INTRODUCTION

The architects of the original *Diagnostic and Statistic Manual* (American Psychiatric Association, 1952) were more or less doomed from the start. The definition of mental illness and criteria for the specific disorders were vague at best, and the reliability of diagnosis (even for the major and common disorders) was so poor that it called into serious question the validity of the entire classification system itself. The problem of nebulous and inadequately specified criteria was largely addressed in subsequent revisions of the *DSM*, of which the *DSM-IV-TR* (American Psychiatric Association, 2000) is the current incarnation. Another major contributor to poor reliability of diagnosis was the lack of uniformity of diagnostic questions to evaluate psychiatric symptoms and arrive at a formal diagnosis. This problem has been addressed by the development of structured interviews. Undeniably, since the 1970s the ability of clinicians and researchers to accurately diagnose psychiatric disorders has improved in quantum leaps, and structured interviews have contributed significantly to the advancement in diagnostic clarity and precision. In this chapter, we provide an overview of the major features and advantages of structured interviews, followed by a discussion of the most popular multidisorder structured interviews. We conclude with a review of
more specialized structured interviews to provide the reader with an appreciation for the diversity of clinical information that structured interviews can provide.

BASIC ISSUES REGARDING STRUCTURED INTERVIEWS

Traditional assessment methods have relied on the unstructured, or "clinical," interview, behavioral and observational assessment, and psychometric testing. Regarding the interview component, there are important differences between an unstructured interview and a structured one. For starters, unstructured clinical interviews are heavily influenced by the individual client's needs, the client's responses, and the clinician's intuitions. With unstructured interviews, clinicians are entirely responsible for asking whatever questions they decide are necessary for them to reach a diagnostic conclusion. In fact, any type of question (relevant or not) can be asked in any way that fits the mood, preferences, training, and philosophy of the clinician. The amount and specific kind of information gathered during an interview are largely determined by the clinician's theoretical model (e.g., psychoanalytic, behavioral), view of psychopathology, training, knowledge base, and interpersonal style. As a consequence, one can imagine the kind of inconsistency and variability in an interview from one clinician to another.

Structured interviews, on the other hand, conform to a standardized list of questions, including follow-up questions, a standardized sequence of questioning, and, finally, systematized ratings of the client's responses. In fact, the impetus for the development of structured interviews was generated by the need to standardize questions and provide explicit guidelines for categorizing or coding responses. Adoption of such procedures serves to: (1) increase coverage of many disorders that otherwise might be overlooked, (2) enhance the diagnostician's ability to accurately determine if a particular symptom is present or absent, and (3) reduce variability among interviewers (i.e., reduce unreliability). These features of structured diagnostic interviews add much to developing clinical psychology into a true science (i.e., structured interviews are subject to evaluation and statistical analysis, and they are modified and improved based on the emerging database of the field).

It is also important to emphasize that the term structured interview is a broad one and that the actual amount of "structure" provided by an interview varies considerably. Basically, structured interviews can be divided into one of two types: fully structured and semistructured. In a fully structured interview, questions are asked verbatim to the respondent, the wording of probes used to follow up on initial questions is specified, and interviewers are trained not to deviate from this very specific format. In a semistructured interview, although the initial questions for each symptom are specified and are typically asked verbatim to the respondent, the interviewer has substantial latitude to follow up on responses. For example, the interviewer can modify existing questions and probes in any way
and even devise completely new, innovative questions to more accurately rate specific symptoms. The amount of structure provided in a structured interview clearly impacts the extent of clinical experience and judgment needed to administer the interview appropriately: Semistructured interviews require clinically experienced examiners to administer the interview and to make diagnoses, whereas fully structured interviews can be administered by nonclinicians who receive training on the specific instrument, thus making fully structured interviews economical to use, especially in large studies.

Structured interviews are used in many different venues and for many different purposes. Application of structured interviews falls into three broad areas: research, clinical, and training use. The research domain is probably the most common for structured interviews, in which the interview is used to diagnose participants accurately so that etiology, comorbidity, and treatment approaches (among other topics) can be studied for a particular diagnosis or group of diagnoses. Certainly, good research requires that individuals assigned a diagnosis truly meet full criteria for that diagnosis. In clinical settings, structured interviews may be used as part of a comprehensive and standardized intake evaluation. A variation on this theme is that a structured interview may be used to clarify and confirm diagnoses based on an initial unstructured interview. Use of structured interviews for training in the mental health field is an ideal application because interviewers have the opportunity to learn (through repeated administrations) specific questions and follow-up probes used to elicit information and evaluate specific diagnostic criteria provided by the DSM system. Modeling one's own questions and flow of the interview from a well-developed structured interview can be an invaluable source of training for the clinician.

Over the past several decades, proliferation of structured interviews has been steady. Structured interviews have been created to assist with the differential diagnosis of all major Axis I (clinical) and all standard Axis II (personality) disorders. These structured interviews used for diagnosis are typically aligned with the DSM system and therefore assess the formal diagnostic criteria specified in the manual. But structured interviews for differential diagnosis are not the only kind of structured interviews; other structured interviews are more narrow in focus, for example, to assess a specific problem or form of psychopathology (e.g., eating disorders, borderline personality features) in great depth.

