



Eavesdropping on Health: A Naturalistic Observation Approach for Social Health Research

Matthias R. Mehl*

University of Arizona

Abstract

This article provides an overview of a novel ecological momentary assessment method called the electronically activated recorder or EAR. The EAR is a portable audio recorder that periodically records snippets of ambient sounds from participants' momentary environments. In tracking moment-to-moment ambient sounds, the EAR yields an acoustic log of a person's day as it naturally unfolds. As a naturalistic observation method, it provides an observer's account of daily life and is optimized for the assessment of audible aspects of participants' daily social environments and interactions. The article discusses the EAR method conceptually and methodologically and identifies three important ways in which it can enrich social health research. Specifically, it can help cross-validate research findings independent of self-reports, calibrate psychological measures against behavioral markers of real-world social functioning, and further our understanding of the role that people's mundane social interactions and language use play in coping and health.

Over the last two decades, psychology has experienced an impressive growth in research investigating the social underpinnings of health. Striking advances have been made in unraveling mind–body connections and identifying ways in which our social lives exert traceable effects on bodily systems and influence important health outcomes (Cohen, 2004; Kiecolt-Glaser, McGuire, Robles, & Glaser, 2002; Sapolsky, 2004). Yet, there is a growing concern that progress in the field is constrained by the fact that social health research lags behind other areas in psychology in implementing a multimethod approach on both sides of the 'social health equation'. Whereas the measurement of health – typically on the criterion side – has developed rapidly with progress in noninvasive physiological, endocrinological, and immunological assessment (e.g., Cacioppo, Tassinary, & Berntson, 2007), the measurement of health-relevant social constructs – typically on the predictor side – continues to be based predominantly on one method: the participants' self-report (Coyne & Gottlieb, 1996; Somerfield & McCrae, 1997).

When self-reports are not systematically complemented by other assessment methods, the scientific picture that emerges lacks dimensionality; it depicts what participants think of themselves and how they construe their social behaviors and relationships. Although these are undoubtedly important dimensions, it is hardly the complete phenomenon. Human memory is not a perfect storage system that precisely and permanently archives all perceptions and experiences, and the human cognitive apparatus is not a perfect information processor that provides an unbiased, objective account of our social worlds (Dunning, 2005; Schwarz & Sudman, 1993; Stone et al., 2000). Thus, social health research would benefit from ‘expanding the aperture of psychological assessment’ (Trull, 2007) to capture a broader spectrum of social information with a wider array of methods, and to develop a more complete, multidimensional understanding of the social architecture of health.

Since the early 1980s, momentary assessment methods – which in health research are commonly referred to as *ecological momentary assessment* (EMA; Stone & Shiffman, 1994) and in other contexts are also known as *experience sampling methods* (Hektner, Schmidt, & Csikszentmihalyi, 2006), *diary methods* (Bolger, Davis, & Rafaeli, 2003), or *ambulatory assessment methods* (Fahrenberg, Myrtek, Pawlik, & Perrez, 2007) – have been developed to bypass methodological concerns around global and retrospective self-reports. Over the last decade, EMA methods have effectively become the gold standard for assessing psychological aspects of daily life (Conner, Barrett, Tugade, & Tennen, 2007; Moskowitz & Young, 2006; Piasecki, Hufford, Solhan, & Trull, 2007; Reis & Wheeler, 1991; Suls & Martin, 1993).

With EMA, participants report their thoughts, feelings, and behaviors over time directly in their natural settings (Conner et al., 2007). In some cases, they provide (electronic or paper-and-pencil) end-of-day diaries; in other cases, they receive a personal digital assistant (PDA; i.e., Palm Pilot or Pocket PC) and are prompted repeatedly over a day to provide instant reports of what they are doing and how they are feeling. The impact that EMA has had on psychology is hard to overstate. Through EMA it has become possible to study psychosocial phenomena ‘in vivo’ and ‘in real time’. EMA methods have been widely used in social health research (Smyth & Stone, 2003; Stone, Shiffman, Atienza, & Nebeling, 2007; Tennen, Affleck, & Armeli, 2003).

Figure 1 shows a simple method matrix. It organizes types of methods used in social health research according to whether data collection is based on self-reports or behavioral observation and whether it takes place in the laboratory or in the participants’ natural environments (for sake of simplicity, it focuses only on psychological assessments and excludes physiological assessments). The upper left quadrant shows the generic global/retrospective self-report questionnaire (e.g., a standardized coping checklist). Within this matrix, EMA methods, that is momentary self-reports,

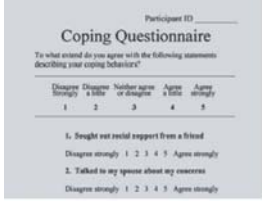



Data collection	In the laboratory	In the natural environment
Via self-report		
Via behavioral observation		

Figure 1 A method matrix for social health research.

Note: The picture illustrating behavioral observation in the laboratory was retrieved with permission from Dr. Giese-Davis' website at <http://coding-lab.giese-davis.com/process.php>.

are located in the upper right quadrant. The lower left quadrant contains laboratory-based observational methods such as the video-taping of couple interactions (e.g., Heyman, 2001) or patient support groups (e.g., Giese-Davis, Piemme, Dillon, & Twirbut, 2005).

As Figure 1 makes apparent, the lower right quadrant, behavioral observation in the natural environment, is not well represented. In fact, in psychology, extremely few studies have employed person-centered, naturalistic observation methods (e.g., Barker & Wright, 1951; Craik, 2000). Funder (2007) pointed out that, among other reasons, this is the case because it is not straightforward how one would go about collecting truly naturalistic behavioral data. Essentially, it seems, it would require a 'detective's report [that] would specify in exact detail everything the participant said and did, and with whom, in all of the contexts of the participant's life' (p. 41). Because this is ultimately impossible, EMA methods are generally considered the best available proxy of behavioral observation in the field (Conner et al., 2007).

From a multimethod perspective, however, momentary and global/retrospective self-reports share important method variance because both derive their data from participants' reports of their introspections and perceptions – in the method matrix both are located within the same row.

