difficulties that a participant may encounter and make necessary adjustments to the research plan instead of subjecting participants to impossible tasks. This experience also allows the researcher to better train participants, and with better training, greater compliance will follow. Compliance is especially important in ESM studies because the ecological strength of ESM is dependent on this compliance. But if the participant is compliant only under certain circumstances, such as when he or she is bored and has nothing else to do, or when he or she is in a good mood, then the assumption of ecological validity is weakened.

In sum, while ESM provides the researcher with a helpful tool with which one can further one's research goals, like any other tool, its utility can only be measured by the care with which the researcher plans and conducts the research, analyzes the findings, and interprets the results. ESM is one of the more effective tools in a growing repertory of means from which the researcher can delve into important psychological questions.

REFERENCES

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

Ayllon, T. and E. Haughton: 1964, 'Modification of symptomatic verbal behavior of mental patients', Behavioral Research Therapy 2, pp. 87–98.

Bank, L., T.J. Dishion, M. Skinner and G.R. Patterson: 1990, 'Method variance in structural equation modeling: Living with "glop,"' in G.R. Patterson (ed.), Depression and Aggression in Family Interaction. Advances in Family Research (Lawrence Erlbaum Associates Inc, Hillsdale, NJ), pp. 247–279.

Barnes, R.M.: 1956, Work Sampling (Wiley, New York).

Beidel, D.C., A.M. Neal and A.S. Lederer: 1991, 'The feasibility and validity of a daily diary for the assessment of anxiety in children', Behavior Therapy 22, pp. 505–517.

Brandstaetter, H.: 1983, 'Emotional responses to other persons in everyday life situations', Journal of Personality and Social Psychology 45, pp. 871–883.
Brunswik, E.: 1949, Systematic and Representative Design of Psychological Experiments (University of California Press, Berkeley, CA).
Campbell, D.T. and D.W. Fiske: 1959, 'Convergent and discriminant validation by the multitrait-multimethod matrix, Psychological Bulletin 56, pp. 81–105.
Case Institute of Technology: 1958, An operations research study of the scientific acitivy of chemists (Operations Research Group, Cleveland, Ohio).
Cattell, R.B.: 1957, Personality and Motivation: Structure and Measurement (World Book, New York).
Cerin, E., A. Szabo and C. Williams: 2001, 'Is the experience sampling method (ESM) appropriate for studying pre-competitive emotions?', Psychology of Sport and Exercise 2, pp, 27–45.
Conner, T., L. Feldman Barrett, E. Bliss-Moreau, K. Lebo and C. Kaschub: 'A practical guide to experience-sampling procedures', Journal of Happiness Studies (this issue).
Cote, S. and D.S. Moskowitz: 1998, 'On the dynamic covariation between interpersonal behavior and affect: Prediction from neuroticism, extroversion, and agreeableness', Journal of Personality and Social Psychology 75, pp. 1032–1046.
Cruise, C.E., J. Broderick, L. Porter, A. Kaell and A.A. Stone: 1996, 'Reactive effects of diary self-assessment in chronic pain patients', Pain 67, pp. 253–258.
Csikszentmihalyi, M. and R. Larson: 1984, Being Adolescent: Conflict and Growth in the Teenage Years (Basic Books, New York).
Csikszentmihalyi, M. and R. Larson: 1987, 'Validity and reliability of the experience sampling method', Journal of Nervous and Mental Disease 175, pp. 526–537.
Csikszentmihalyi, M., R. Larson and S. Prescott: 1977, 'The ecology of adolescent activity and experience', Journal of Youth and Adolescence 6, pp. 281–294.
Cutler, S.E., R.J. Larsen and S.C. Bruce: 1996, 'Repressive coping style and the experience and recall of emotion: A naturalistic study of daily affect', Journal of Personality 64, pp. 379–405.
delle Fave, A. and F. Massimini: 1992, 'The ESM and the measurement of clinical change: A case of anxiety disorder', in M.W. deVries (ed.), The Experience of Psychopathology: Investigating Mental Disorders in Their Natural Settings (Cambridge University Press, New York) pp. 280–289.
Diener, E. and R. Biswas-Diener: 2000, New Directions in Subjective Well-Being Research: The Cutting Edge. Unpublished manuscript. University of Illinois, Urbana-Champaign.
Diener, E. and C. Diener: 1996, 'Most people are happy', Psychological Science 7, pp. 181–185.
Diener, E. and R.A. Emmons: 1985, 'The independence of positive and negative affect', Journal of Personality and Social Psychology 47, pp. 1108–1117.
Diener, E. and F. Fujita: 1995, 'Resources, personal strivings, and subjective well-being: A nomthetic and idiographic approach', Journal of Personality and Social Psychology 68, pp. 926–935.