In the last two decades, evidence-based practice has become the cornerstone of clinical psychology. With the field's increasing emphasis on establishing psychotherapy as an empirically supported method of treating mental illness, the reliability and validity of our methods of diagnosis have become all-important components of treatment and clinical research. In order to justify the use of psychotherapy, we must justify the methods by which we classify and ascertain diagnoses. This becomes blatantly problematic when two different clinicians, interviewing the same client, furnish two different diagnoses, because they chose to ask completely different sets of questions. As noted earlier, one solution to this quandary was the development of structured interviews.
ADVANTAGES AND DISADVANTAGES OF STRUCTURED AND SEMISTRUCTURED INTERVIEWS

Structured interviews were developed to improve our ability to accurately diagnose individuals, in both clinical and research settings. By systemizing the questions clinicians ask and the way answers to those questions are recorded and interpreted, structured interviews minimize needless variability in diagnostic evaluations (Morrison, 1988; Rogers, 2001; Rubinson & Asnis, 1989; Segal & Coolidge, 2003). Thus, by increasing reliability, structured interviews also increase validity of diagnosis, assuming the criteria on which diagnosis is based are also valid. In addition, structured interviews can be invaluable training tools for laypersons and clinicians-in-training (Morrison).

Despite the advantages of structured interviews, their use is not without controversy. Critics of structured interviews maintain that utilizing such interviews can severely damage rapport and, thus, the therapeutic relationship essential to psychotherapy (Rogers, 2001; Rubinson & Asnis, 1989). Some critics also argue that the validity of structured interviews is questionable, because such interviews are based on diagnostic criteria that have not been established as valid in all cases (Rogers; Rubinson & Asnis). Finally, structured interviews may be limited because they sacrifice either the breadth or the depth of information attained (Rogers; Rubinson & Asnis). Next, we examine each of the arguments in more detail. A brief summary of advantages and disadvantages is presented in Table 6.1.

ADVANTAGES OF STRUCTURED AND SEMISTRUCTURED INTERVIEWS

Increased Reliability

Perhaps the most important advantage of structured interviews centers on increased reliability. By systemizing and standardizing the questions interviewers ask and the way those questions are presented, structured interviews improve reliability in a variety of ways. First, structured interviews decrease the amount of information variance in interviews (Segal & Coolidge, 2003). That is, structured interviews decrease the chances that two different interviewers will elicit different information from the same client, which may result in different diagnoses. Interviewers may arrive at different information from clients for a variety of reasons. For example, they may ask different questions, they may cover different criteria for specific disorders, they may ask questions in different sequences, they may rate the intensity of client's reported symptoms in different ways, or they may record the information clients give them differently (Rogers, 2001). Structured interviews help eliminate this variability by standardizing all of these aspects of a diagnostic evaluation. Thus, interrater reliability, or the like-
### Table 6.1 Advantages and Disadvantages of Structured Interviews

<table>
<thead>
<tr>
<th>Advantages of Structured Interviews</th>
<th>Disadvantages of Structured Interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Increased Reliability:</strong> Because questions are standardized, structured interviews result in decreased variability among interviewers, which enhances interrater reliability. Structured interviews also increase the reliability of assessment for a client's symptoms across time as well as the reliability between client report and collateral information.</td>
<td><strong>May Hinder Rapport:</strong> Use of structured interviews may damage rapport because they are problem centered, not person centered, and poorly trained interviewers may neglect to use their basic clinical skills during the assessment.</td>
</tr>
<tr>
<td><strong>Increased Validity:</strong> Structured interviews ensure that diagnostic criteria are covered systematically and completely. This is important because it serves to increase the validity of diagnosis.</td>
<td><strong>Limited by the Validity of the Classification System Itself:</strong> Structured interviews used for diagnosis are inherently tied to diagnostic systems. Thus, they are only as valid as the systems on which they are based. Furthermore, it is difficult to establish the validity of particular structured interviews because there is no gold standard in psychiatric diagnosis.</td>
</tr>
<tr>
<td><strong>Utility as Training Tools:</strong> Nonclinicians can easily be trained to administer fully structured interviews, which can be cost effective in both research and clinical settings. In addition, structured interviews are excellent training tools for clinicians-in-training because structured interviews promote the learning of specific diagnostic questions and probes used by experienced clinical interviewers.</td>
<td><strong>Breadth versus Depth:</strong> Structured interviews are limited because they cannot cover all disorders or topic areas. When choosing a structured interview, one must think carefully about the trade-offs of breadth versus depth of assessment.</td>
</tr>
</tbody>
</table>

Increased interrater reliability has broad implications in both clinical and research settings. Because many methods of psychological treatment are intimately tied to diagnoses, it is imperative that diagnoses be accurate (Segal & Coolidge, 2001). Thus, if different clinicians interviewing the same client arrive at different diagnostic conclusions, it would be challenging at best to make a definitive decision about treatment. Similarly, accurate diagnosis is also essential for some types of clinical research, for example, that addressing the causes and treatments of specific forms of psychopathology (Segal & Coolidge). Imagine a study examining different treatments for Major Depressive Disorder. In such a study, it would be imperative to be certain that those in the treatment groups...
actually suffer from depression. Indeed, we must be able to diagnose participants with depression accurately and definitively before we can even begin to examine effectiveness of treatment for the disorder or theories of etiology.

