Introduction to the Special Section on Self-Monitoring: A Major Assessment Method in Clinical Psychology

John D. Cone
United States International University

One of the most frequently used assessment methods, self-monitoring (SM), is reviewed in this Special Section. Despite its popularity with applied psychologists, relatively little research on its methodological adequacy or clinical utility has occurred in the past 20 years. The 8 articles invited for this Special Section are designed to renew interest in studying SM. In this regard, their authors have focused on SM as an assessment method, devoting relatively little attention to its use as a treatment procedure. Issues of terminology, SM parameters, and criteria for evaluating the quality of SM data are covered. In addition, descriptions of the use of SM with various clinical problems and populations, along with suggestions for motivating the collection of high-quality data are presented in the articles. Finally, the authors share their impressions of what is needed next if the scientific soundness of the method and its clinical utility are to continue to evolve.

Self-monitoring is an assessment method used extensively in clinical psychology (Elliot, Miltenberger, Kaster-Bundgaard, & Lumley, 1996). Though its use is widespread, information as to its psychometric adequacy and research on its clinical utility are not. For a brief period in the late 1960s through the early 1980s, research on self-monitoring flourished. More recently, however, methodological research has waned. A purpose of the present Special Section of articles is to take stock of the current status of self-monitoring (SM) research and to stimulate more of it. The Special Section's contributors have been chosen to represent a broad spectrum of established researchers and clinicians with extensive expertise in both assessment and intervention. Each treats a particular subset of the literature dealing with SM as an assessment method and emphasizes issues of relevance to that particular subset. The authors were encouraged to focus on assessment rather than treatment applications of SM. Though each contribution is unique, some concerns cut across the different articles and are discussed briefly in this introductory article.

Terminological Issues

Just exactly what SM is and how it is distinct from other assessment methods never has been clearly settled. First, it is not the personality construct popularized by Snyder (1987). For Snyder, SM reflects the extent to which a person "treats interactions with others as dramatic performances designed to gain attention, make impressions, and at times entertain" (p. 178). As an assessment method, SM has been defined as the act of systematically observing and recording aspects of one's own behavior (Bornstein, Hamilton, & Bornstein, 1986, p. 176). For the purposes of this Special Section, this definition can be expanded to include con-textual variables relevant to the behavior as well. Thus, SM can be seen as "the act of systematically observing and recording aspects of one's own behavior and internal and external environmental events thought to be functionally related to that behavior." This incorporates the many instances in which the person is asked to note and record events occurring immediately prior to and after the behavior, such as whether the urge to smoke occurred in the context of drinking alcohol with friends.

Terminological clarity might be purchased by ceding the term SM to personality assessors and theorists, referring hereafter to self-observation. This is logically defensible given that one of the two major components of the method involves observing, detecting, or sensing one's own behavior and/or events related to it. Further, the term *self-observation* highlights parallels between SM and direct-observation assessment methods. For example, recording formats (e.g., event recording or momentary time sampling) are comparable across the two methods, as are numerous other issues such as response definition, observer training, and observer accuracy. Unfortunately, SM has been used long enough and consistently enough to have gathered momentum that makes changing terms unlikely at this point.

Distinguishing SM from self-report assessment is useful and relatively easily accomplished. Indeed, when placed along the directness continuum, self-report is closer to the indirect end, and SM is closer to the direct (Cone, 1978). The three aspects of the continuum (i.e., topographic, temporal, and spatial) provide clear separation between the two methods. SM relies on observations of *the* behavior of clinical interest (i.e., topographically identical responses) at the *time* (i.e., temporally contiguous) and *place* (i.e., natural setting) of its actual occurrence. In contrast, self-report uses stands-in or surrogates (verbal descriptions, reports) of the behavior of interest that are obtained at a time and place different from the time and place of the behavior's actual occurrence. Some of the articles that follow refer to "self-reported estimates" or simply "estimates" to designate self-report methods (e.g., Korotitsch & Nelson-Gray, 1999). Craske and Tsao (1999) use the term "liberal self-monitoring" when describing the recall of past events.

Correspondence concerning this article should be addressed to John D. Cone, Department of Psychology and Family Studies, United States International University, 10455 Pomerado Road, M-15, San Diego, California 92131. Electronic mail may be sent to jcone@usiu.edu.

Presumably "conservative self-monitoring" applies to assessment by the assessee in which there is minimal time lapse between the observing and recording components, identified by Bornstein et al. (1986). My own preference is to use this condition (i.e., minimal time lapse) as a distinguishing one, which along with temporal concurrence and topographic equivalence makes up the essential characteristics of the method.