Diener, E. and R.J. Larsen: 1984, 'Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses', Journal of Personality and Social Psychology 47, pp. 871–883.

Diener, E., R.J. Larsen and R.A. Emmons: 1984, 'Person × Situation interactions: Choice of situations and congruence response models', Journal of Personality and Social Psychology 47, pp. 580–592.

Diener, E., R.J. Larsen, S. Levine and R.A. Emmons: 1985, 'Intensity and frequency: Dimensions underlying positive and negative affect', Journal of Personality and Social Psychology 48, pp. 1253–1265.

Diener, E., R.E. Lucas, S. Oishi and E.M. Suh: 2002, 'Looking up and looking down: Weighting good and bad information in life satisfaction judgments', Personality and Social Psychology Bulletin 28, pp. 437–445.

Diener, E., H. Smith and F. Fujita: 1995, 'The personality structure of affect', Journal of Personality and Social Psychology 69, pp. 130–141.

Egloff, B., A. Tausch, C.-W. Kohlmann and H.W. Krohne: 1995, 'Relationships between time of day, day of the week, and positive mood: Exploring the role of the mood measure', Motivation and Emotion 19, pp. 99–110.

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

Epstein, S.: 1980, 'The stability of behavior: Implications for psychological research', American Psychologist 9, pp. 790–806.

Fahrenberg, J. and M. Myrtek (eds): 2001, Progress in Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies (Hogrefe and Huber Publishers, Kirkland, WA).

Feldman Barrett, L.: 1997, 'The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion', Personality and Social Psychology Bulletin 23, pp. 1100–1110.

Fiske, D.: 1971, Measuring the Concepts of Personality (Aldine, Chicago).

Flory, J.D., K. Raeikkoenen, K.A. Matthews and J.F. Owens: 2000, 'Self-focused attention and mood during everyday social interactions', Personality and Social Psychology Bulletin 26, pp. 875–883.

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

Funder, D.C.: 2001, 'Personality', Annual Review of Psychology 52, pp. 197–221.

Han, J.J., M.D. Leichtman and Q. Wang: 1998, 'Autobiographical memory in Korean, Chinese, and American children', Developmental Psychology 34, pp. 701–713.

Heiland, R.E. and W.J. Richardson: 1957, Work Sampling (McGraw-Hill, New York).

Henry, B., T. Moffitt, A. Caspi, J. Langley and P.A. Silva: 1994, 'On the "remembrance of things past": A longitudinal evaluation of the retrospective method', Psychological Assessment 6, pp. 92–101.

Hinricks, J.R.: 1964, 'Communications activity of industrial research personnel', Personnel Psychology 17, pp. 193–204.

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.

Isen, A.M.: 1970, 'Success, failure, attention, and reaction to others: The warm glow of success', Journal of Personality and Social Psychology 15, pp. 294–301.

Isen, A.M and P.F. Levin: 1972, 'Effect of feeling good on helping: Cookies and kindness', Journal of Personality and Social Psychology 21, pp. 384–388.

Ji, L.-J., N. Schwarz and R.E. Nisbett: 2000, 'Culture, autobiographical memory, and behavioral frequency reports: Measurement issues in cross-cultural studies', Personality and Social Psychology Bulletin 26, 585–593.

Kahneman, D.: 1999, 'Objective happiness', in D. Kahneman, E. Diener and N. Schwarz (eds), Well-Being: The Foundations of Hedonic Psychology (Russell Sage Foundation, New York), pp. 3–25.

Klinger, E.: (1978–1979), 'Dimensions of thought an dimagery in normal waking states', Journal of Altered States of Consciousness 4, 97–113.

Lamiell, J.T.: 1997, 'Individuals and the differences between them', in R. Hogan, J.A. Johnson and S.R. Briggs (eds), Handbook of Personality Psychology (Academic Press, San Diego, CA) pp. 117–141.

Larsen, R.J.: 1987, 'The stability of mood variability: A spectral analytic approach to daily mood assessments', Journal of Personality and Social Psychology 52, pp. 1195–1204.

Larsen, R.J. and E. Diener: 1987, 'Affect intensity as an individual difference characteristic: A review', Journal of Research in Personality 21, pp. 1–39.