Therefore, some of the concerns raised for global/retrospective self-reports potentially also apply to momentary self-reports (Conner et al., 2007; Piasecki et al., 2007). Thus, to complete the social health researcher's toolkit, it would be desirable to fill the lower right quadrant by complementing momentary self-report data with momentary observational data.

A Method for the Naturalistic Observation of Daily Life: The Electronically Activated Recorder

Over the last years, we have developed the electronically activated recorder (EAR; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001), a method that unobtrusively samples (acoustic) observations of participants directly from the natural flow of their daily lives.

What is the EAR?

Technically, the EAR is a portable audio recorder that periodically records brief snippets of ambient sounds. Participants wear it attached to their belt or in a purse-like bag while going about their daily lives. In tracking moment-to-moment ambient sounds around the participants, it yields acoustic logs of their days as they naturally unfold. In sampling only a fraction of the time instead of recording continuously it makes large-scale naturalistic observation studies feasible.

The initial EAR system consisted of a microcassette recorder triggered by a microchip. It was programmed to record for 30 seconds every 12.5 minutes and produced roughly 70–80 recordings per day (Mehl et al., 2001). This first analog EAR was used from 1998 to 2000 in a psychometric study of people's daily social environments and interactions (Mehl & Pennebaker, 2003b). However, the system was not user friendly. It was susceptible to technical problems, bulky, and required participants to flip the tape after 24 hours of monitoring.

The second-generation EAR system remedied these problems. It consisted of a digital voice recorder triggered by a microchip. Going digital revolutionized EAR research. The new system was small and handy, operated with high precision and reliability, and allowed for several days of nonstop monitoring. It also automatically time stamped every recording. The digital EAR was used in a series of studies between 2001 and 2004 (e.g., Mehl & Pennebaker, 2003a; Mehl, 2006b; Mehl, Gosling, & Pennebaker, 2006).

In 2005, we replaced the digital EAR device with a third-generation system that runs on a PDA (Pocket PC). The PDA-based EAR system has several critical advantages: (i) it is software-based (i.e., requires no hardware modifications); (ii) it is available at a lower price; (iii) it runs on different commercial devices; (iv) it allows for freely programmed recording schedules (e.g., 30 seconds every 12.5 minutes, 5 minutes every hour) as well as blackout periods with no recordings (e.g., over night). Finally,



Figure 2 Pictures illustrating how the PDA-based EAR system is worn by a person. Note: The depicted individuals consented to the use of their pictures in publications.

because now both the traditional, self-report-based EMA methods and the EAR use the same electronic device, it is possible to merge both methodologies. Figure 2 illustrates how the PDA-based EAR system is worn by a person.

How does the EAR compare to traditional, self-report-based EMA methods?

As a psychological real-time data capture method, the EAR compares most directly to self-report-based EMA methods (Bolger et al., 2003; Conner et al., 2007; Stone et al., 2007). Table 1 summarizes important similarities and differences between the two methods.

Both methodologies are naturalistic in their approach and dynamic and process oriented in their focus (Fahrenberg et al., 2007; Laurenceau & Bolger, 2005). Whereas for traditional EMAs both paper-and-pencil and PDA-based versions are available (Conner et al., 2007), the EAR runs only electronically. The most important difference between the two methods lies in the fact that traditional, psychological EMA methods are based on momentary self-reports whereas the EAR is based on momentary behavioral observation. The two types of methods, hence, adopt different assessment perspectives: the one of the self or agent with the corresponding subjective, experiential account of an event versus the one of a bystander

Table 1 A comparison between traditional, self-report-based ecological momentary assessment methods and the EAR method

	Self-report-based EMA methods	EAR-method
Approach	Naturalistic	Naturalistic
Focus	Dynamic/process	Dynamic/process
Medium	Paper and pencil, electronic (PDA)	Electronic (PDA)
Method	Self-report	Behavioral observation
Perspective	Self (agent)	Other (observer)
Awareness of assessment	High	Low after habituation
Burden for participant	Practical (interruption of daily life)	Psychological (intrusion of privacy)
Data collection limited by	Response burden	Privacy considerations
<i>Optimized for assessment of</i>	<i>Subjective experiences and perceptions</i>	<i>Objective social environments and interactions</i>

or observer with the corresponding objective (i.e., 'person as object'), behavioral account.

Self-report-based EMAs by nature require participants' awareness of the assessment. In contrast, the EAR operates imperceptibly; participants never know when the recorder is on or off. Furthermore, after an initial period of device-induced self-awareness, participants' generally habituate to wearing the EAR and often report forgetting about it for extended periods of time.

The two methods further differ in the burden they place on participants. Self-report-based EMAs come with the practical burden of requiring participants to intermittently interrupt the flow of their daily lives to answer a series of questions. This practical burden creates an upper limit for the number of prompts that can be implemented per day, and the number of questions that can be asked upon each prompt (Bolger et al., 2003). The practical burden of the EAR consists of wearing the device; this burden is relatively low and independent of the specified sampling rate or the amount of information that is extracted from the sound files. However, the EAR places a different burden on participants: the psychological discomfort of knowing one is intermittently recorded (sometimes also referred to as evaluation apprehension). Therefore, EAR data collection is limited to sampling rates that result in privacy intrusions that are tolerable for participants. It is critical for EAR research to have policies in place that effectively protect participants' privacy and ensure the confidentiality of the data.

Taken together, these practical and conceptual differences between traditional self-report-based EMAs and the EAR suggest that the two methodologies are suited for slightly different assessments. In capturing the agent's 'insider' perspective, self-report-based EMAs are optimized for

the assessment of participants' subjective experiences and perceptions (e.g., thoughts, feelings, appraisals). In capturing the 'outsider' perspective of an unobtrusive observer, the EAR is optimized for assessing audible aspects of participants' objective social environments and interactions (e.g., social settings, communication behaviors, language use).

What psychological information can be derived from the EAR recordings?

