

Chapter 3

Validity and Reliability of the Experience-Sampling Method

Mihaly Csikszentmihalyi and Reed Larson

To understand the dynamics of mental health, it is essential to develop measures for the frequency and the patterning of mental processes in everyday-life situations. The Experience-Sampling Method (ESM) is an attempt to provide a valid instrument to describe variations in self-reports of mental processes. It can be used to obtain empirical data on the following types of variables: (a) frequency and patterning of daily activity, social interaction, and changes in location; (b) frequency, intensity, and patterning of psychological states, i.e., emotional, cognitive, and conative dimensions of experience; (c) frequency and patterning of thoughts, including quality and intensity of thought disturbance. The article reviews practical and methodological issues of the ESM and presents evidence for its short- and long-term reliability when used as an instrument for assessing the variables outlined above. It also presents evidence for validity by showing correlation between ESM measures on the one hand and physiological measures, one-time psychological tests, and behavioral indices on the other. A number of studies with normal and clinical populations that have used the ESM are reviewed to demonstrate the range of issues to which the technique can be usefully applied.

This study was supported by the U.S. Public Health Service, the National Institute of Mental Health, and the Spencer Foundation. *Journal of Nervous and Mental Disease*, 175(9), 526–536. Copyright © 1987 Wolters Kluwer Health—Republished with permission.

M. Csikszentmihalyi (✉)
Division of Behavioral & Organizational Science, Claremont Graduate University,
Claremont, CA, USA
e-mail: miska@cgu.edu

R. Larson
Department of Psychology, University of Illinois, Urbana, IL, USA

Sampling of Experience

In recent years a growing number of investigators have sought information on the daily events and experiences that make up people's lives. Pervin (1985) identified the "increasing use of beeper technology" as a research methodology in which signaling devices carried by respondents are used to elicit self-report data at randomized points in time. One of the earliest lines of investigation using pagers to stimulate self-reports began at the University of Chicago in 1975, under the name of "Experience Sampling Method" (Csikszentmihalyi et al. 1977). The general purpose of this methodology is to study the subjective experience of persons interacting in natural environments, as advocated by Lewin (1936) and Murray (1938), in a way that ensures ecological validity (Brunswick 1952). The need for this kind of approach arises from dissatisfaction with a large body of research demonstrating the inability of people to provide accurate retrospective information on their daily behavior and experience (Bernard et al. 1984; Mischel 1968; Yarmey 1979). Its goal is similar to the one Fiske (1971, p. 179) set out for psychology as a whole: "to measure ... the ways a person usually behaves, the regularities in perceptions, feelings and actions."

The objective of the research described in this paper is to sample experience systematically, hence the name Experience-Sampling Method (ESM). The present article describes ESM and reports on its reliability and validity, using findings from a number of studies. (Readers may also wish to refer to two earlier and more restricted reviews of the methodology i.e., Hormuth 1986; Larson and Csikszentmihalyi 1983.)

The development of experience sampling responds to a number of currents within psychology and the social sciences. Two methodological traditions are the most direct ancestors of the present approach; first, research in the allocation of time to everyday activities (time budgets); and second, research measuring psychological reactions to everyday activities and experiences. Time budget studies have typically assessed time investment in different activities by different categories of persons (Altschuller 1923; Bevens 1913; Robinson 1977; Sorokin and Berger 1939; Szalai 1972; Thorndike 1937; Zuzanek 1980). Another approach has been that of the Kansas School of Ecological Psychology, which observationally investigated behavioral settings and focused on time use in the socialization of children (Barker 1968; Barker and Wright 1955; Barker et al. 1961). These studies were later extended cross-culturally (Johnson 1973; Munroe and Munroe 1971a, b; Rogoff 1978).

A second tradition of research focused on the impact of everyday life situations on psychological states, such as "psychopathology and coping" (Gurin et al. 1960); and "well-being" (Bradburn 1969). These studies provided data on global psychological states in representative populations. Researchers in the field of social psychology, on the other hand, performed experimental studies in which respondents were asked to imagine themselves in various life situations and then to report their psychological reactions (Endler and Magnusson 1976; Magnusson and Endler 1977).

Although these methods have enriched our understanding of individual lives, they have several short-comings: Imagery evoked in laboratory studies is not necessarily typical of experience encountered in real-life situations. In quality-of-life studies, only global assessments of extremely complex phenomena are presented. The data are gathered in retrospect, outside of the context of the situation, thus permitting distortions and rationalizations to become important. Time budget studies have been obtained from observer data or from diaries that do not provide direct access to the subjects' internal states. Nor in these studies is it clear what the link is between behavior and psychological states or between time use and experience. The ESM, which assesses subjects in real time and context, attempts to overcome some of these shortcomings.

Conceptually, the ESM "exposes" regularities in the stream of consciousness, such as states of heightened happiness or self-awareness, extreme concentration experienced at work, and symptoms of illness. The research aim is to relate these regularities to characteristics of the person (e.g., age, aptitude, physiological arousal, medical diagnosis), of the situation (e.g., the challenges of a job, the content of a TV show), or of the interaction between person and situation (e.g., the dynamics of a conversation with a friend, the circumstances that lead to a specific event). The objective is to identify and analyze how patterns in people's subjective experience relate to the wider conditions of their lives. The purpose of using this method is to be as "objective" about subjective phenomena as possible without compromising the essential personal meaning of the experience.