There are at least four important components of SM that are given more or less emphasis in the articles that follow. First, observing, detecting, or sensing the response or environmental event must occur. Second, some notation or record of the occurrence is made. Third, analysis of the notations is undertaken. Fourth, the results of the analysis are displayed. Remembering to look at or notice one's own behavior or contextual events surrounding it cannot be assumed when using SM. Various clever ways of signaling clients to observe have been reported, including remotely activating a beeper to cue recording (e.g., Csikszentmihalyi & Larson, 1987) or having palm-top computers beep to promote observation and data entry (Shiffman et al., 1997). Barton, Blanchard, and Veazy (1999) discuss both equipment-assisted and equipment-mediated forms of SM in terms of sensing and notation functions. Recording can be done with paper-and-pencil formats as well, and these can be of varying degrees of structure, from diary-like narratives expected of the client to preprinted checklists the client merely marks. Wilson and Vitousek (1999) provide an example of a recording form used to track food use by persons with eating disorders. When handheld computers are used for the cuing and recording functions, it is a small step to using them for analysis and presentation or display of results as well. Thus, each entry can be added instantly to prior ones, and graphs are then automatically updated and displayed, thereby providing ongoing information to the client of progress with behavior change efforts.

Parameters of Self-Monitoring

SM parameters depend heavily on the uses to be made of the data and the type of problem being treated. To illustrate, though it is best to make observations during periods when the behavior is most likely to occur (e.g., mealtimes, when alone in the evenings), functional analyses can be facilitated by observing and recording environmental antecedents and consequences at times when the behavior does not occur as well. Important considerations when using SM include the following: (a) when and how often the client is expected to observe and note relevant behavior, (b) over what period of time SM is to be used, (c) what is supposed to be recorded, (d) how the recording is to occur (i.e., the format of the record), (e) whether cueing is to be used and how, (f) the number of different items to be monitored, and (g) any means to be used to assure compliance.

Whereas observation and notation of every occurrence of a behavior can be a gold standard worth aspiring to with many clients, Wilson and Vitousek (1999) warn in their article that overly frequent weighing is to be avoided with some forms of eating disorder. If lots of different events are monitored, then observer load can be moderated through the use of momentary time sampling in which the client is signaled to note occurrence or nonoccurrence for each event at specific moments in time. Presumably, these are randomly selected to provide representative estimates of occurrence. As for the time period over which SM should be used, rules of thumb have yet to be established. If it is used primarily for descriptive purposes to inform functional analyses for planning treatment, then it might be discontinued when functional relationships are identified. If it is subsequently used to monitor treatment effectiveness, then its use covers the same period of time as treatment itself. In their article, Barton et al. (1999) cite evidence that the time period to be covered with SM depends on the behavior monitored and on the purposes for which the data are to be put. With tension headaches, a 2-week baseline might be sufficient for research purposes, for example, with only 1 week needed for clinical purposes. With migraines, longer baselines might be needed for both purposes.

Exactly what is to be recorded requires care in selection. Illustrative is Wilson and Vitousek's (1999) admonition to avoid calorie counting with persons with anorexia or bulimia. In a similar vein, Craske and Tsao (1999) note that SM is rarely used to assess ruminative and ritualistic behavior in persons with obsessive–compulsive disorder because of the high likelihood of such persons becoming compulsive about SM. Response property or dimensional quantity (Johnston & Pennypacker, 1993) is relevant, with frequency being used most often. Duration and magnitude (or intensity) are sometimes more appropriate, however, or at least should be recorded along with frequency. Foster, Laverty-Finch, Gizzo, and Osantowski (1999) describe varying recording formats, including continuous noting of all frequencies over a time period (event recording) or in brief intervals throughout the period (whole interval or partial interval recording). Participants can be given preprinted forms to mark as behaviors occur or they can be cued randomly to respond to a series of questions. Children can be prompted to mark one of a graded series of smiley faces to indicate the intensity of emotions being monitored (e.g., Beidel, Neal, & Lederer, 1991; Lang & Cuthbert, 1984). Less structured formats include diaries and journals in which clients are encouraged to note their experiences, as in descriptive experience sampling (Hurlburt, 1997).

Remembering to observe and record one's own behavior is challenging and is assisted in many studies by some sort of auditory signaling device. Users of Csikszentmihalyi's experience sampling method (ESM, Csikszentmihalyi & Figurski, 1982) use telephone pagers to cue participants to note and record events at randomly determined times. Shiffman et al. (1997) used palm-top computers to cue participants to record their cigarette smoking. Simple timers can be used as well, with participants trained to reset them to go off at random intervals.