In addition to increasing interrater reliability, structured interviews increase the likelihood that the diagnosis is reliable across time and across different sources of information (Rogers, 2001). In many clinical and research settings, clients or participants are in fact assessed on different occasions. The danger in making multiple assessments is that if an interviewer evaluates a client in a different manner with different questions on different occasions, the client’s presentation may be completely different, not because the client’s symptoms or diagnosis has changed but rather because the way the client is asked about those symptoms has changed. Utilizing a standardized interview for multiple assessments helps ensure that if a client’s presentation has changed, it is because his or her symptoms are actually different, not because of variance in interviews (Rogers). Likewise, in many settings, clinicians conduct collateral interviews with important people in the client’s life to get a broader picture of the client’s symptoms, problems, and experiences. Using a structured interview for both a client and a collateral source may greatly increase the chances that discrepancies between client and collateral interviews are real, rather than just by-products of interviewing styles (Rogers).

**Increased Validity**

Validity of diagnosis has to do with the meaningfulness or usefulness of the diagnosis, and reliability is a required prerequisite for validity. Thus, by virtue of the fact that structured interviews greatly increase reliability of diagnosis, they also increase the likelihood that a diagnosis is valid. Structured interviews also improve the validity of diagnoses in other ways. The systematic construction of structured interviews lends a methodological validity to these types of assessments, compared to unstructured approaches. Because structured interviews are designed to assess well-defined diagnostic criteria thoroughly and accurately, they are often better assessments of those criteria than are unstructured interviews (Rogers, 2001; Segal & Coolidge, 2003). According to Rogers, clinicians who use unstructured interviews sometimes diagnose too quickly and may miss comorbid diagnoses. Thus, because structured interviews “force” clinicians to assess all of the specified criteria for a broad range of diagnoses, they offer a more thorough and valid assessment of many disorders, compared to unstructured interviews.

To elaborate, some unstructured interviews may provide information about the presence or absence of only a few common mental disorders. Coverage of other disorders may be neglected during an unstructured interview if, for example, the interviewer is unfamiliar with the specific criteria of some disorders. Some unstructured interviews may also provide limited information about whether comorbid psychopathology exists as well as inconsistent information about the severity of the psychopathology. Structured interviews, because they incorporate
systematic ratings, easily provide information that allows for the determination of the level of severity and the level of impairment associated with a particular diagnosis, and structured interviews provide the same information about any comorbid conditions.

Utility as Training Tools

Structured interviews can be invaluable training tools for both laypersons working in the field and clinicians-in-training (Morrison, 1988; Segal & Coolidge, 2003). Structured interviews can be a useful means of ensuring that laypersons who are making preliminary mental health assessments, for example, intake staff at hospitals, are evaluating individuals thoroughly and accurately. In the case of nonclinician interviewers, fully structured interviews are advisable because they minimize the amount of clinical judgment needed for accurate administration. Structured interviews can be invaluable tools in the training of mental health professionals as well. Becoming familiar with structured interviews may help inexperienced clinicians develop an understanding of the flow, format, and questions inherent in a good diagnostic interview.

DISADVANTAGES OF STRUCTURED AND SEMISTRUCTURED INTERVIEWS

May Hinder Rapport

The most common criticism of structured interviews is that their use may significantly damage rapport (Rogers, 2001; Rubinson & Asnis, 1989; Segal & Coolidge, 2003). Critics of structured interviews argue that reliable and accurate diagnosis of a client is useless if it comes at the cost of the development of the therapeutic alliance, which is the basis of psychotherapy. Structured interviews may damage rapport because they are problem centered rather than person centered. There is a danger that interviewers may get so wrapped up in the protocol of their interview that they fail to demonstrate the warmth and genuine regard necessary to form an alliance. In addition, interviewers who are overly focused on the questions they must “get through” in an interview may miss important behavioral cues or other information that could prove essential to the case (Rogers; Segal & Coolidge).

Proponents of structured interviews point out that the problem of rapport building during a structured interview can be overcome with training and experience (Rogers, 2001). Structured interviews can be conducted in such a way that they help to establish rapport and understanding of the client, especially if interviewers make an effort to utilize their basic clinical skills. In order to ensure that this is the case, however, interviewers must be aware of the potential negative effects of structured interviews on rapport building and make building a therapeutic alliance a prominent goal during an interview, even when they are also focused on following protocol. It behooves those who use structured interviews to engage
their respondents in a meaningful way during the interview and to avoid a rotelike interviewing style that may serve to alienate. On the other hand, some clients actually like the structured interview approach to assessment because it is perceived as thorough and detailed; in these cases, rapport is easily attained.

**Limited by the Validity of the Classification System Itself**

Earlier, we noted that structured interviews frequently offer a more valid assessment of diagnostic criteria than unstructured interviews. Thus, proponents of structured interviews claim, structured interviews are more valid in general. The assumption inherent in this argument is that our diagnostic criteria are valid. Some would argue, however, that this assumption is a false one (Morrison, 1988; Rogers, 2001). Diagnostic criteria, in particular the criteria of the DSM, have often been criticized for lack of validity. Although each successive edition of the DSM has been better grounded in scientific research, critics have maintained that the criteria for some diagnoses are not well examined enough to constitute any sort of validity (Segal & Coolidge, 2003). This point is evidenced by the fact that the criteria for many disorders have changed significantly from one edition to another in the evolution of the DSM. Furthermore, research suggests that the criteria in the DSM are often severely limited by culture (Rogers). Thus, the criteria may only be valid for a particular group of individuals at a particular point in time. All in all, the way we conceptualize diagnoses, while improving, is far from perfect. And, because structured interviews are intimately tied to diagnostic criteria, they are, by definition, limited by the same inadequacies inherent in those criteria.