Methods

Instruments

Signaling device. To obtain representative self-reports of experiential states, the ESM relies on an electronic instrument that emits stimulus signals according to a random schedule. Different sound or vibration signaling devices have been used by different research groups: pagers such as those used to page doctors in the hospital (e.g., Csikszentmihalyi and Graef 1980; Csikszentmihalyi and Figurski 1982; Larson and Csikszentmihalyi 1983), programmed pocket calculators (e.g., Massimini, this issue, p. 545), programmed wrist watch terminals (e.g., deVries 1983; Brandstatter 1983), and other devices that signal at random intervals, such as those used in thought-sampling research (Hurlburt 1979; Klinger et al. 1980) and occupational studies (Divilbliss and Self 1978; Spencer 1971). Some studies have had the subjects themselves set watches according to predetermined timetables (Brandstatter 1983; Diener et al. 1984) but most program the signal devices for the respondents. The signal devices can also be programmed simultaneously and this provides special opportunities for analysis of the interdependence of experience in studies of couples, families, or friendship groups. Pawlik and Buse (1982) have

pioneered data-recording units that, in addition to signaling, are also able to record coded responses directly on tape. At the University of Heidelberg, Hormuth (1985, 1986) built a device that can be programmed from a portable Epson computer to signal 128 times over a period of 8 days. Different devices have different advantages and disadvantages. The choice also depends on the groups of subjects under investigation (e.g., the vibration option is more practical for research with adolescents, who spend part of their day at school; the sound of wrist watch terminals can be too soft to page the elderly, etc.). The function of the signaling device is always to cue respondents to report their activities, thoughts, and inner states at unexpected random times in natural settings. Any technique providing this function is adequate.

Experience-Sampling Form (ESF)

At the signal, the respondent writes down information about his or her momentary situation and psychological state on the ESF self-report questionnaire (see Appendix). This record becomes the basic datum of the ESM. The ESF is typically designed so that it will take no more than 2 min to complete. Respondents usually carry a full packet of the forms in a booklet.

Items contained in the form vary depending on the investigator's goals. In the authors' research, the objective has been to seek comprehensive coverage of the respondent's external and internal situation at the time of the signal. Hence, the ESF includes open questions about location, social context, primary and secondary activity, content of thought, time at which the ESF is filled out, and a number of Likert scales measuring several dimensions of the respondent's perceived situation including affect (happy, cheerful, sociable, and friendly), activation (alert, active, strong, excited), cognitive efficiency (concentration, ease of concentration, self-consciousness, clear mood), and motivation (wish to do the activity, control, feeling involved). The sample ESF used in a study of adolescents is reproduced in the Appendix.

Variants of this form have been used with other samples. Some researchers have focused exclusively on thought content (Hurlburt 1979, 1980; Klinger et al. 1980) or on emotional experience (Diener et al. 1984). Other investigators have developed specialized item sets on self-image and self-awareness (Franzoi and Brewer 1984; Savin-Williams and Demo 1983), adjustments to changes in residence (Hormuth 1986), intervening daily events (Greene 1985), binge eating (Johnson and Larson 1982), alcohol and drug consumption (Filstead et al. 1985), thought disorders (deVries, 1983, 1984; deVries et al. 1986) and the special problems of physically disabled children and adolescents,¹ among others.

¹ Marta Wenger and colleagues of the Collaborative Study of Children with Special Needs at the Children's Hospital Medical Center in Boston have used ESM with several samples of physically disabled children.

Procedures

Scheduling of signals. In the majority of ESM studies respondents received seven to 10 signals per day for 7 consecutive days. Usually the scheduling of signals is a variant of what Cochran (1953) describes as “systematic sampling,” a procedure that typically obtains more precise estimates of population characteristics than a purely random sample. Within the 15–18 target hours each day, one signaling time is selected for every block of 90–120 min, using a table of random numbers, with the provision that no signals should occur within 15 min of each other. The length of the reporting period and the timing of reports again depend on the scope and aims of the investigation. Filstead and colleagues (1985) asked alcoholics coming out of treatment to respond to a schedule of four signals per day for 3 months. In a study of the menstrual cycle, LeFevre et al. (1985) signaled married couples three times a day for 1 month. At the other extreme, studies of thought content have had signals with a mean delay of only 30 min for 3 days (Hurlburt 1979). A concentration of signals during crucial points of the day or in different situations has also been used (deVries 1983). However, the more that researchers have demanded of respondents, the fewer people have been willing to take part in the research. Generally, when frequency has increased, shortening the duration of the total sampling period gives augmented compliance.

Before beginning experience sampling, respondents receive instructions on the use of the signaling device and are instructed on how to fill out the ESF. They are told that they should fill out the form as soon as possible after each signal; typical situations in which this might be difficult (driving a car, playing sports) are discussed. Subjects are asked to keep the beeper turned on and to respond to all signals received, unless they go to bed or they “really need privacy.” The general purposes of the research are described, and the necessity for the respondents to report their life as it actually is—with all its joys and problems—is stressed.

Respondents should be given a chance to fill out a practice ESF. Subsequently, each subject is provided with a bound booklet of about 40–60 ESFs and a telephone number where someone can be reached in case any question or complication arises. When the week is completed, each subject is debriefed and the ESF booklets are collected. Additional interview or questionnaire data may be obtained before or after the test week.

Coding

Most of the data consists of self-scoring rating scales. The open-ended items may be coded in different ways, depending on the goals of the study. In some studies (Buse and Pawlik 1984; Hormuth 1984), activity and thought categories are provided on each ESF for the respondent to check, thus eliminating the need for the

researchers to code open-ended responses. In the Chicago research each variable was coded in fine detail, although codes in most analyses are aggregated in larger categories.

Activity codes. For the adult sample the answers to the item “What was the main thing you were doing?” were initially coded in one of 154 activity categories (e.g., “operating a typewriter,” “playing with a child,” “planning a meal”). In most analyses, however, activities were combined in 16 larger groups (e.g., “working at work,” “other at work,” “transportation”). Sometimes only three global activity areas are contrasted; work, maintenance, and leisure. Agreement between two coders at the level of the 154 categories was 88 %; at the level of the 16 variable groupings it was 96 %.