Observer load is an important consideration, with overall data accuracy and reliability likely to be inversely related to the number of different events a person is monitoring (Hayes & Cavior, 1977, 1980). Craske and Tsao (1999) provide useful suggestions for minimizing overload, including restricting SM to the first N minutes of an observation interval, restricting it to the first N occurrences of an event, and defining recordable events as those that include certain critical features, excluding all others that might be similar but are lacking those features. Baird and Nelson-Gray (1999) suggest starting with a few behaviors and gradually fading in additional ones. Designing a system with careful attention to other SM parameters can lead to the successful collection of more data than is likely if multiple parameters are ignored. For example, a momentary time-sampling format in which participants are cued

at random times to record which of 12 different behaviors they are engaging in at the moment might yield data with different psychometric characteristics than a continuous observation format that requires the behaviors to be tracked on an ongoing basis. Illustrative are recent findings by Schimmack and Diener (1997) of positive correlations between intensity ratings of pleasant and unpleasant affect when event recording is used, and substantially lower correlations when momentary time sampling is used.

It is also likely that compliance with SM requests interacts with a number of these parameters. For example, Hermann, Peters, and Blanchard (1995) found that recording headaches at specific times occurred more consistently when clients were told the palm-top computers they were to use had internal clocks to determine when their entries were made. Barton et al. (1999) report a number of clever ways of enhancing compliance with SM requests.

Evaluating the Adequacy of Self-Monitored Data

As with any assessment method, numerous variables can affect the adequacy of SM data. These include how carefully people are trained to monitor their behavior, the number of concurrent response requirements (e.g., number of different events recorded, dimensional quantities of each), response bias (e.g., social desirability), demand characteristics (e.g., "vital for treatment to work" vs. "might be useful ..."), whether recording accuracy is consequated, the length of the recording period, reactivity, awareness that accuracy checks can be made on a random basis (e.g., through collaterals, behavioral byproducts, biological indicators), availability heuristics, and the degree of adherence to SM requirements. SM must be evaluated against stringent psychometric criteria just as any other assessment method must be. However, depending on whether cognitive, motor, or physiological behavior is being observed, conventional approaches to accuracy, reliability, and validity will be more or less applicable. Jackson (1999) treats these issues in her article, noting that the appropriateness of psychometric concepts also varies depending on whether behavior per se or hypothetical constructs indicated by behavior are being assessed. Foster et al. (1999) cite studies showing variable agreement between self-observers and external observers and make five recommendations for maintaining data quality when using SM assessment. Craske and Tsao (1999) offer important comparisons between SM and retrospective estimating (self-reports), discussing response demand biases, memory biases, the inflation of aversive experiences, and the impact of availability heuristics. With respect to the last, they note that available or easy-to-recall events tend to be judged more numerous. Thus, salient events such as panic attacks being more retrievable can be susceptible to overestimation. Presumably, SM data are less affected by availability heuristics, though relatively few studies have compared SM and self-report methods directly. As Barton et al. (1999) observe, those that have compared SM and self-reports present conflicting results. Some show overestimates of global self-reports compared with SM; however, others show the reverse. Wilson and Vitousek (1999) comment that clients with bulimia nervosa tend to overreport amounts of food eaten 24 hr earlier when compared with more precise information available from laboratory assays. Unfortunately, they can only speculate whether self-reports such as these would be surpassed in accuracy by SM. More research showing the

independent and comparative psychometric adequacy of SM is clearly warranted (Taylor, 1999).

In the managed cost context characterizing contemporary behavioral health services, it will be important to examine the clinical utility of SM more thoroughly. Are higher quality service outcomes achieved more rapidly and with greater duration when SM is used? How does the overall clinical utility of SM compare with other assessment methods? Which of the clinical benefits associated with SM are due to the quality of its assessment data on the one hand and its reactive effects on the other?

Uses of Self-Monitoring

The articles in this Special Section reflect the extremely broad uses to which SM has been put in recent years. It is evident that SM can facilitate multiple aspects of assessment activities aimed at designing effective interventions. In addition to its use to obtain precise information about behavior, and relevant antecedent and consequent environmental events necessary to inform functional analyses of the behaviors, it can be used to monitor the effects of ongoing treatment and provide data critical to treatment implementation. It can also facilitate or potentiate the effects of treatment. To illustrate, it can serve as a discriminative stimulus to use aspects of the treatment on an ongoing basis, or to bring more client focus and self-management to the therapy process (Wilson & Vitousek, 1999). Moreover, SM can be used as a criterion for evaluating the adequacy of other assessment methods. Delespaul (1995) suggested the ESM be used in this way, for example, and Craske and Tsao (1999) cite research in which SM was used to validate self-report measures of obsessive–compulsive behavior.