In addition to this problem with diagnostic criteria, structured interviews have other problems with validity. Specifically, it is difficult to establish the validity of any particular structured interview (Morrison, 1988). Our best means of establishing validity of a structured interview is to compare diagnoses obtained from such interviews to diagnoses obtained by expert clinicians or by other structured interviews. This is inherently problematic because we cannot be certain that diagnoses by experts or other structured interviews are themselves valid (Morrison; Rogers, 2001; Segal & Coolidge, 2003).

**Breadth versus Depth**

A final criticism of structured interviews centers on the fact that no one structured interview can be all things in all situations. A particular structured interview cannot cover all disorders and eventualities (Morrison, 1988; Rogers, 2001; Segal & Coolidge, 2003). For example, if a structured interview has been designed to cover an entire diagnostic system (like the DSM, which identifies over several hundred specific disorders), then the inquiries about each disorder must be limited to a few inclusion criteria. Thus, the fidelity of the official diagnostic criteria has been compromised for the sake of a comprehensive interview. If the fidelity of the criteria is not compromised, then the structured interview
becomes unwieldy, in terms of time and effort, on the part of both the interviewer and interviewee. Most structured interviews attempt some kind of compromise between these two approaches.

Thus, as for breadth versus depth of approach, users of structured interviews are forced to make a choice about what is most useful in a given situation. Both choices have their limitations. If a clinician or researcher decides to utilize an interview that provides great breadth of information, it is ensured that a wide range of disorders and a great many different areas of a respondent's life are assessed. However, one may not have the depth of information needed to fully conceptualize a case. On the other hand, deciding to utilize an interview focused on one or two specific areas will provide clinicians and researchers with a wealth of information about those specific areas, but it may result in missing information that could lead to an additional diagnosis or a different case conceptualization. Thus, it is essential to understand that when choosing a particular structured interview, there are often trade-offs regarding breadth and depth of information.

WEIGHING BOTH ADVANTAGES AND DISADVANTAGES

Our examination of the arguments for and against the use of structured interviews highlights the importance of carefully contemplating what is needed in a particular clinical or research situation before choosing to utilize a structured interview. Structured interviews can be invaluable tools in both clinical and research work; however, it is essential that one not employ such tools without accounting for some of the problems inherent in their use. Another perspective voiced by Rogers (2001) is that it would be unwise to view the interviewing process as an either/or proposition (i.e., unstructured vs. structured interview). In certain situations, unstructured interviews may meet the objectives of a particular clinical inquiry more efficiently than a structured interview. For example, in a crisis situation, flexibility on the part of the clinician is needed to meet the pressing demands of this fluid and potentially volatile interaction. However, structured interviewing allows for the assessment of the reliability of the interviewing process itself, which speaks to the validity of psychiatric diagnoses as well as of the entire classification system.

In summary, it is apparent that introduction of operationalized, specified, empirically derived, and standardized criteria for mental disorders in conjunction with introduction of standardized structured diagnostic interviews has revolutionized the diagnostic process and vastly improved reliability and validity. A conclusion that one can draw about the specific impact of structured interviews is that they have greatly improved clinical and research endeavors by providing a more standardized, scientific, and quantitative approach to the evaluation of mental disorders and clinical problems. Specific interviews are discussed next.
STRUCTURED AND SEMISTRUCTURED INTERVIEWS FOR DIFFERENTIAL DIAGNOSIS

Interviews discussed in this section are the Diagnostic Interview Schedule for DSM-IV, the Schedule for Affective Disorders and Schizophrenia, the Structured Clinical Interview for DSM-IV Axis I Disorders, and the Structured Clinical Interview for DSM-IV Axis II Personality Disorders. All assess a variety of disorders and therefore can assist in the important task of differential diagnosis (i.e., discriminating which disorder(s) among similar ones the respondent meets criteria for) while also providing for a full assessment of comorbid psychopathology.

THE DIAGNOSTIC INTERVIEW SCHEDULE FOR DSM-IV

The Diagnostic Interview Schedule for DSM-IV (DIS-IV; Robins et al., 2000) is designed to ascertain the presence or absence of major psychiatric disorders of the DSM-IV (American Psychiatric Association, 1994). It is unique among the multidisorder diagnostic interviews in that it is a fully structured interview specifically designed for use by nonclinician interviewers, whereas the other interviews are semistructured. By definition, a fully structured interview clearly specifies all questions and probes and does not permit deviations. Thus, the DIS-IV, by virtue of its structure, minimizes the amount of clinical judgment and experience required to administer it.

To ensure standardized administration of the DIS, the paper-and-pencil version of the instrument is no longer recommended, due to the complicated format. Instead, a computerized version of the DIS-IV (C-DIS) is recommended. Computerized administration may be interviewer administered or self-administered. In both formats, the exact wording of all questions and probes are presented to the respondent in a fixed order on a computer screen, and rephrasing of questions is discouraged, although DIS interviewers can repeat questions as necessary to ensure that they are understood by the respondent. All questions are written to be closed-ended, and replies are coded with a forced choice “yes” or “no” format, which eliminates the need for clinical judgment to rate responses. The DIS gathers all necessary information about the subject from the subject, and collateral sources of information are not used. The DIS is self-contained and covers all necessary symptoms to make many DSM-IV diagnoses. The coded responses are entered directly into a database during the interview, and the diagnosis is made according to the explicit rules of the DSM-IV diagnostic system.