Researchers can adopt a psychological rating or a behavior coding approach to deriving information from the sampled ambient sounds (Sillars, 1991). With the psychological rating approach, expert raters listen to the full set or selected segments of participants' sound files and judge the degree to which they indicate the presence of a construct of interest. For example, relationship experts could listen to participants' sound files and rate them on relationship satisfaction, social support, expressed emotions, or protective buffering (Kerig & Baucom, 2004). In this case, information is extracted at a molar, psychological level. Reliability can be determined from the consensus among the expert raters, and the construct validity of the ratings emerges from comparisons with established criterion measures (e.g., self- or spousal reports).

In our research, we have primarily worked with behavior codings. With this approach information is extracted at the molecular level of the raw behavior. Trained coders listen to all of a participant's EAR recordings and code each sound file using a standardized coding system. Over the last years, we have developed and refined the Social Environment Coding of Sound Inventory (SECSI; Mehl & Pennebaker, 2003b; Mehl et al., 2006) to capture basic aspects of participants' momentary social environments and interactions. The SECSI comprises four category clusters: (i) the person's current *location* (e.g., in apartment, outdoors, in transit), (ii) *activity* (e.g., listening to music, watching TV, eating), (iii) *interaction* (e.g., alone, talking, on the phone), and (iv) *mood* (e.g., laughing, crying, arguing). Conceptually, it captures information about how individuals (i) select social environments (e.g., displaying a preference for spending time in private versus public settings) and (ii) interact with their social environments (e.g., displaying a preference for dyadic versus group conversations).

Table 2 shows a sample excerpt of an EAR transcript and coding sheet. Each line in the table represents a sound file (with the recording date and time) and each column represents a SECSI category. '1' is coded if a category applied and '0' if it did not apply during a 30-second sound file. The raw codings can then be converted into relative frequencies by calculating the percentage of a person's waking EAR recordings in which a category applied (e.g., in which the person was talking or laughing). The coders routinely also transcribe all of the participants' utterances captured by the EAR. The verbatim transcripts can then be content analyzed (Mehl, 2006a). In addition to the SECSI, we have

Table 2 Sample excerpt of an EAR transcript and coding sheet taken from the EAR monitoring of a student participant

File no.	Date	Time	Transcript	Technical problem	EAR conversation	Interaction			Location				Mood		
						Other people present	Talking	On the phone	In apartment	Outdoors	In transit	In public place	Laughing	Crying	Sighing
023	11/7/2001	2:15 p.m.	Excuse me. That's alright. Bye.	0	0	0	1	1	0	0	0	1	1	0	0
024	11/7/2001	2:27 p.m.		1	0	0	0	0	0	0	0	0	0	0	0
025	11/7/2001	2:38 p.m.		0	0	0	0	0	0	0	1	0	0	0	0
026	11/7/2001	2:50 p.m.	They tell you? Hey, what exactly does that do? I mean what kind of medication do they give me? Test medication? But it will work, right? Okay.	0	0	0	1	1	1	0	0	0	0	0	0
027	11/7/2001	3:02 p.m.	Is mom there? Okay I have to talk to her. Hey. I talked to the tooth people just now, the tooth people. Anyway, I have to have proof of social security. So I need you to fax it to me or mail it to me. Uh. No.	0	0	0	1	1	1	0	0	0	0	0	1
028	11/7/2001	3:13 p.m.		0	0	0	0	0	1	0	0	0	0	0	0
029	11/7/2001	3:25 p.m.		0	0	0	0	0	0	0	1	0	0	0	0

Table 2 (Continued)

File no.	Date	Time	Transcript	Interaction			Location				Mood				
				Technical problem	EAR conversation	Other people present	Talking	On the phone	In apartment	Outdoors	In transit	In public place	Laughing	Crying	Sighing
030	11/7/2001	3:36 p.m.		0	0	0	0	0	0	0	1	0	0	0	0
031	11/7/2001	3:48 p.m.		0	0	0	0	0	0	1	0	0	0	0	0
032	11/7/2001	4:00 p.m.	Hi, how are you doing? Yeah. I am in a psychology experiment and they asked me to wear this device.	0	1	1	1	0	0	0	0	1	0	0	0
033	11/7/2001	4:11 p.m.	Do you know if we have laboratory today? You don't? Oh, oh, oh no, she has had it done.	0	0	1	1	0	0	0	0	1	1	0	0
034	11/7/2001	4:23 p.m.	By bus. Number 5 takes you there. I can show you exactly where it is. It leaves at 6:08, the one I want to catch. I missed it yesterday so. Do we keep these, turn these in?	0	0	1	1	0	0	0	0	1	0	0	0

Note: For illustrative purposes only selected coding categories are shown.

developed a coding system to capture the topics of students' daily conversations and we are currently developing a coding system to capture specific coping-relevant aspects of participants' mundane social environments and interactions.

To assess the psychometric properties of the EAR data, we obtain estimates of intercoder reliability by having all coders of a study code a standard set of training EAR recordings. Consistent with the specific, concrete, and behavioral nature of the codings (e.g., 'talking' or 'laughing'), intercoder reliabilities tend to be high. The majority of the SECSI categories have reliabilities that exceed $r = 0.80$ (Mehl et al., 2006; Mehl & Pennebaker, 2003b). It is an advantage of the coding approach that behavior codings at the molecular level (e.g., 'talking') are less susceptible to interpretational ambiguity than psychological ratings at the molar level (e.g., 'relationship satisfaction'). Furthermore, the coding approach yields measures that are based on a nonarbitrary, naturally meaningful, and real-life relevant metric: act frequencies of daily behavior (e.g., percentage of time spent alone or on the phone; cf. Kazdin, 2006).

What are ethical considerations around the EAR method?

Recording ambient sounds around participants raises ethical and legal questions (Fritsche & Linneweber, 2006). EAR studies conducted in our laboratory routinely implement a series of safeguards to protect participants' privacy and to ensure the confidentiality of the data (Mehl et al., 2001; Mehl & Pennebaker, 2003b). We have found these safeguards to be highly effective at alleviating concerns that participants may have about the method. First, the EAR is programmed to record only a fraction of a person's day. Our default pattern, 30 seconds every 12.5 minutes, records less than 5% of the time and leaves more than 95% of participants' days private in the first place. Second, the recordings are kept short; 30-second recordings are long enough to reliably extract basic behavioral information, yet they are short enough to capture little contextualized personal information. Finally, and most importantly, all participants can listen to their EAR recordings and delete parts they do not want on record before the investigators access the data. In a recent study (Mehl et al., 2006), 19 out of 96 participants (19.8%) reviewed their recordings, but only three erased sound files (10 in total).