Of course, subjects may vary in how they report an activity. The same activity can be described by one person as “I was typing at my desk,” by another as “I was helping out my boss,” and by still another as “I was waiting for 5 o’clock to come so I could leave the office.” Action identification theorists (Harré and Secord 1972; Wegner and Vallacher 1977) find important differences associated with different ways of segmenting and labeling behavior. In our studies activity reports were collapsed into functional categories (e.g., work or leisure) without concern for structural characteristics such as the “level of identity” of the response, although such coding could also be attempted.

Thoughts. We used the same codes as for activities to code the content of thought, with only a few additions. The categories were then aggregated into fewer functionally equivalent groups (e.g., thoughts about work, family, or self). Other researchers have developed schemes for coding thoughts that are specifically designed to study the stream of consciousness and psychopathology (deVries et al. 1986; Hurlburt, 1979; Klinger et al. 1980).

Data Structure

Data are typically stored in two major computer files: a “beeper” and a “person” file. The first contains the data from each separate ESF in its entirety. The second contains percentages, means, standard deviations, and other aggregate scores for different variables, compiled by respondent, as well as information from interviews and questionnaires. The person file may also contain various aggregate scores for each individual, such as the percentage of time in various activities or mean self-rating in different social contexts or locations. For each new analysis the relevant data are aggregated and then transferred from the beeper to the person file. The addition of a “day level” file has been added to the beeper and person files by deVries et al. (1986) to study psychopathology. The day level file contains an overall assessment of the daily routine and so-called “Zeitgebers” (time waking up, going to sleep, meals).

Because the ESM obtains random samples of daily experience, the data base in these files is a representative record that can be accessed over and over again, to

test any number of hypotheses formulated at the time of collection or 20 years later. One can think of the data base as a permanent laboratory in which an almost unlimited number of relationships may be tested (e.g., Graef et al. 1983). To the extent that new records are continuously being added and the number of observations in each cell increases, ever more refined questions can be asked of the data.

Compliance

Volunteer rates. The ESM can be used with a variety of populations, provided that they can write and that a viable research alliance can be established. To date the youngest respondents have been 10 years old, and the oldest 85. ESM research has been carried out on people with schizophrenia (Delespaul and deVries, this issue, p. 537; deVries et al. 1986), anxiety disorders (Dijkman and deVries (1987), this issue, p. 550), multiple personality disorders (Loewenstein et al. 1987), bulimia (Johnson and Larson 1982), alcoholism (Filstead et al. 1985) and paraplegia.² In our studies we have been able to include adults who spoke little English and had only a few years of grammar school education, but the rate of volunteering from unskilled blue collar workers was extremely low (12 % of the target sample), and of those who volunteered only half completed 30 ESFs or more. It was clear that for them the task was unusual and difficult to handle. At the other extreme, among clerical workers and technicians, 75 % of the eligible population volunteered and completed the study. Among randomly selected fifth and eighth graders in a current study the rate was 91 %. Among high school students and adults, females have been more wilting participants than males.³

Response frequency. Respondents varied in their rate of compliance with the method. Blue-collar workers have responded to 73 % of the signals on the average, clerical and managerial workers to 85 and 92 %, respectively, for an overall average of 80 %. High school students had a median response rate of 70 %. Pawlik and Buse (1982) and deVries et al. (1986) reported an 86 % completion rate, and Hormuth's study (1985) had a median completion rate of 82 %.

Missing signals occur for a variety of reasons. From the debriefing interviews these appear to be due primarily to technical problems such as beeper malfunction or reception difficulties. The second most frequent source of attrition was forgetting the pager or the ESFs at home. A third source was related to the nature of the activity at the time the signal was transmitted, such as being in church, in the swimming pool, or in bed.

The frequency of delay in response to the pager is typically quite small. In our study of 75 adolescents (Csikszentmihalyi and Larson, 1984), 64 % responded

² See footnote 1

³ Larson (1979) The significance of time alone in adolescents' lives. Doctoral Dissertation, The University of Chicago.

immediately when they were signaled and 87 % responded within 10 min. In Hormuth's (1985) study of 101 adult Germans ($N = 5145$ observations), 50 % of the signals were responded to immediately, 80 % within 5 min of the signal, and 90 % within 18 min. Similar results were obtained in other studies. Delays are usually due to being engaged in activities that cannot be interrupted, for example, taking a test in school, driving a car, or talking to a customer. In most studies, ESFs filled out more than 20 min after the signal were discarded.

Experimental effects. Analyses of debriefing responses suggest that the intrusiveness of the method is not felt to be excessive. Among U.S. adults, 32 % said that the beeper was getting disruptive or annoying by the end of the week; 22 % of the German adults complained that it disrupted daily routine (Hormuth 1985). Ninety percent of the Americans and 80 % of the Germans felt that the reports captured their week well. When asked whether they would participate again in such a study, 75 % of Hormuth's subjects answered yes.

Reliability of ESM Measures

Sampling Accuracy

Because the ESM obtains a systematic random sample of daily life, it provides a measure of how people spend their time during a typical week. This measure is both an important measure in its own right, indicating the frequency of different activities, and a means for determining sampling accuracy of the ESM reports. A comparison with diary records from time budget studies (Robinson 1977; Szalai 1972) shows that the frequency of activities measured with the ESM correlates well with the rank of time budget activity frequencies ($r = 0.93$). Although diaries and the ESM provide very similar measures of activity frequency, a few discrepancies between the two are worth mentioning because they seem to be exceptions that prove the accuracy of the ESM. Respondents report to be 'idling' over 5 % of the time, while this category does not even appear in the time budgets. Idling was coded when the respondents reported staring out of the window or standing about without doing or thinking anything. Apparently this type of behavior is drastically underrepresented in retrospective reports. Overall, however, the two measures produce almost identical values of time allocation for different daily activities. The duration and sequence of activities, however, are more clearly calculated with diary approaches.