In 1978, development of the original DIS was begun by researchers at the Washington University Department of Psychiatry in St. Louis at the request of the National Institutes of Mental Health (NIMH). At that time, the NIMH Division of Biometry and Epidemiology was planning a set of large-scale, multicenter epidemiological investigations of mental illness in the general adult population in the United States as part of its Epidemiological Catchment Area Program. Variables under study included incidence and prevalence of many
psychiatric disorders and utilization profiles of health and mental health services. With this impressive purpose in mind, development of a structured interview that could be administered by nonclinicians was imperative, due to the prohibitive cost of using professional clinicians as interviewers. As a result, the DIS was designed as a fully structured diagnostic interview, and it was explicitly crafted so that it can be administered and scored by nonclinician interviewers.

The DIS has undergone several major revisions since its inception. For example, the original DIS (Robins, Helzer, Croughan, & Ratcliff, 1981) covered criteria for DSM-III (American Psychiatric Association, 1980) disorders. DIS questions and diagnostic algorithms were revamped to establish compatibility with DSM-III-R (American Psychiatric Association, 1987); this is called Version DIS-III-R (Robins, Helzer, Cottler, & Goldring, 1989). The current version of the DIS (Version IV; Robins et al., 2000) is closely tied to the DSM-IV system; to this end, DSM diagnostic criteria for the disorders have been faithfully turned into specific questions on the DIS.

Because the DIS was designed for epidemiological research with normative samples, interviewers do not elicit a presenting problem from the respondent, as would be typical in unstructured clinical interviews. Rather, DIS interviews begin by asking questions about symptoms in a standardized order. Like other structured interviews, the DIS has sections that cover different disorders. Each diagnostic section is independent, except where one diagnosis preempts another. Once a symptom is reported to be present, further closed-ended questions are asked about diagnostically relevant information, such as severity, frequency, time frame, and possibility of organic etiology of the symptom. The DIS includes a set of core questions that are asked of each respondent. Core questions are followed by contingent questions that are administered only if the preceding core question is endorsed. DIS interviewers utilize a "probe flowchart" that indicates which probes to use in which circumstances.

For each symptom, the respondent is asked to state whether it has ever been present and how recently. All data about the presence or absence of symptoms and time frames of occurrence are coded and entered into the computer. Consistent with its use of nonclinician interviewers who may not be overly familiar with the DSM-IV or psychiatric diagnosis, the diagnostic output of the DIS is generated by a computer program that analyzes data from the completed interview. The output provides estimates of prevalence for two time periods: current and lifetime.

Due to its highly structured format, full administration of the DIS-IV typically requires between 90 and 150 minutes. To shorten administration time, the modular format makes it possible to drop evaluation of disorders that are not of interest in a particular study. Another option is to drop further questioning for a particular disorder once it is clear that the threshold number of symptoms needed for diagnosis will not be met. Although designed for use by nonclinician administrators, training for competent administration of the DIS is necessary. Trainees typically attend a 1-week training program at Washington University, during
which they review the DIS manual, listen to didactic presentations about the structure and conventions of the DIS, view videotaped vignettes, complete workbook exercises, and conduct several practice interviews, followed by feedback and review. Additional supervised practice is also recommended.

The psychometric properties of the original DIS and its revisions are excellent, and such data have been documented in an impressive array of studies. The interested reader is referred to Compton and Cottler (2004) for an excellent summary of the psychometric characteristics of the DIS. Overall, the DIS has proven to be a popular and useful diagnostic assessment tool, especially for large-scale epidemiological research. The DIS has been translated into over a dozen languages. It is used in countries across the globe for epidemiological research and served as the basis for the Composite International Diagnostic Interview used by the World Health Organization. Presently, the DIS-IV is the only well-validated case-finding strategy that can make DSM-IV diagnoses in large-scale epidemiological research. Like earlier versions, the DIS-IV can be expected to enjoy widespread application in psychiatric research, service, and training. For information on DIS materials, training, and developments, the interested reader may consult the DIS Web site, http://epi.wustl.edu.

THE SCHEDULE FOR AFFECTIVE DISORDERS AND SCHIZOPHRENIA

The Schedule for Affective Disorders and Schizophrenia (SADS; Endicott & Spitzer, 1978) is a semistructured diagnostic interview designed to evaluate a range of Axis I clinical disorders, with a focus on mood and psychotic disorders. Ancillary coverage is provided for anxiety symptoms, substance abuse, psychosocial treatment history, and antisocial personality features. The SADS provides in-depth but focused coverage of the mood and psychotic disorders and also supplies meaningful distinctions of impairment in the clinical range for these disorders.

The original SADS focused on psychiatric symptoms as specified by the Research Diagnostic Criteria (RDC; Spitzer, Endicott, & Robins, 1978), which made available specific inclusion and exclusion criteria for many psychiatric disorders. The RDC predated publication of the DSM-III (American Psychiatric Association, 1980) and was a significant predecessor of that system. Many of the specified criteria described in the RDC were adopted for inclusion in DSM-III. As such, much information derived from SADS interviews can be applied to make DSM-based diagnoses.