However, the more serious concerns revolve not around the participants themselves but around bystanders who are not directly involved in the study but whose behaviors are captured by the EAR. In the USA, there are very few restrictions about recording people's utterances in public places. The situation concerning the recording of private conversations is more ambiguous. In most parts of the USA, including Texas and Arizona where the studies from our laboratory have been conducted, recordings can be made legally if at least one of the interactants has knowledge of the

recording device. A small number of states only allow recordings if all interactants have knowledge of the recording device. Even in the most legally restrictive states, however, unauthorized recordings are only problematic if they are personally identifiable.

In EAR studies from our laboratory, participants are encouraged to wear the microphone visibly and to openly mention the EAR in conversations with others. Irrespective of such notification, anonymity of other people's utterances is of paramount importance, because their behavior is collected without informed consent. As mentioned above, the brief recording snippets minimize the chance that personally identifying information about a third person is captured. For further protection, the sound files are coded by staff that is trained and certified for research with human subjects. In the coding process, then, any identifying information is omitted from the transcripts. Finally, participants always have the option of erasing any sound files before the researchers can access them. It is thus highly unlikely that the EAR paradigm as we have established it violates privacy rights of people who are inadvertently recorded.

How obtrusive is the method and how well do participants comply with wearing the EAR device?

The EAR method requires participants to tolerate being intermittently observed without exactly knowing when. This can create self-awareness and result in behavioral reactance (i.e., artificial behavior) and noncompliance (i.e., not wearing the EAR). Therefore, it is critical to estimate how obtrusive the method is in daily life and how well participants comply with it.

Mehl and Holleran (2007) recently addressed these questions by analyzing measures of both self-reported and behaviorally assessed EAR obtrusiveness and compliance in two samples: a short-term (2 days) and a longer-term (10–11 days) monitoring. Self-reported obtrusiveness was measured with items such as 'To what degree were you generally aware of the EAR?' or 'To what degree did the EAR impede on your daily activities?' As a behavioral measure of obtrusiveness, the coders counted in how many sound files participants mentioned the EAR with others. As a self-report measure of compliance, participants reported what percentage of the day they were wearing the EAR. Finally, as a behavioral compliance measure, the coders counted the number of sound files that indicated that participants were not wearing the EAR. 'Not wearing the EAR' was coded if over a 30-second recording period no ambient sounds whatsoever, that is not even sounds of breathing or clothes rubbing against the microphone, were recorded.

The analyses painted the following picture about the method's obtrusiveness: Immediately after receiving the EAR, participants go through a brief period of heightened self-awareness in which conversations about

the EAR are frequent. Within two hours of wearing the device, however, most participants habituate to the method and rarely mentioned it with others anymore. This habituation effect was found not only for the short-term monitoring but also for the longer-term monitoring. In the longitudinal sample, some individuals initially talked about the method more than others; yet after 5–6 days of wearing the device, virtually all participants had adjusted to it and barely mentioned it anymore in their daily conversations.

The study further yielded the following findings about participants' compliance: In the short-term monitoring, participants' compliance was very high in the first hours after they had received the EAR. Noncompliance gradually increased over time and leveled off at about 10–12% on the second day of monitoring. Compliance in the longer-term monitoring was high for at least 6 days. After that, variability in noncompliance increased suggesting that some participants' tolerance threshold may have been reached.

Taken together, these analyses indicate (i) that EAR compliance and obtrusiveness can be reliably assessed, (ii) that compliance is generally high and comparable to what has been reported for self-report-based EMAs (Green, Rafaeli, Bolger, Shrouf, & Reis, 2006), and (iii) that after an initial habituation period of about 2 hours, the method operates fairly unobtrusively and does not impede participants much in their normal activities.

What things are captured on the sound files? To what extent does the EAR reveal real life?

As the metaphorical researcher's ear on the participant's lapel, the EAR essentially eavesdrops on people's daily lives. That naturally begs the question of what things are captured on the sound files and to what extent the EAR reveals real life. One of my first 'aha!' experiences when we started doing EAR research was how ordinary and mundane real life really is. The sound files we obtained from participants first and foremost documented that for most people most of real life is not thrilling, glittery, and extraordinary. In the end, our daily lives tend to be fairly 'average' (Taylor & Brown, 1988). In the words of one of our participants (after listening to her own sound files), 'I, as probably most people, like to think of myself as interesting and superior. Listening to myself, however, I have concluded that I am most certainly not. I am just like everybody else.' Much of what the EAR captures is either silence (apart from ambient noises) or rather banal and linguistically unrefined conversations that reveal participants engaged in the pursuit of their daily activities (e.g., school, commute, watching TV). In essence, the majority of the sound files speak to the ordinary, humdrum nature of daily life (Craik, 2000).

Yet, with its fine-grained grid of observations, the EAR also regularly captures some of the less publicly presentable aspects of a person's social behavior. For example, the EAR regularly catches intimate conversations as well as emotional outbursts, arguments, and profanity. In addition to documenting a person's behavior 'on stage', it also reveals some of those moments where humans are caught off-guard showing their usually hidden, weak, and unpolished faces (Goffman, 1959). This potential of the EAR to capture 'off-stage' behavior can, for example, be used by social health researchers to get a better handle on the assessment of theoretically important but methodologically notoriously difficult to measure negative support behaviors such as criticizing/blaming, hostility, and withdrawal (cf. Manne & Zautra, 1989).

How can researchers get the EAR? What practical things are important to know to use it?

Recently, there has been a change in the logistics of the EAR. The software-based system was originally programmed and distributed by John Price, our long-term EAR engineer (Mehl et al., 2001). As of now, my laboratory has taken over its distribution and maintenance. One implication of this change is that we are now making the software available for academic use at no cost. Researchers can obtain a copy of the software directly from my laboratory. All we ask for in return is for users to share their experiences with us and to provide feedback on how the system can be improved. Our hope is that this change will lower the psychological and economic hurdles for researchers who are interested in the EAR and foster a more widespread use of the method.