Larsen, R.J., M. Kasimatis: 1990, 'Individual differences in entrainment of mood to the weekly calendar', Journal of Personality and Social Psychology 58, pp. 164–171.

Larson, R.W., M.H. Richards and M. Perry-Jenkins: 1994, 'Divergent worlds: The daily emotional experience of mothers and fathers in the domestic and public spheres', Journal of Personality and Social Psychology 67, pp. 1034–1046.

Larson, R. and P.A.E.G. Delespaul: 1992, 'Analyzing experience sampling data: A guidebook for the perplexed', in M.W. deVries (ed.), The Experience of Psychopathology: Investigating Mental Disorders in their Natural Settings (Cambridge University Press, New York), pp. 58–78.

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

Lucas, R.E.: 2000, 'Pleasant affect and sociability: Towards a comprehensive model of extraverted feelings and behaviors', Unpublished Doctoral Dissertation, University of Illinois, Urbana-Champaign.

Lynn, P.: 2001, 'The impact of incentives on response rates to personal interview surveys: Role and perceptions of interviewers', International Journal of Public Opinion Research 13, pp. 326–336.

Mathews, A. and C. MacLeod: 1994, 'Cognitive approaches to emotion and emotional disorderes', Annual Review of Psychology 45, pp. 25–50.

McAdams, D.P.: 1995, 'What do we know when we know a person?' Journal of Personality 63, pp. 365–396.

McFarland, C., M. Ross, N. DeCourville: 1989, 'Women's theories of menstruation and biases in recall of menstrual symptoms', Journal of Personality and Social Psychology 57, pp. 522–531.

Mehl, M.R., J.W. Pennebaker, M. Crow, J. Dabbs and J. Price: 2001, 'The electronically activated recorder (EAR): A device for sampling naturalistic daily activities and conversations', Behavioral Research Methods, Instruments, and Computers, Vol. 33, pp. 517–523.

Mischel, W.: 1968, Personality and Assessment (Wiley, New York).

Nolen-Hoeksema, S.: 1993, 'Sex differences in control of depression', in P.D.M. Wegner and J.W. Pennebaker (eds), Handbook of Mental Control (Prentice-Hall, New Jersey), pp. 306–324.

Nowlis, D.P. and A.Y. Cohen: 1968, 'Mood-reports and the college natural setting: A day in the lives of three roommates under academic pressure', Psychological Reports 23, pp. 551–566.

Oishi, S.: 2002, 'The function of daily vs. retrospective judgments of satisfaction in relationship longevity' (Manuscript in preparation).

Oishi, S., E. Diener, C.N. Scollon and R. Biswas-Diener.: 2002, 'Cross-situational consistency of affective experiences across cultures', Journal of Personality and Social Psychology (submitted).

Pavot, W., E. Diener and F. Fujita.: 1990, 'Extraversion and happiness', Personality and Individual Differences 11, pp. 1299–1306.

Pelham, B.W.: 1993, 'The idiographic nature of human personality: Examples of the idiographic self-concept', Journal of Personality and Social Psychology 64, pp. 665–677.

Ptacek, J.T., R.E. Smith, K. Espe and B. Raffety.: 1994, 'Limited correspondence between daily coping reports and retrospective coping recall', Psychological Assessment 6, pp. 41–49.

Redelmeier, D.A. and D. Kahneman.: 1996, 'Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures', Pain 66, pp. 3–8.

Reis, H.T. and S.L. Gable.: 2000, 'Event-sampling and other methods for studying everyday 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).

Robinson, D. and G.L. Clore.: 2002, 'Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes', Journal of Personality and Social Psychology 83, pp. 198–215.

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

Rusting, C.L. and R.J. Larsen.: 1998, 'Diurnal patterns of unpleasant mood: Associations with neuroticism, depression, and anxiety', Journal of Personality, 66, pp. 85–103.

Ruble, D.N.: 1977, 'Premenstrual symptoms: A reinterpretation', Science 197, pp. 291–292.

Schimmack, U.: 'Affect measurement in experience sampling research', Journal of Happiness Studies (this issue).

Schimmack, U.: 'Frequency judgments of emotions: The cognitive basis of personality assessment', in P. Sedelmeier and T. Betsch (eds), Frequency Processing and Cognition (Oxford University Press, Oxford).

Schimmack, U. and E. Diener.: 1997, 'Affect intensity: Separating intensity and frequency in repeatedly measured affect', Journal of Personality and Social Psychology 73, pp. 1313–1329.