The SADS is intended to be used with adult respondents and to be administered by trained mental health professionals. It focuses heavily on the differential diagnosis of mood and psychotic disorders, with great depth of assessment in these areas. In the beginning of the interview, a brief overview of the respondent's background and psychiatric problems is elicited in an open-ended inquiry. The SADS is then divided into two parts, each focusing on a different time period.
Part I provides for a thorough evaluation of current psychiatric problems and concomitant functional impairment. A unique feature of the SADS is that for the current episode, symptoms are rated when they were at their worst levels, to increase diagnostic sensitivity and validity. In contrast, Part II provides a broad overview of past episodes of psychopathology and treatment. Overall, the SADS covers over 20 diagnoses in a systematic and comprehensive fashion and provides for diagnosis of both current and lifetime psychiatric disorders. Some examples include Schizophrenia (with six subtypes), Schizoaffective Disorder, Manic Disorder, Hypomanic Disorder, Major Depressive Disorder (with 11 subtypes), Minor Depressive Disorder, Panic Disorder, Obsessive-Compulsive Disorder, Phobic Disorder, alcoholism, and Antisocial Personality Disorder (Endicott & Spitzer, 1978).

In the SADS, questions are clustered according to specific diagnoses, which improves the flow of the interview. For each disorder, standard questions are specified to evaluate specific symptoms of that disorder. Questions are either dichotomous or rated on a Likert scale, which allows for uniform documentation of levels of severity, persistence, and functional impairment associated with each symptom. To supplement patient self-report and obtain the most accurate symptom picture, the SADS allows for consideration of all available sources of information (i.e., chart records, input from relatives). In addition to the standard questions asked of each respondent, optional probes may be selectively used to clarify responses, and unstructured questions may be generated by the interviewer to augment answers to the optional probes. Thus, considerable clinical experience and judgment are needed to administer the SADS. To reduce the length of administration and evaluation of symptoms that are not diagnostically significant, many diagnostic sections begin with screening questions that provide for "skip-outs" to the next section if the respondent shows no evidence of having the disorder. Administration of the SADS typically takes between 1 1/2 and 2 1/2 hours. Formal diagnostic appraisals are made by the interviewer after the interview is completed. At present, no computer scoring applications have been designed, due to the complex nature of the diagnostic process and the strong reliance on clinical judgment.

As noted earlier, the SADS was designed for use by trained clinicians. Considerable clinical judgment, interviewing skills, and familiarity with diagnostic criteria and psychiatric symptoms are requisite for competent administration. As such, it is recommended that the SADS be administered only by professionals with graduate degrees and clinical experience, such as clinical psychologists, psychiatrists, and psychiatric social workers (Endicott & Spitzer, 1978). Training in the SADS is intensive and can encompass several weeks. The process includes reviewing the most recent SADS manual and practice in rating written case vignettes and videotaped SADS interviews. Additionally, trainees typically watch and score live interviews as if participating in a reliability study with a simultaneous-rating design. Throughout, discussion and clarification with expert interviewers regarding diagnostic disagreements or difficulties add to the experience.
Finally, trainees conduct their own SADS interviews, which are observed and critiqued by the expert trainers.

Numerous additional versions of the SADS have been devised, each with a distinct focus and purpose. Perhaps the most common is the SADS-L (Lifetime version), which can be used to make both current and lifetime diagnoses but has significantly fewer details about current psychopathology than the full SADS and results in a quicker administration time. The SADS-L generally is used with nonpsychiatric samples in which there is no assumption of a significant current psychiatric problem. The SADS—Change Version is also popular and consists of 45 key symptoms from the SADS Part 1. Extensive study of the SADS suggests that it possesses excellent psychometric characteristics, and the interested reader is referred to Rogers, Jackson, and Cashel (2004) for a comprehensive review of these data.

The SADS has been translated into several languages, but its primary use has been in North America. The SADS has been widely used in clinical research over the past three decades and consequently has a large body of empirical data associated with it. As such, it is often the instrument of choice for clinical researchers desiring in-depth assessment of depression and schizophrenia. The extensive subtyping of disorders provided by the SADS is also highly valued by clinical researchers. However, due to its length and complexity, the SADS is less often chosen for use in typical clinical settings.

THE STRUCTURED CLINICAL INTERVIEW FOR DSM-IV AXIS I DISORDERS

The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) is a flexible, semistructured diagnostic interview designed for use by trained clinicians to diagnose many adult DSM-IV Axis I mental disorders. The current version is the product of many prior editions, which were updated and modified over time. With each revision, the SCID has been reworked to enhance accuracy and ease of use, culminating in the February 2001 revision, when the SCID was updated to match the DSM-IV-TR (American Psychiatric Association, 2000). The SCID-I has widespread popularity as an instrument to obtain reliable and valid psychiatric diagnoses for clinical, research, and training purposes, and it has been used in more than 1,000 studies.