Stability of Activity Estimates

How stable are these measures over time? To answer this question, the week has been divided into two halves, and the frequency of activities for the group was computed within each period. By and large, activities are reported with the same frequency in the two halves of the reporting period. For a sample of 107 adults the difference was not significant ($\chi^2[15] = 8.1$), in spite of the enormous number. For a sample of 75 students, on the other hand, the differences were significant ($\chi^2[13] = 33.4$; $p < 0.01$). The primary difference was the greater percentage of time at work during the second half of the week, and it is attributable to the fact that more people, especially in the adolescent sample, started the experiment toward the end of the week (Thursday and Friday), so that the first half included more weekend self-reports.

Stability of Psychological States

Another issue is whether self-reports of affect, activation, motivation, and cognitive efficiency are stable over the testing period, or whether the pattern of responses changes over time as a result of the measurement procedure. To answer this question the mean scores for each individual from the first half of the week were compared with those from the second half.

None of the averaged individual mean response variables showed a significant change from the week's beginning to its end. It should be added that for both adolescents ($N = 75$) and adults ($N = 107$) the most extreme change was on the variable free vs. constrained: adolescents (difference = 0.26, $p = 0.002$) and adults (difference = 0.16, $p = 0.006$) felt more constrained in the second half of the week, probably in reaction to the method itself.

The variance in responses around an individual's mean diminished from the first to the second half of the week. For adults, the decrease in variance was always significant, whereas for adolescents the variance in affect and in strength was not.

Does this mean that in the course of the reporting period people become stereotyped in their responses and fail to differentiate between situations? To answer that question, we compared the first and the second halves of the week to determine whether the amount of variance accounted for by activities diminished with time, by calculating the variance attributable to the person's own response pattern and that to his or her activities (see Table 3.1). To save space, only one variable for each of the four dimensions is presented here. Personal effects were in general more powerful, accounting for between one fifth and one third of the variance, whereas activity effects explained only about 5%. In the second half of the week all the personal effects increased, indicating that with time individual responses become more predictable. However, activity effects did not show a comparable

Table 1 Proportion of variance in psychological states accounted for by persons and activities in the first and second half of the week of ESM self-reporting

Item	Adolescents				Adults			
	Persons (<i>N</i> = 75)		Activities (<i>N</i> = 14)		Persons (<i>N</i> = 107)		Activities (<i>N</i> = 16)	
	1st half	2nd half	1st half	2nd half	1st half	2nd half	1st half	2nd half
Affect: happy	0.18	0.23	0.05	0.05	0.25	0.34	0.04	0.04
Arousal: active	0.15	0.20	0.10	0.10	0.16	0.29	0.11	0.07
Cognitive efficiency: concentration	0.18	0.28	0.07	0.07	0.32	0.41	0.07	0.06
Motivation: wish to be doing	0.16	0.24	0.13	0.13	0.20	0.25	0.05	0.09

decline over time for either adolescents or adults. Thus the reduction in variance does not imply a lessened sensitivity to environmental effects, but a more precise self-anchoring on the response scales.

Individual Consistency Over the Week

Contrary to a one-time measure, the ESM is not based on the assumption that people are going to be entirely consistent in their responses. Person A might be happier than person B on Monday, but on Thursday B could be happier than A, depending on intervening experiences. Because the technique was devised to measure the effects of life situations on psychological states, perfect reliability would in fact defeat its purpose. In general, however, it was expected that relative differences between respondents would tend to persist over time. To check on individual response consistency, each subject's mean and standard deviation in the first half of the week were correlated with the means and standard deviations in the second half.

All correlations were significant, for the means as well as for the standard deviations, indicating that levels of both response and variability are fairly stable individual characteristics. The median correlation coefficient on the eight variables was 0.60 for the adolescents and 0.74 for the adults, suggesting that better anchored psychological states develop with age.

One investigator⁴ added five items intended to measure self-esteem to each ESM and collected 2287 observations from 49 mothers of small children. The Cronbach alpha for the set of items was 0.94, and coefficient of the correlation between the mean self-esteem scores in the first half of the week and the second half was 0.86 ($p = 0.0001$).

Pawlik and Buse (1982), studying 135 high school students in Hamburg, also correlated the frequency of responses between the 3651 protocols in the first half

⁴ Wells A (1985) Variations in self-esteem in the daily life of mothers, Doctoral Dissertation, The University of Chicago.

Table 2 Changes in immediate experience from time 2° table is based on the average of the approximately 31 ESM self-reports for each of 27 adolescents at two points in time. From Freeman, Larson, and Csikszentmihalyi, 1966. Reprinted with permission

Affective dimensions	Correlations between times 1 and 2	Mean of individual means		<i>t</i>
		Time 1	Time 2	
Affect				
Happy (versus sad)	0.77***	4.83	4.81	-0.20
Friendly (versus angry)	0.67***	4.88	4.77	-1.16
Cheerful (versus irritable)	0.72***	4.72	4.66	-0.64
Sociable (versus lonely)	0.50***	4.82	4.66	-1.45
Total affect scale	0.77***	19.29	18.90	-1.38
Activation				
Alert (versus drowsy)	0.62***	4.65	4.69	-0.36
Strong (versus weak)	0.60***	4.36	4.28	-0.75
Active (versus passive)	0.44*	4.22	4.14	-0.74
Involved (versus detached)	0.52**	4.56	4.52	-0.31
Excited (versus bored)	0.38*	3.90	4.00	-0.90
Total activation scale	0.62***	21.81	21.63	-0.37

**N* = 27
 **P* < 0.05
 ***p* < 0.01
 ****p* < 0.001

of the week and the 3729 protocols in the second half. Their subjects reported only the presence or absence of various subjective states, instead of using Likert scales. The correlation coefficients were 0.57 for locations, 0.76 for moods, and 0.80 for motives, quite similar to the ones reported above despite the difference in methods.