SM has been used with an extraordinarily wide range of problem behaviors and medical conditions, as the articles in this Special Section document. Shapiro and Cole (1999) provide excellent examples of SM's application with children, especially in school settings. Peterson and Tremblay (1999) show its versatility in behavioral medicine with children, and Barton et al. (1999) do the same for adults.

Summary

Research and applied activities of psychologists are based on data of the highest quality available. This means data from the most direct forms of assessment method. SM is the most practical direct assessment method, and it's the only direct approach for tapping private events or cognitive behavior. The articles that follow provide a comprehensive overview of the state of the art with respect to SM. They are meant to revive interest in research on the methodology and clinical utility of the approach in hopes that its continued popularity with applied psychologists will be grounded in scientific evidence of its effectiveness. Careful consideration of the issues raised in these articles and their authors' recommendations will provide a solid foundation for anyone planning to use SM, whether for research or practical purposes.

References

Baird, S., & Nelson-Gray, R. O. (1999). Direct observation and self-monitoring. In S. C. Hayes, D. H. Barlow, & R. O. Nelson-Gray (Eds.). *The scientist practitioner: Research and accountability in the age of managed care* (2nd ed., pp. 353–386). Boston: Allyn & Bacon.

Barton, K. A., Blanchard, E. B., & Veazy, C. (1999). Self-monitoring as an assessment strategy in behavioral medicine. *Psychological Assessment, 11,* 490–497.

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

Bornstein, P. H., Hamilton, S. B., & Bornstein, M. T. (1986). Self-monitoring procedures. In A. R. Ciminero, C. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 176–222). New York: Wiley.

Cone, J. D. (1978). The Behavioral Assessment Grid (BAG): A conceptual framework and a taxonomy. *Behavior Therapy, 9,* 882–888.

Craske, M. G., & Tsao, J. C. I. (1999). Self-monitoring with panic and anxiety disorders. *Psychological Assessment, 11,* 466–479.

Csikszentmihalyi, M., & Figurski, T. J. (1982). Self-awareness and aversive experience in everyday life. *Journal of Personality, 50,* 15–28.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease, 175,* 526–536.

Delespaul, P. A. E. G. (1995). *Assessing schizophrenia in daily life: The experience sampling method.* Maastricht, the Netherlands: Universitaire Pers Maastricht.

Elliot, A. J., Miltenberger, R. G., Kaster-Bundgaard, J., & Lumley, V. (1996). A national survey of assessment and therapy used by behavior therapists. *Cognitive and Behavioral Practice, 3,* 107–125.

Foster, S. L., Laverty-Finch, C., Gizzo, D. P., & Osantowski, J. (1999). Practical issues in self-observation. *Psychological Assessment, 11,* 426–438.

Hayes, S. C., & Cavior, N. (1977). Multiple tracking and the reactivity of self-monitoring: I. Negative behaviors. *Behavior Therapy, 8,* 819–831.

Hayes, S. C., & Cavior, N. (1980). Multiple tracking and the reactivity of self-monitoring: II. Positive behaviors. *Behavioral Assessment, 2,* 283–296.

Hermann, C., Peters, M. L., & Blanchard, E. B. (1995). The use of hand-held computers for symptom-monitoring: The case of chronic headache. *Mind Body Medicine, 1,* 59–70.

Hurlburt, R. T. (1997). Randomly sampling thinking in the natural environment. *Journal of Consulting and Clinical Psychology, 65,* 941–949.

Jackson, J. L. (1999). Psychometric considerations in self-monitoring assessment. *Psychological Assessment, 11,* 439–447.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment, 11,* 415–425.

Lang, P. J., & Cuthbert, B. N. (1984). Affective information processing and the assessment of anxiety. *Journal of Behavioral Assessment, 6,* 369–395.

Peterson, L., & Tremblay, G. (1999). Self-monitoring in behavioral medicine: Children. *Psychological Assessment, 11,* 458–465.

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

Shapiro, E. S., & Cole, C. L. (1999). Self-monitoring in assessing children's problems. *Psychological Assessment, 11,* 448–457.

Shiffman, S., Hufford, M., Hickcox, M., Paty, J. A., Gnys, M., & Kassel, J. D. (1997). Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of Consulting and Clinical Psychology, 65,* 292–300.

Snyder, M. (1987). *Public appearances, private realities: The psychology of self-monitoring.* New York: Freeman.

Taylor, S. (1999). Behavioral assessment: Review and prospect. *Behaviour Research & Therapy, 37,* 475–482.

Wilson, G. T., & Vitousek, K. M. (1999). Self-monitoring in the assessment of eating disorders. *Psychological Assessment, 11,* 480–489.