The original SCID was designed for application in both research and clinical settings. Recently, the SCID has been split into two distinct versions: the Research Version and the Clinician Version. The Research Version covers more disorders, subtypes, and course specifiers than the Clinician Version and therefore takes longer to complete. The benefit, however, is that it provides a wealth of diagnostic data that are particularly valued by clinical researchers. The research version is distributed by the Biometrics Research Department of the New York State Psychiatric Institute.
The Clinician Version of the SCID (SCID-CV; First, Spitzer, Gibbon, & Williams, 1997a) is designed for use in clinical settings. It has been trimmed down to encompass only those DSM-IV disorders that are most typically seen in clinical practice and can further be abbreviated on a module-by-module basis. The SCID-CV contains six self-contained modules of major diagnostic categories (Mood Episodes, Psychotic Symptoms, Psychotic Disorders, Mood Disorders, Substance Use Disorders, and Anxiety and Other Disorders). Table 6.2 provides a list of specific DSM-IV diagnoses covered by the SCID-CV.

**Table 6.2 Diagnostic Coverage of the SCID—Clinician Version**

<table>
<thead>
<tr>
<th>Mood Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar I Disorder</td>
</tr>
<tr>
<td>Bipolar II Disorder</td>
</tr>
<tr>
<td>Bipolar Disorder NOS</td>
</tr>
<tr>
<td>Cyclothymic Disorder</td>
</tr>
<tr>
<td>Dysthymic Disorder</td>
</tr>
<tr>
<td>Major Depressive Disorder</td>
</tr>
<tr>
<td>Depressive Disorder NOS</td>
</tr>
<tr>
<td>Mood Disorder Due to a GMC</td>
</tr>
<tr>
<td>Substance-Induced Mood Disorder</td>
</tr>
<tr>
<td>Psychotic Disorders</td>
</tr>
<tr>
<td>Schizophrenia (with five subtypes)</td>
</tr>
<tr>
<td>Schizophreniform Disorder</td>
</tr>
<tr>
<td>Schizoaffective Disorder</td>
</tr>
<tr>
<td>Delusional Disorder</td>
</tr>
<tr>
<td>Brief Psychotic Disorder</td>
</tr>
<tr>
<td>Psychotic Disorder Due to a GMC</td>
</tr>
<tr>
<td>Substance-Induced Psychotic Disorder</td>
</tr>
<tr>
<td>Psychotic Disorder NOS</td>
</tr>
<tr>
<td>Substance Use Disorders</td>
</tr>
<tr>
<td>Alcohol Abuse/Alcohol Dependence</td>
</tr>
<tr>
<td>Amphetamine Abuse/Amphetamine Dependence</td>
</tr>
<tr>
<td>Cannabis Abuse/Cannabis Dependence</td>
</tr>
<tr>
<td>Cocaine Abuse/Cocaine Dependence</td>
</tr>
<tr>
<td>Hallucinogen Abuse/Hallucinogen Dependence</td>
</tr>
<tr>
<td>Opioid Abuse/Opioid Dependence</td>
</tr>
<tr>
<td>Phencyclidine Abuse/Phencyclidine Dependence</td>
</tr>
<tr>
<td>Sedative, Hypnotic, or Anxiolytic Abuse/Sedative, Hypnotic, or Anxiolytic Dependence</td>
</tr>
<tr>
<td>Other (or Unknown) Substance Abuse/Other (or Unknown) Substance</td>
</tr>
<tr>
<td>Anxiety Disorders</td>
</tr>
<tr>
<td>Panic Disorder with Agoraphobia</td>
</tr>
<tr>
<td>Panic Disorder without Agoraphobia</td>
</tr>
<tr>
<td>Obsessive-Compulsive Disorder</td>
</tr>
<tr>
<td>Post-traumatic Stress Disorder</td>
</tr>
<tr>
<td>Anxiety Disorder Due to a GMC</td>
</tr>
<tr>
<td>Substance-Induced Anxiety Disorder</td>
</tr>
<tr>
<td>Anxiety Disorder NOS</td>
</tr>
</tbody>
</table>

(continues)
TABLE 6.2 (continued)

Adjustment Disorders
- Adjustment Disorder with Depressed Mood
- Adjustment Disorder with Anxiety
- Adjustment Disorder with Mixed Anxiety and Depressed Mood
- Adjustment Disorder with Disturbance of Conduct
- Adjustment Disorder with Mixed Disturbance of Emotions and Conduct
- Unspecified Adjustment Disorder

Disorders Covered in Summarized Fashion (without specific diagnostic criteria)
- Agoraphobia without History of Panic Disorder
- Social Phobia
- Specific Phobia
- Generalized Anxiety Disorder
- Somatization Disorder
- Undifferentiated Somatoform Disorder
- Hypochondriasis
- Body Dysmorphic Disorder
- Anorexia Nervosa
- Bulimia Nervosa

*NOS = Not Otherwise Specified; GMC = General Medical Condition.

The modular design of the SCID is a major strength of the instrument because administration can be customized easily to meet the unique needs of the user. For example, the SCID can be shortened or lengthened to include only those categories of interest, and the order of modules can be altered. The format and sequence of the SCID were designed to approximate the flowchart and decision trees followed by experienced diagnostic interviewers. The SCID begins with an open-ended overview portion, during which the development and history of the present psychological disturbance are elicited and tentative diagnostic hypotheses are generated. Then the SCID systematically presents modules that allow for assessment of specific disorders and symptoms. Most disorders are evaluated for two time periods: current (meets criteria for the past month) and lifetime (ever met criteria).