Individual Consistency Over Two Years

Test-retest data are available for 28 adolescents (Freeman et al. 1980) who took part in ESM for a week first in their freshman or sophomore years in high school, and again, 2 years later, in their junior or senior years. The stability in their responses ranged from *r* = 0.45 (*p* = 0.05) for the variables active, to *r* = 0.77 (*p* = 0.001) for happy (Table 3.2).

Internal Consistency

Because individual ESM items are administered many times, it is less important for most analyses to have multiple items measuring a single construct as in a traditional paper and pencil measure. Indeed, it would be sketching respondents' patience to ask

them to fill out a 20-item instrument 50 times in 1 week. For data reduction, however, researchers have factor-analyzed the ESM mood items and constructed scales by summing small numbers of items. Alpha levels for scales of affect (0.57) and arousal (0.48) are acceptable for measures computed from only four items.

Validity of ESM Measures

The reliability data on the convergence of diary and ESM measures also provide information on the validity of the ESM for measuring time usage. Here we will focus on its use for assessing internal states.

In general, the data suggest (a) that ESM reports of psychological states covary in expected ways with the values for physical conditions and with situational factors such as activity, location, and social context; (b) that measures of individual differences based on the ESM correlate with independent measures of similar constructs; and (c) that the ESM differentiates between groups *expected* to be different, e.g., patient and nonpatient groups or gifted and average mathematics students.

Situational Validity

Eight subjects wearing heart rate and activity monitors as well as ESM pagers were asked to supplement each self-report with a rating on a 10-point scale on “How physically active have you been in the past 3 min?”.⁵ This self-reported item predicted heart rate as well as readings from the ankle and wrist activity monitors: the correlation of self-ratings with heart rate was $r = 0.41$ ($p < 0.0001$), and with monitor readings, $r = 0.36$ ($p < 0.0001$). Substantial individual differences in this relationship, ranging from $r = 0.61$ to $r = 0.16$, were significantly correlated with other personality variables that suggested important individual differences in how physically aware the subjects were.

In addition, the physical activity self-ratings differentiated very highly between four body positions. When respondents were lying down, the mean z-score for self-reported physical activity was -1.47 , when sitting, it was 0.34 , when standing, it was 0.43 , and when walking, it was 1.03 (analysis of variance, $F[3,268] = 41.7$, $p < 0.0001$). Heart rate also varied in relation to body position ($F[3,268] = 6.95$, $p < 0.0001$).

Another example of an expected relationship between activity and experience is the association between what people do and their level of motivation. When

⁵ Hoover MD (1983) Individual differences in the relation of heart rate to self-reports. Doctoral Dissertation, The University of Chicago.

activity categories are arranged on an obligatory-discretionary axis, the productive ones (work) are rated as obligatory 80 % of the time. Maintenance activities are obligatory less often, from cleaning house (54 %) to shopping (37 %), and leisure is rarely seen as obligatory, from socializing (15 %) to watching TV (3 %).

In the study of working women with small children mentioned above, it was found that the self-esteem of mothers was much higher when they were working or involved in leisure than when they were taking care of the house or of their children (analysis of variance, $F[1,46] = 13.77, p < 0.0005$). Additional expected relationships between activity and experience can be found over a wide range from states related to drug and alcohol use or binge eating (Johnson and Larson 1982; Larson et al. 1984) to emotions when alone on Friday or Saturday night (Larson et al. 1982), or to stress experienced in anxiety-provoking situations (Dijkman and deVries 1987, p. 550; Margraf et al. 1987). this issue, p. 558).

Individual Characteristics and Variation in Experience

In addition to using ESM data to assess regularities in how people experience different daily situations, they may also be useful when the *person* is the unit of analysis. For example, a number of researchers have found correlations between participants' responses on the ESM and their scores on other psychometric instruments.

Giannino et al. (1979) entered ESM scores of a workers' sample into a regression analysis with 27 predictor items. The variable that accounted for the largest proportion of variance in the affect dimension (13 %, $p < 0.0001$) was the alienation-from-self subscale of Maddi's Alienation Test (Maddi et al. 1979). In other words, the best predictor of positive affect on the ESM was the absence of alienation from the self. Moreover, workers who scored high on a work satisfaction test scored much higher on the item "involved" when they were alone and actually working on their jobs than did subjects who scored low on work satisfaction. Satisfied workers had higher scores on concentration, skills, alertness, and motivation (in each case, $p < 0.0001$).

The strength of the need for intimacy was measured by McAdams and Constantian (1983) using projective techniques and comparing them with ESM measures. People with a high need for intimacy reported more thoughts about people and relationships ($r = 0.52, p < 0.001$), more conversation with others ($r = 0.40, p < 0.01$), higher affect when with others ($p < 0.001$), and a lower rate of wishing to be alone when with others ($r = -0.32, p < 0.05$).

Hamilton et al. (1984) developed a questionnaire scale for measuring the amount of enjoyment subjects report in their daily lives. The amount of intrinsic enjoyment was related to several ESM variables such as motivation, wish to be doing the activity ($p < 0.001$), concentration, ease of concentration, control, and activity/potency ($p < 0.05$ in each case).

One-time assessments on the Rosenberg Self-Esteem scale (RSES) were compared with the average of a repeated self-esteem scale (four items from the RSES included in the ESF) and with the average of five ESM items related to self-esteem. The one-time RSES correlated with the repeated RSES items $r = 0.62$ ($N = 49, p < 0.0001$) and with the ESM self-esteem items $r = 0.42$ ($p < 0.002$). This latter correlation varied considerably depending on the social context: When alone, subjects' responses on the ESM self-esteem items correlated with the one-time RSES at only $r = .26$ (NS), when only children were present at $r = 0.36$ ($p < 0.05$), and when adults were also present at $r = 0.50$ ($p < 0.001$). In other words, self-esteem as measured by a one-time traditional test corresponds to the self-esteem people report when they are in public.