Schimmack, U., U. Bockenholt and R. Reisenzein.: 2002, 'Response styles in affect ratings: Making a moutain out of a molehill', Journal of Personality Assessment 78, pp. 461–483.

Schwarz, N.: 1994, 'Judgment in a social context: Biases, shortcomings, and the logic of conversation', in M. Zanna (ed.), Advances in Experimental Social Psychology, vol. 26 (Academic Press, San Diego, CA) pp. 123–162.

Schwarz, N.: 1999, 'Self-reports: How the questions shape the answers', American Psychologist 54, pp. 93–105.

Shedler, J., M. Mayman and M. Manis.: 1993, 'The illusion of mental health', American Psychologist 48, pp. 1117–1131.

Scollon, C.N., E. Diener, S. Oishi and R. Biswas-Diener.: 2002, 'Culture, self-concept, and memory for emotions' (Manuscript under review).

Scollon, C.N., E. Diener and S. Oishi.: 2002, 'An Experience Sampling and Cross-Cultural Investigation of Dialectical Emotional Experience' (Manuscript in preparation).

Snijders, T.A.B. and R.J. Bosker.: 1999, Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, (Sage Publications, Thousand Oaks, CA).

Stone, A.A., R.C. Kessler and J.A. Haythornwaite.: 1991, 'Measuring daily events and experiences: Decisions for the researcher', Journal of Personality 59, pp. 575–607.

Stone, A.A., J.E. Schwartz, J.M. Neale, S. Shiffman, C.A. Marco, M. Hickcox, J. Paty, L.S. Porter and L.J. Curise.: 1998, 'A comparison of coping assessed by ecological momentary assessment and retrospective recall', Journal of Personality and Social Psychology 74, pp. 1670–1680.

Stone, A.A., S.S. Shiffman and M.W. DeVries.: 1999, 'Ecological momentary assessment', in D. Kahneman, E. Diener and N. Schwarz (eds), Well-Being: Foundations of a Hedonic Psychology (Russell Sage Foundation, New York), pp. 26–39.

Tennen, H., J. Suls and G., Affleck.: 1991, Personality and daily experience: The promise and the challenge. Journal of Personality 59, pp. 313–337.

Thomas, D.L. and E. Diener.: 1990, 'Memory and accuracy in the recall of emotions', Journal of Personality and Social Psychology 59, pp. 291–297.

Tversky, A. and D. Kahneman.: 1973, Judgment Under Uncertainty: Heuristics and Biases (Oregon Research Institute, Eugene, OR).

Wang, Q.: 2001, Culture effects on adults' earliest childhood recollection and self-description: Implications for the relation between memory and the self. Journal of Personality and Social Psychology 81, pp. 220–233.

Wessman, A.E. and D.F. Ricks.: 1966, Mood and Personality (Holt, Rinehart and Winston, New York).

West, S.G. and J.T. Hepworth.: 1991, Statistical issues in the study of temporal data: Daily experiences. Journal of Personality 59, pp. 609–662.

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

Wilson, K.C.M., R. Hopkins, M.W. deVries and J.R.M. Copeland.: 1992, 'Research alliance and the limit of compliance: Experience sampling with the depressed elderly', in M.W. deVries (ed.), The Experience of Psychopathology: Investigating Mental Disorders in their Natural Settings (Cambridge University Press, New York).

Winkielman, P., B. Knäuper and N. Schwarz.: 1998, 'Looking back at anger: Reference periods change the interpretation of (emotion) frequency questions', Journal of Personality and Social Psychology 75, pp. 719–728.

Wirtz, D., J. Kruger, C.N. Scollon and E. Diener.: 2002, 'What to do on Spring Break? The role of predicted, online, and remembered experience in future choice' Psychological Science (in press).

Wong, M., and M. Csikszentmihalyi.: 1991, 'Motivation and academic achievement: The effects of personality traits and the quality of experience', Journal of Personality 59, pp. 539–574.

Zelenski, M. and R.J. Larsen.: 2000, 'The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data', Journal of Research in Personality 34, pp. 178–197.

Zirkel, S. and N. Cantor.: 1990, 'Personal construal of life tasks: Those who struggle for independence', Journal of Personality and Social Psychology 58, pp. 172–185.

Address for Correspondence:
CHRISTIE N. SCOLLON
603 *East Daniel Street, Champaign*
IL 61820, USA

E-mail: scollon@s.psych.uiuc.edu