It is beyond the scope of this article to provide all the necessary practical information for running an EAR study. However, we recently created a researcher's guide with this purpose in mind. This guide is available from my laboratory and contains (i) hardware recommendations (e.g., the software runs only on Windows Mobile operated PDAs), (ii) instructions for how to install and use the software, (iii) a sample consent form, (iv) a set of standard questionnaires (e.g., EAR compliance and obtrusiveness questionnaire), and (v) a script for how to administer the EAR. Apart from providing this guide, we are always happy to help 'jump start' and troubleshoot.

Implications of the EAR Method for Social Health Research

In the remainder of the article I outline important implications of the EAR method for social health research. In its basic approach, the EAR has the same goal as traditional EMA methods: to study psychosocial processes where they naturally occur – in the flow of people's daily lives. Yet, as a naturalistic observation method, it offers unique potentials. Specifically, I propose that it can enrich social health research in at least

three important ways: it can help (i) cross-validate research findings independent of self-reports, (ii) calibrate psychological measures against behavioral markers of real-world social functioning, and (iii) further our understanding of the role that people's mundane social interactions and language use play for coping and health.

Cross-validating research findings independent of self-reports

At a basic level, the EAR provides an opportunity for social health researchers to cross-validate important research findings with a method that is independent of participants' self-reports. Frequently in health research, both the psychological construct to be predicted (e.g., health) and the construct used as predictor (e.g., social support) are assessed with questionnaires. However, when predictor and criterion share important method variance, they are susceptible to similar response biases. This can undermine the validity of a substantive research finding.

Unfortunately, this is not an esoteric measurement issue for devoted methodologists. Rather, it is an issue that has repeatedly surfaced in psychology and has had broad consequences for the field. For example, to what extent does a negative correlation between self-reported neuroticism and self-reported health indicate that neurotic individuals are less healthy or that neurotic individuals perceive their lives, including their health, overly negatively (Watson & Pennebaker, 1999). Similarly, to what extent does a positive correlation between self-reported positive illusions about the self and self-reported mental health indicate that positive illusions foster mental health or that individuals who hold positive illusions in various domains of life also hold positive illusions about their mental health (Colvin & Block, 1994). In essence, correlations within methods run the risk of confounding response content and response style.

Traditional EMA methods have done much to alleviate this problem. For example, when participants' levels of depression are assessed with a standardized depression measure and their daily interactions with momentary self-reports, any emerging relationship among the two cannot simply be due to generalized negative self-schemas (Nezlek, Imbrie, & Shean, 1994). Yet, as mentioned before, global or retrospective and momentary self-reports still share method variance; both are based on participants' reports of their introspections and perceptions. Therefore, both are similarly susceptible to potential methodological confounds such as self-deception, socially desirable responding, and mood, salience, or context effects (Conner et al., 2007; Piasecki et al., 2007).

For example, depressed individuals may experience more negative mood in daily life that may lead them to perceive their social interactions overly negatively. EAR-derived observer ratings (e.g., rated interaction quality) or behavior codings (e.g., coded laughing) could counteract this effect (cf. Mehl, 2006b). Similarly, when defense mechanisms are at work, discrepancies

between momentary self-report and momentary observational data can be psychologically informative. For example, narcissists may report glorious interactions with others in their daily diaries but observer ratings may paint a more modest picture (cf. Clifton, Turkheimer, & Oltmanns, 2004).

In essence, then, the EAR can help determine convergent and discriminant validity in the study of real-world psychological processes. Often, traditional EMAs and the EAR will yield similar findings; at times, they will reveal critical discrepancies – both outcomes are informative.

Calibrating psychological measures against behavioral markers of real-world social functioning

Another persistent problem that the EAR can help get a better handle on is the calibration of psychological measures against important real-world behaviors. More than a decade ago, Sechrest, McKnight, and McKnight (1996) pointed out that ‘very few psychological measures of any kind are expressed in a metric that is intuitively or immediately meaningful’ (p. 1065) and that ‘science, understanding of behavior, and ability to intervene to help people would be advanced by a better understanding of the measures by which the phenomena which we concern ourselves are gauged’ (p. 1068).

In psychology, the vast majority of measures use arbitrary metrics, such as 5- or 7-point rating scales. Measures with arbitrary metrics have no clear referents that inform about the implications that scoring at a certain level (e.g., a ‘4’ on optimism) has for how a person fares in important domains of life. How much more time does a person with a ‘2’ on an extraversion scale spend alone than a person with a ‘4’? How much less time does a person with a ‘2’ on a Conscientiousness scale spend in class than a person with a ‘4’? And, how does a person’s daily life change if an intervention reduces the person’s depression score by 7 points (or, say, half a standard deviation)? Finding answers to questions like these is crucial for understanding and communicating the practical implications of psychological effects. Yet, the field has struggled greatly with accomplishing this. Blanton and Jaccard (2006) and Kazdin (2006) recently reminded that the issue of arbitrary metrics is as urgent today as it was 10 years ago.

How can the EAR help with calibrating psychological measures? As described above, the EAR-derived behavioral codings can be readily converted into a metric that is nonarbitrary, intuitively meaningful, and inherently real-world relevant. If the EAR captures a person talking in 40 out of 120 recordings, one can estimate that the person spent about a third of the day (or about 5 hours) talking. Similarly, if TV sounds are found in 20% of the recordings, one can estimate that the person was (actively or passively) watching TV 20% of the day (or about 3.5 hours). By linking EAR-derived frequencies of daily behavior to the metric of measures, a better understanding of the real-world implications of psychological effects can be obtained.