Clinically, Loewenstein et al. (1987) reported the use of ESM with a woman with a multiple personality disorder. They found that the alternates displayed quantitative differences on the affect and motivational scales comparable to those observed between separate individuals,

Differences in Experience Between Groups

The ESM has also differentiated well between the responses of groups with distinctive behavioral patterns and groups with psychopathology. For example, deVries and associates (deVries 1983; deVries et al. 1986) coded the thoughts reported on the ESFs of Dutch schizophrenic and nonschizophrenic mental patients; the schizophrenics generally suffered more severe thought disorders, whereas the other patients suffered mainly from affective disorders. He found that nonschizophrenic patients evidenced congruence between thoughts and actions 75 % of the time, whereas the schizophrenics did the same only half of the time ($t = 2.82, p < 0.005$). The thoughts reported by the schizophrenics were also coded as disordered much more often than were those of the control group ($t = 9.13, p < 0.0005$). On the other hand, the schizophrenics reported a more positive average affect ($t = 1.78, p < 0.05$).

In their investigation of women with eating disorders, Johnson and Larson (1982) found that bulimic women were involved in food-related behavior or thought on the average of 38 % of their waking time, as opposed to the 14 % average for a comparison group of women. They also found that the overall level of positive affect was lower for bulimics than for normal women (happy, $t = 4.66, p < 0.001$; cheerful, $t = 4.14, p < 0.001$; sociable, $t = 4.12, p < 0.001$).

Overview

Since its introduction 10 years ago, the ESM has proved to be a useful tool for psychological research. Its main contribution has been to make the variations of daily experience, long outside the domain of objectivity, available for analysis,

replication, and falsifiability, thus opening up a whole range of phenomena to systematic observation.

The most heuristic usefulness of the ESM lies in its *description of the patterns of an individual's daily experience*. Because the method yields repeated measurements of a person's activities, feelings, thoughts, motivations, and medical symptoms over time, questions such as the following can be answered: How much of the person's variation in happiness (or any other state) is related to what the person does; to the company he or she keeps; to the time of day; to intervening events? By the same token, the ESM can reveal subjective effects of major life changes that might otherwise be hidden from consciousness by distortion or inaccurate recollection. The comparison of responses before and after a job or family change, a clinical intervention, or other changes in life situations can reveal what impact such transitions have on a person's daily life.

Adding up patterns within a person, it becomes possible to use ESM *to evaluate the common experience of situations*. For instance, is solitude, housework, or marijuana smoking experienced similarly by different individuals? It follows that the ESM can be used to compare the subjective experience of different events. Do men and women differ in their daily emotions? What states, feelings, attitudes differentiate talented achievers from talented nonachievers? How can we compare the experience of physically handicapped children with that of other children? Likewise, ESM data can reveal whether changes in life situations elicit consistent changes in experiences of people in general.

A final use of ESM is *to study the dynamics of emotions and other subjective states*. The study of consciousness has lagged behind other fields of psychology. We know little about the structure of emotions and less about how other dimensions of our psychological state (e.g., concentration, involvement, motivation) ebb and flow in daily experience. ESM data allow examination of the magnitude, duration, and sequences of states, as well as an investigation of correlations between the occurrences of different experiences. For example, one can examine whether concentration is typically associated with positive affect, how long it lasts, and what factors are related to its ending.

The major limitation of the ESM is the obvious one: its dependence on respondents' self-reports. This limitation becomes a concern in situations in which it is conceivable that a large segment of one's sample provided inaccurate or distorted data. For example, if an employer used the method to study his employees' productivity, the accuracy of self-reports' related to working would be suspect, as would the ESM results in an investigation of private, sensitive, or illegal activities.

When self-reports deal with the immediate, however, they have been found to be a very useful source of data (Ericsson and Simon 1980; Mischel 1981). In this paper we have presented ample evidence indicating that they typically provide a plausible representation of reality.

Appendix: Experience-Sampling Form

Date: _____ Time Beeped: _____ am/pm Time Filled Out __ am/pm

As you were beeped...

What were you thinking about? _____

Where were you? _____

What was the MAIN thing you were doing? _____

What other things were you doing? _____

WHY were you doing this particular activity?

I had to I wanted to do it I had nothing else to do

	Not at all	Some what	Quite	Very
How well were you concentrating?	0	1 2 3	4 5 6	7 8 9
Was it hard to concentrate?	0	1 2 3	4 5 6	7 8 9
How self-conscious were you?	0	1 2 3	4 5 6	7 8 9
Did you feel good about yourself?	0	1 2 3	4 5 6	7 8 9
Were you in control of the situation?	0	1 2 3	4 5 6	7 8 9
Were you living up to your own expectations?	0	1 2 3	4 5 6	7 8 9
Were you living up to expectations of others?	0	1 2 3	4 5 6	7 8 9

Describe your mood as you were beeped:

	Very	Quite	Some	Neither	Some	Quite	Very	
Alert	0	o	.	—	.	o	0	Drowsy
Happy	0	o	.	—	.	o	0	Sad
Irritable	0	o	.	—	.	o	0	Cheerful
Strong	0	o	.	—	.	o	0	Weak
Active	0	o	.	—	.	o	0	Passive
Lonely	0	o	.	—	.	o	0	Sociable
Ashamed	0	o	.	—	.	o	0	Proud
Involved	0	o	.	—	.	o	0	Detached
Excited	0	o	.	—	.	o	0	Bored
Closed	0	o	.	—	.	o	0	Open
Clear	0	o	.	—	.	o	0	Confused

(continued)

(continued)

Tense	0	o	.	—	.	o	0	Relaxed
Competitive	0	o	.	—	.	o	0	Cooperative

Did you feel any physical discomfort as you were beeped:

Overall pain or discomfort	none			slight			bothersome			severe
	0	1	2	3	4	5	6	7	8	9

Please specify: _____

Who were you with?

<input type="checkbox"/> Alone	<input type="checkbox"/> Friend(s)	How many?
<input type="checkbox"/> Mother		_____
	Female <input type="checkbox"/>	Male <input type="checkbox"/>
<input type="checkbox"/> Father	<input type="checkbox"/> Strangers	
<input type="checkbox"/> Sister(s) or brother(s)	<input type="checkbox"/>	
	Other _____	

Indicate how you felt about your activity:

	low		high
Challenges of the activity	0	1 2 3 4 5 6 7 8 9	
Your skills in the activity	0	1 2 3 4 5 6 7 8 9	
	not at all		very much
Was this activity important to you?	0	1 2 3 4 5 6 7 8 9	
Was this activity important to others?	0	1 2 3 4 6 6 7 8 9	
Were you succeeding at what you were doing?	0	1 2 3 4 5 6 7 8 9	
Do you wish you had been doing something else?	0	1 2 3 4 5 6 7 8 9	
Were you satisfied with how you were doing?	0	1 2 3 4 5 6 7 8 9	
How important was this activity in relation to your overall goals	0	1 2 3 4 5 6 7 8 9	
<i>If you had a choice...</i>			
Who would you be with?	_____		
What would you be doing?	_____		

Since you were last beeped has anything happened or have you done anything which could have affected the way you feel?

Nasty cracks, comments, etc. *****

References

- Altschuller, M. I. (1923). *Byudzhet vremeni (Time Budgets)*. Perm: USSR.
- Barker, R. G. (1968). *Ecological psychology*. Stanford: Stanford University Press.
- Barker, R. G., & Wright, H. F. (1955). *Midwest and its children*. New York: Row, Peterson, Evanston.
- Barker, R. G., Wright, H. F., Barker, L. S., & Schoggen, M. F. (1961). *Specimen records of American and English children*. Lawrence: University of Kansas Press.
- Bernard, H. R., Killworth, P., Kronenfeld, D., & Sailer, L. (1984). On the validity of retrospective data: The problem of informant accuracy. *Annual Review of Anthropology*, *13*, 495–517.
- Bevans, G. E. (1913). *How workingmen spend their spare time*. New York: Columbia University Press.
- Bradburn, N. (1969). *The structure of psychological well-being*. Chicago: Aldine.
- Brandstatter, H. (1983). Emotional responses to other persons in everyday life situations. *Journal of Personality and Social Psychology*, *46*, 871–883.
- Brunawick, E. (Ed.). (1952). *The conceptual framework of psychology. International encyclopedia of unified science* (Vol. 1(10)). Chicago: University of Chicago Press.
- Buse, L., & Pawlik, K. (1984). Inter-Setting-Korrelationen und Setting-Persönlichkeits-Wechselwirkungen: Ergebnisse einer Felduntersuchung zur Konsistenz von Verhalten und Erleben. *Zeitschrift für Sozialpsychologie*, *15*, 44–59.
- Csikszentmihalyi, M., & Figurski, T. J. (1982). Self-awareness and aversive experience in everyday life. *Journal of Personality*, *50*, 14–26.
- Csikszentmihalyi, M., & Graef, R. (1980). The experience of freedom in everyday life. *American Journal of Community Psychology*, *18*, 401–414.
- Csikszentmihalyi, M., & Larson, R. (1984). *Being adolescent: Conflict and growth in the teenage years*. New York: Basic.
- Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, *6*, 281–294.
- deVries, M. (1983). *Temporal patterning of psychiatric symptoms*. Vienna: World Psychiatric Association Congress.
- deVries, M. (1984). Temperament and infant mortality. *American Journal of Psychiatry*, *141*, 1189–1194.
- deVries, M., Delespaul, P., Dijkman, C., & Theunissen, J. (1986). *Temporal and situational aspects of severe mental disorders*. Milan: Angeli: L'Esperienza quotidiana.
- Diener, E., Larson, R. J., & Emmons, R. A. (1984). Person X situation interactions: Choices of situations and congruence response models. *Journal of Personality and Social Psychology*, *47*, 592–680.
- Dijkman, C. I. M., & deVries, M. W. (1987). The social ecology of anxiety: Theoretical and quantitative perspectives. *Journal of Nervous and Mental Disease*, *175*, 550–557.
- Divilbiss, J., & Self, P. (1978). Work analysis by random sampling. *Bulletin of the Medical Library Association*, *66*, 19–32.
- Endler, N. S., & Magnusson, D. (Eds.). (1976). *Interactional psychology and personality*. New York: Wiley.
- Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review*, *87*, 216–251.
- Filstead, W., Reich, W., Parrella, D., & Rossi, J. (1985). *Using electronic pagers to monitor the process of recovery in alcoholics and drug abusers*. Paper presented at the 34th international congress on alcohol, drug abuse and tobacco, Calgary, Alberta, Canada.
- Fiske, D. (1971). *Measuring the concept of personality*. Chicago: Aldine.
- Franzoi, S. L., & Brewer, L. C. (1984). The experience of self-awareness and its relation to level of self-consciousness: An experiential sampling study. *Journal of Research in Personality*, *18*, 522–540.
- Freeman, M., Larson, R., & Csikszentmihalyi, M. (1980). Immediate experience and its recollection. *Merrill Palmer Quarterly*, *32*, 167–185.