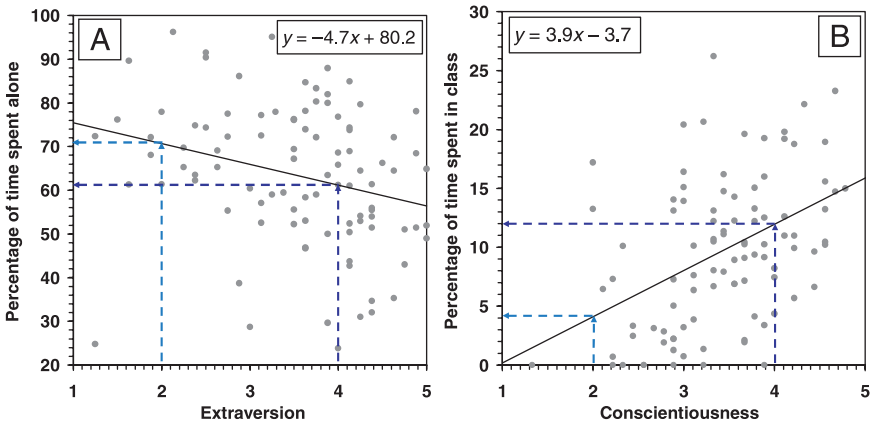


Figure 3 Calibration of two personality measures: Regressing participants' Extraversion (panel A) and Conscientiousness (panel B) scores on their EAR-assessed daily behaviors (panel A: percentage of time spent alone; panel B: percentage of time spent in class).

Note: The scatter plots and regression equations are based on data reported in Mehl et al. (2006); Extraversion and Conscientiousness were assessed with the Big Five Inventory (John & Srivastava, 1999) using a 5-point unipolar scale; dashed arrows indicate the estimated behavioral frequencies for individuals one point below (2) and one point above (4) the midpoint of the scale.

For example, in a study on behavioral manifestations of personality in daily life, Mehl et al. (2006) found that extraversion was correlated $r = -0.27$ with EAR-assessed time spent alone and Conscientiousness $r = 0.42$ with EAR-assessed time spent in class. Figure 3 illustrates what these correlations convey about the real-world implications of the two personality measures. The regression equations reveal that individuals who marked a '4' on the 5-point extraversion scale spent almost 10% less time alone than individuals who marked a '2' (70.8% versus 61.4%, Panel A). And, individuals who marked a '4' on the 5-point Conscientiousness scale spent about three times more time in class than individuals who marked a '2' (11.9% versus 4.1%; Panel B).

Similarly, Mehl (2006b) found in a college student sample that, among moderately and severely depressed participants, higher scores on the Beck Depression Inventory (short form; Beck & Beck, 1972) were related to less social engagement. Individuals at the cutoff for severe depression spent almost 14% more time alone than individuals at the cutoff for moderate depression (64.4% versus 50.6%). Interestingly, among non- or mildly depressed participants, Beck Depression Inventory scores were unrelated to students' levels of social engagement.

Finally, in response to a highly publicized but empirically unsupported claim that women use about 20,000 words per day while men use only 7,000 words, Mehl, Vazire, Ramirez-Esparza, Slatcher, and Pennebaker (2007) reanalyzed data from six EAR studies to estimate how many words

men and women really use every day. Given that the EAR was set to record 4% of the time, it was straightforward how to estimate daily word use from the number of recorded words. The analyses showed that men and women both used about 16,000 words per day – with large individual differences around the mean but no evidence of a reliable sex difference (the least talkative participant used an estimated 695 words; the most talkative an impressive 47,016 words). The number of words spoken per day is a powerful nonarbitrary metric for measuring talkativeness and is likely partly responsible for why the study received broad public attention.

Taken together, the fact that the EAR tracks people's daily lives with a fine-meshed grid, combined with the interpretative advantage of its measures of daily behavior makes the method well suited for the calibration of psychological scales and effects against real-world behavior. Because the EAR assesses audible aspects of a person's social environment and interactions, it most directly serves as a behavioral criterion of real-world social functioning.

Exploring the role of people's mundane social interactions in the context of coping and health

A third important potential of the EAR for social health research stems from its ability to provide researchers with a representative sample of a person's real-world conversations and language use. Peoples' daily interactions with others are arguably the most important platform on which health-relevant social processes unfold. How does being socially integrated protect people's immune system and health if not through the interactions they have with their social network (Cohen, 2004)? How can the emotional support that a breast cancer patient receives protect her mental health if not through the interactions she has with significant others (Weihs et al., 2005)? And, how can marital quality affect long-term survival after heart failure if not through aspects of couples' daily interactions (Rohrbaugh, Shoham, & Coyne, 2006)? In essence, then, daily social interactions are a core 'carrier' of the health effects of social variables.

The significance of studying how people interact with important others is widely recognized. It is an established paradigm in relationship research to bring couples to the laboratory to observe how they discuss areas of disagreement (Heyman, 2001; Margolin et al., 1998). Also, EMA researchers have long analyzed people's real-world interactions using diary methods (Laurenceau & Bolger, 2005; Reis & Wheeler, 1991). Yet, the naturalistic observation of ordinary, mundane interactions has received little scientific attention.

This is important because by method constraint, both existing approaches, that is, the observations of interactions in the laboratory and the study of daily interactions using diaries, tend to focus on interactions that are extended in duration, important in content, and focused in nature. In other words,

most research so far has looked at accentuated interactions that stand out as clear figures on the ground of the number of short and seemingly trivial social encounters people have every day (Mehl & Pennebaker, 2003b). To better understand the health implications of people's social interactions, it is important to also study their ordinary, mundane conversations. The EAR samples these conversations representatively and documents them word by word.

For example, Mehl & Pennebaker (2003a) used the EAR to track students' conversations in the aftermath of the attacks of September 11, 2001, and to test how aspects of their interactions predicted their adjustment to the events. Over the first 10 days after 9/11, students gradually shifted their conversations from group-based to dyadic (i.e. one-on-one) interactions. Interestingly, and consistent with the idea that ordinary interactions have psychological relevance apart from focused 'coping interactions' about the stressor, it was the degree to which students increased in their overall one-on-one interactions, not the amount of 9/11 conversations they had, that predicted their psychological adjustment at follow-up 2 weeks later.