- Giannino, S., Graef, R., & Csikszentmihalyi, M. (1979) *Well-being and the perceived balance between opportunities and capabilities*. Paper presented at the 87th convention of the American Psychiatric Association, New York, New York.
- Graef, R., Csikszentmihalyi, M., & Gianinno, S. (1983). Measuring intrinsic motivation in everyday life. *Leisure Studies*, 2, 155–168.
- Greene, A. (1985). *Self-concept and life transitions in early adolescence*. Paper presented at the biannual meeting of the society for research on child development, Toronto, Ontario, Canada.
- Gurin, G., Veroff, J., & Field, S. (1960). *Americans view their mental health*. New York: Basic.
- Hamilton, J. A., Hater, R. J., & Buchsbaum, M. S. (1984). Intrinsic enjoyment and boredom coping scales: Validation with personality, evoked potential and attention measures. *Personality and Individual Differences*, 5, 183–193.
- Harre, R., & Secord, P. F. (1972). *The explanation of social behavior*. Oxford: Blackwell.
- Hormuth, S. E. (1984). Transitions in commitments to roles and self-concept change; recreation as a paradigm. In V. L. Allen & E. van de Vliert (Eds.), *Role transitions: Explorations and explanations*. New York: Plenum.
- Harmuth, S. E. (1985). *Methaden fur psychologische Forschung im Pels (Diskussionspapier Nr 43)*. Heidelberg: Psychologischen Institut der Universitat Heidelberg.
- Hormuth, S. E. (1986). The sampling of experiences in situ. *Journal of Personality*, 54, 262–293.
- Hurlburt, R. T. (1979). Random sampling of cognitions and behavior. *Journal of Research in Personality*, 13, 103–111.
- Hurlburt, R. T. (1980). Validation and correlation of thought sampling with retrospective measures. *Cognitive Therapy and Research*, 4, 235–238.
- Johnson, A. (1973). Time allocation in a machiguenga community. *Ethology*, 14, 301–310.
- Johnson, C., & Larson, R. (1982). Bulimia: An analysis of moods and behavior. *Psychosomatic Medicine*, 44, 341–351.
- Klinger, E., Barta, S., & Mexeimer, M. (1980). Motivational correlates of thought content frequency and commitment. *Journal of Personality and Social Psychology*, 39, 1222–1237.
- Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. Reis (Ed.), *New directions for naturalistic methods in the behavioral sciences*. San Francisco: Jossey-Bass.
- Larson, R., Csikszentmihalyi, M., & Freeman, M. (1984). Alcohol and marijuana use in adolescents' daily lives: A random sample of experiences. *The International Journal of the Addictions*, 19, 367–381.
- Larson, R., Csikszentmihalyi, M., & Graef, R. (1982). Time alone in daily eiperience: loneliness or renewal? In L. A. Peplau & D. Perlman (Eds.), *Loneliness; A sourcebook of research and theory*. New York: Wiley.
- LeFevre, J., Hendricks, C., Church, R., & McClintock, M. (1985). *Psycho-logical and social behavior in couples over menstrual cycle: 'On-the-spot' sampling from everyday life*. Paper presented at the 6th conference of the society for menstrual cycle research, Galveston.
- Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill.
- Loewenstein, R. J., Hamilton, J., Magna, S. et al. (1987). Experiential sampling in the study of multiple personality disorder. *American Journal of Psychiatry*, 144, 19–24.
- Maddi, S. R., Kobasa, S. C., & Hoover, M. (1979). An alienation test. *Journal of Humanistic Psychology*, 19, 73–76.
- Magnusson, D., & Endler, N. S. (1977). Interactional psychology: Present status and future prospects. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillside: Erlbaum.
- Margraf, J., Taylor, C. B., Ehlers, A. et al. (1987). Panic attacks in the natural environment. *Journal of Nervous and Mental Disease*, 175, 558–565.
- McAdams, D., & Constantian, C. A. (1983). Intimacy and affiliation motives in daily living: An experience sampling analysis. *Journal of Personality and Social Psychology*, 45, 851–861.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Mischel, W. (1981). A cognitive-social learning approach to assess-ment. In T. Merluzzi, C. Glass M. Genest (Eds.), *Cognitive assessment*. New York: Guilford.

- Munroe, R. H., & Munroe, R. L. (1971a). Household density and infant care in an East African society. *Journal of Social Psychology*, 83, 9–13.
- Munroe, R. H., & Munroe, R. L. (1971b). Effect of environmental experience on spatial ability in an East African society. *Journal of Social Psychology*, 83, 15–22.
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Pawlik, K., & Buse, L. (1982). Rechnergestützt verhaltensregistrierung im Feld: Beschreibung und erst psychometrische Überprüfung einer neuen Erhebungsmethode. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 3, 101–118.
- Pervin, L. A. (1985). Personality: Current controversies, issues, and directions. *Annual Review of Psychology*, 36, 83–114.
- Robinson, J. (1977). *How Americans use time*. New York: Praeger.
- Rogoff, B. (1978). Spot observation: An introduction and examination. *Quarterly Newsletter of the Institute for Comparative Human Development*, 2, 21–26.
- Savin-Williams, R. C., & Demo, D. H. (1983). Situational and transnational determinants of adolescents' self-feelings. *Journal of Personality and Social Psychology*, 44, 824–833.
- Sorokin, P., & Berger, C. (1939). *Time budgets and human behavior*. Cambridge: Harvard University Press.
- Spencer, C. (1971). Random time sampling with self-observation for library cost studies: Unit costs of interlibrary loans and photocopies at a regional medical library. *Journal of the American Society for Information Science*, 32, 153–160.
- Szalai, A. (1972). *The use of time*. Mouton: The Hague.
- Thorndike, E. L. (1937) How we spend our time and what we spend it for. *Scientific Monthly*, 44(5), 464–469.
- Wegner, D. M., & Vallacher, R. R. (1977). *Implicit psychology an introduction to social cognition*. New York: Oxford University Press.
- Yarmey, D. (1979). *The psychology of eyewitness testimony*. New York: Free Press.
- Zuzanek, J. (1980). *Work and leisure in the Soviet Union: A time budget analysis*. New York: Praeger.