In a series of ongoing studies, we are currently following up on this finding and exploring what kind of and how dyadic interactions facilitate adjustment to major upheavals, such as life-changing or life-threatening events. Specifically, we are zooming in on the daily interactions that patients diagnosed with a life-changing (i.e. rheumatoid arthritis) or a life-threatening (i.e. breast cancer) illness have with members of their social networks. The goal of these studies is to identify aspects of people's mundane conversations that can account for the salutary effects that social variables such as social support and interpersonal disclosure have in these coping contexts. Preliminary analyses so far indicate that indeed only a small fraction of participants' daily interactions are about arthritis or breast cancer; the majority of the captured conversations are interpersonal but nonfocal in nature. This suggests that coping constructs such as interpersonal disclosure and social support are probably best understood as broad phenomena that, in addition to direct, illness-focused conversations encompass various aspects of people's seemingly trivial but psychologically consequential, ordinary, humdrum, social encounters.

Summary and Conclusions

The purpose of this article was to provide a conceptual and methodological discussion of a relatively novel, observational momentary assessment method in the context of social health research. As the metaphorical researcher's ear on the participant's lapel, the EAR eavesdrops on people's daily lives and provides highly naturalistic, vivid, and psychologically rich information about their moment-to-moment (acoustic) social worlds.

The most immediate potential of the method for social health research probably lies in its ability to track people's everyday conversations and

language use. The words people naturally use have been found to have important health implications (Pennebaker, Mehl, & Niederhoffer, 2003). To highlight only selected recent findings, the use of first person singular pronouns ('I', 'me', 'my') has been found to be related to the number of depressive symptoms (Mehl, 2006b; Rude, Gortner, & Pennebaker, 2004). The use of first person plural pronouns ('we', 'us', 'our') has been found to be related to how individuals cope with a collective upheaval (Cohn, Mehl, & Pennebaker, 2004). And the use of social words (such as references to family and friends) has been found to be related to longevity (Pressman & Cohen, 2007). These are intriguing findings; yet, overall we still know very little about how word use and health are related. The EAR can capture people's word use when talking to critical members of their social networks. That way, it is a powerful tool to advance the study of the health implications of everyday language use.

In many ways the method is still in its infancy. Yet, after close to a decade of development and research with it, it has sufficiently matured to be of value to the field. It is encouraging that its use has by now spread beyond the group of original developers; about a handful of 'early adopter' laboratories have implemented it in studies. Interestingly, and consistent with the potentials outlined in this article, the majority of these studies have used it within a multimethod approach to come to a more complete understanding of the role that people's daily social environments, interactions, and language use play in the psychosocial regulation of health.

Acknowledgments

The author thanks James Pennebaker for his mentoring and support over the years and John Price for the technical implementation of the EAR method and his support over the years. Thanks also to Megan Robbins for her comments on a previous draft of this manuscript. The preparation of this article was supported by an American Cancer Society Institutional Research Grant (IRG-74-001-28).

Short Biography

Matthias Mehl's research is located at the intersection of social/personality psychology and health psychology. He is interested in the psychological implications of people's daily social lives. In his health research, he investigates how people use their daily interactions with others to cope with life-changing and life-threatening events. A central thread that runs through his work is the development of alternative assessment methods that can complement psychology's reliance on self-reports. His research in this area has focused on naturalistic observation, quantitative text analysis, informant reports, and physiological measurement. He has authored or coauthored articles in various scientific journals, including *Science*,

Psychological Science, Journal of Personality and Social Psychology, Annual Review of Psychology, Psychological Assessment, Journal of Psychosomatic Research, Journal of Research in Personality, and Behavioral Research Methods, Instruments, and Computers. He completed his undergraduate education at the Friedrich-Alexander University in Erlangen, Germany, and received his PhD in Social/Personality Psychology from the University of Texas at Austin in 2004. Since then he has been an Assistant Professor of Psychology at the University of Arizona where he is currently also an Adjunct Assistant Professor in the Department of Communications and an Associate Investigator at the Arizona Cancer Center.

Endnote

* Correspondence address: Department of Psychology, PO Box 210068, University of Arizona Tucson, AZ 85721, USA. Email: mehl@email.arizona.edu.

References

- Barker, R. G., & Wright, H. F. (1951). *One Boy's Day: A specimen Record of Behavior*. New York: Harper & Brothers.
- Beck, A. T., & Beck, R. W. (1972). Screening depressed patients in family practice: A rapid technique. *Postgraduate Medicine*, **52**, 81–85.
- Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, **61**, 27–41.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, **54**, 579–616.
- Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of Psychophysiology*. New York: Cambridge University Press.
- Clifton, A., Turkheimer, E., & Oltmanns, T. F. (2004). Contrasting perspectives on personality problems: Descriptions from the self and others. *Personality and Individual Differences*, **36**, 1499–1514.
- Cohen, S. (2004). Social relationships and health. *American Psychologist*, **59**, 676–684.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, **15**, 687–693.
- Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin*, **116**, 3–20.
- Conner, T., Barrett, L. F., Tugade, M. M., & Tennen, H. (2007). Idiographic personality: The theory and practice of experience sampling. In R. W. Robins, R. C. Fraley, & R. Kreuger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 79–96). New York: Guilford Press.
- Coyne, J. C., & Gottlieb, B. H. (1996). The mismeasure of coping by checklist. *Journal of Personality*, **64**, 959–991.
- Craik, K. H. (2000). The lived day of an individual: A person-environment perspective. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person-Environment Psychology: New Directions and Perspectives* (pp. 233–266). Mahwah, NJ: Lawrence Erlbaum Associates.
- Dunning, D. (2005). *Self-Insight: Roadblocks and Detours on the Path to Knowing Thyself*. New York: Psychology Press.
- Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory assessment – monitoring behavior in daily life settings: A behavioral-scientific challenge for psychology. *European Journal of Psychological Assessment*, **23**, 206–213.
- Fritzsche, I., & Linneweber, V. (2006). Nonreactive methods in psychological research. In M. Eid & E. Diener (Eds.), *Handbook of Multimethod Measurement in Psychology* (pp. 189–203). Washington, DC: American Psychological Association.

- Funder, D. C. (2007). *The personality puzzle*. New York: W. W. Norton & Co.
- Giese-Davis, J., Piemme, K. A., Dillon, C., & Twirbutt, S. (2005). Macro-variables in affective expression in women with breast cancer participating in support groups. In J. Harrigan, K. R. Scherer & R. Rosenthal (Eds.), *Nonverbal Behavior in the Affective Sciences: A Handbook of Research Methods* (pp. 399–445). Oxford, UK: Oxford University Press.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday.
- Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, **11**, 87–105.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2006). Experience sampling method: Measuring the quality of everyday life. Thousand Oaks, CA: Sage.
- Heyman, R. E. (2001). Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment*, **13**, 5–35.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of Personality Theory and Research* (pp. 102–138). New York: Guilford Press.
- Kazdin, A. E. (2006). Arbitrary metrics: Implications for identifying evidence-based treatments. *American Psychologist*, **61**, 42–49.
- Kerig, P. K., & Baucom, D. H. (Eds.) (2004). *Couple Observational Coding Systems*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kiecolt-Glaser, J. K., McGuire, L., Robles, T., & Glaser, R. (2002). Emotions, morbidity, and mortality: New perspectives from psychoneuroimmunology. *Annual Review of Psychology*, **53**, 83–107.
- Laurenceau, J.-P., & Bolger, N. (2005). Using diary methods to study marital and family processes. *Journal of Family Psychology*, **19**, 86–97.
- Manne, S. L., & Zautra, A. J. (1989). Spouse criticism and support: Their association with coping and psychological adjustment among women with rheumatoid arthritis. *Journal of Personality and Social Psychology*, **56**, 608–617.
- Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A. M., Ghosh, C. M., et al. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clinical Child and Family Psychology Review*, **1**, 195–213.
- Mehl, M. R. (2006a). Quantitative text analysis. In M. Eid & E. Diener (Eds.), *Handbook of Multimethod Measurement in Psychology* (pp. 141–156). Washington, DC: American Psychological Association.
- Mehl, M. R. (2006b). The lay assessment of subclinical depression in daily life. *Psychological Assessment*, **18**, 340–345.
- Mehl, M. R., & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the Electronically Activated Recorder (EAR). *European Journal of Psychological Assessment*, **23**, 248–257.
- Mehl, M. R., & Pennebaker, J. W. (2003b). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, **84**, 857–870.
- Mehl, M. R., & Pennebaker, J. W., (2003a). The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science*, **14**, 579–585.
- Mehl, M. R., Pennebaker, J. W., Crow, M., Dabbs, J., & Price, J. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, and Computers*, **33**, 517–523.
- Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, **317**, 82.
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, **90**, 862–877.
- Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: What it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry and Neuroscience*, **31**, 13–20.
- Nezlek, J. B., Imbrie, M., & Shean, G. D. (1994). Depression and everyday social interaction. *Journal of Personality and Social Psychology*, **67**, 1101–1111.

- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of language use: Our words, our selves. *Annual Review of Psychology*, **54**, 547–577.
- Piasecki, T. M., Hufford, M. R., Solhan, M., & Trull, T. J. (2007). Assessing clients in their natural environments with electronic diaries: Rationale, benefits, limitations, and barriers. *Psychological Assessment*, **19**, 25–43.
- Pressman, S. D., & Cohen, S. (2007). Use of social words in autobiographies and longevity. *Psychosomatic Medicine*, **69**, 262–269.
- Reis, H. T., & Wheeler, L. (1991). Studying social interaction with the Rochester Interaction Record. *Advances in Experimental Social Psychology*, **24**, 269–318.
- Rohrbaugh, M. J., Shoham, V., & Coyne, J. C. (2006). Effect of marital quality on eight-year survival of patients with heart failure. *American Journal of Cardiology*, **98**, 1069–1072.
- Rude, S. S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression vulnerable college students. *Cognition and Emotion*, **18**, 1121–1133.
- Sapolsky, R. (2004). *Why Zebras Don't Get Ulcers*. New York: Henry Holt & Co.
- Schwarz, N., & Sudman, S. (Eds.) (1993). *Autobiographical Memory and the Validity of Retrospective Self-Reports*. New York: Springer-Verlag.
- Sechrest, L., McKnight, P. E., & McKnight, K. M. (1996). Calibration of measures for psychotherapy outcome studies. *American Psychologist*, **51**, 1065–1071.
- Sillars, A. L. (1991). Behavioral observation. In B. Montgomery & S. Duck (Eds.). *Studying Interpersonal Interaction* (pp. 197–218). New York: Guilford Press.
- Smyth, J., & Stone, A. (2003). Ecological momentary assessment research in behavioral medicine. *Journal of Happiness Studies*, **4**, 35–52.
- Somerfield, M. P., & McCrae, R. R. (Eds.) (2000). Stress and coping research: Methodological challenges, theoretical advances, and clinical applications. *American Psychologist*, **55**, 620–673.
- Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, **16**, 199–202.
- Stone, A. A., Shiffman, S. S., Atienza, A., & Nebeling, L. (Eds.) (2007). *The Science of Real-Time Data Capture: Self-Reports in Health Research*. New York: Oxford University Press.
- Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.) (2000). *The Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Suls, J., & Martin, R. M. (1993). Daily recording and ambulatory monitoring methodologies in behavioral medicine. *Annals of Behavioral Medicine*, **15**, 3–7.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, **103**, 193–210.
- Tennen, H., Affleck, G., & Armeli, S. (2003). Daily processes at the interface of social and health psychology. In J. Suls & K. Wallston (Eds.), *The Social Foundations of Health and Illness* (pp. 495–529). Oxford, UK: Blackwell Publishers.
- Trull, T. J. (2007). Expanding the aperture of psychological assessment: Introduction to the special section on innovative clinical assessment technologies and methods. *Psychological Assessment*, **19**, 1–3.
- Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity. *Psychological Review*, **96**, 234–254.
- Weihls, K. L., Simmens, S. J., Mizrahi, J., Enright, T. M., Hunt, M. E., & Siegel, R. S. (2005). Dependable social relationships predict overall survival in Stages II and III breast carcinoma patients. *Journal of Psychosomatic Research*, **59**, 299–306.