Consistent with its linkage with DSM-IV, formal diagnostic criteria are included in the SCID booklet, thus permitting interviewers to see the exact criteria to which the SCID questions pertain. This unique feature makes the SCID an excellent training device for clinicians because it facilitates the learning of diagnostic criteria and good questions to assess the criteria. The SCID has many open-ended prompts that encourage respondents to elaborate freely about their symptoms. At times, open-ended prompts are followed by closed-ended questions to fully clarify a particular symptom. Although the SCID provides structure to cover criteria for each disorder, its semistructured format provides significant latitude for interviewers to restate questions, ask for further
clarification, probe, and challenge if the initial prompt was misunderstood by 
the interviewee or clarification is needed to fully rate a symptom. SCID inter-
viewers are encouraged to use all sources of information about a respondent, 
and gentle challenging of the respondent is encouraged if discrepant information 
is suspected.

During administration, each symptom is rated as either absent (or below 
threshold) or present (and clinically significant). A question mark (?) denotes that 
inadequate information was obtained to code the symptom. The SCID flowchart 
instructs interviewers to “skip out” of a particular diagnostic section when essen-
tial symptoms are judged to be below threshold or absent. These skip-outs result 
in decreased time of administration as well as the skipping of items with no diag-
nostic significance. Administration of the SCID is usually completed in one 
session and typically takes from 45 to 90 minutes. Once administration is com-
pleted, all current and past disorders for which criteria are met are listed on a 
Diagnostic Summary sheet.

The SCID is optimally administered by trained clinicians who have knowl-
dge about psychopathology, DSM-IV criteria, and diagnostic interviewing. Due 
to the semistructured format of the SCID, proper administration often requires 
that interviewers restate or clarify questions in ways that are sometimes not 
clearly outlined in the manual in order to judge accurately if a particular diag-
nostic criterion has been met. The task requires that SCID assessors have a 
working knowledge of psychopathology and DSM-IV as well as basic inter-
viewing skills. Standard procedures for training to use the SCID include care-
fully reading the SCID Users Guide (First, Spitzer, Gibbon, & Williams, 1997b), 
reviewing the SCID administration booklet and score sheet, viewing SCID video-
tape training materials that are available from the SCID authors, and conducting 
role-played practice administrations with extensive feedback discussions. Next, 
trainees may administer the SCID to representative participants who are jointly 
rated so that a discussion about sources of disagreements can ensue. In research 
settings, a formal reliability study is advantageous. The reliability and validity of 
the SCID in adult populations with diverse disorders have been evaluated in a 
number of investigations, with generally excellent results among widely varied 
participant samples and experimental designs (see First & Gibbon, 2004; also 
Segal, Hersen & Van Hasselt, 1994).

Overall, the SCID is a widely used and respected assessment tool. It has been 
translated into 12 languages and has been applied successfully in research studies 
and clinical practice in many countries. Computer-assisted clinician-administered 
versions of the SCID-CV and SCID Research Version are available. A self-
administered computerized screening version of the SCID, called the SCID-
Screen-PQ, is also available, but it does not produce final diagnoses. Rather, 
likely diagnoses are further evaluated by a full SCID interview or a clinical eval-
uation. For more information on the SCID, the interested reader may visit the 
THE STRUCTURED CLINICAL INTERVIEW FOR DSM-IV
AXIS II PERSONALITY DISORDERS

To complement the Axis I version of the SCID, a version focusing on Axis II personality disorders according to DSM-IV has been developed, called the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II; First, Gibbon, Spitzer, Williams, & Benjamin, 1997). Instruments such as the SCID-II are particularly important because clinicians and researchers alike have struggled with their ability to accurately diagnose personality disorders and distinguish one personality disorder from another (Coolidge & Segal, 1998; Westen & Shedler, 2000). The SCID-II has a semistructured format similar to that of the SCID Axis I version, but it covers the 10 standard DSM-IV Axis II personality disorders as well as Depressive Personality Disorder and Passive-Aggressive Personality Disorder (which are listed as disorders to be studied further in an appendix of the DSM-IV).

For comprehensive assessment, the SCID-II may be easily used in conjunction with the Axis I SCID, which would be administered prior to personality disorder assessment. This is encouraged so that the respondent's present mental state can be considered when judging the accuracy of self-reported personality traits. The basic structure and conventions of the SCID-II closely resemble those of the SCID-I. One unique feature of the SCID-II is that it includes a self-report Personality Questionnaire, which is a 119-item screening component that can be administered prior to the interview portion and takes about 20 minutes. The purpose of the Personality Questionnaire is to reduce overall administration time, because only those items that are scored in the pathological direction are further evaluated during the structured interview portion.

During the structured interview component, the pathologically endorsed screening responses are further pursued to ascertain whether the symptoms are actually experienced at clinically significant levels. Here, the respondent is asked to elaborate about each suspected personality disorder criterion, and specified prompts are provided. Like the Axis I SCID, the DSM-IV diagnostic criteria are printed on the interview page for easy review, and responses are coded as follows: = inadequate information, 1 = absent or false, 2 = subthreshold, and 3 = threshold or true. Each personality disorder is assessed completely, and diagnoses are made before proceeding to the next disorder. The modular format permits researchers and clinicians to tailor the SCID-II to their specific needs and reduce administration time. Clinicians who administer the SCID-II are expected to use their clinical judgment to clarify responses, gently challenge inconsistencies, and ask for additional information as required to rate accurately each criterion. Collection of diagnostic information from ancillary sources is permitted. Complete administration of the SCID-II typically takes less than 1 hour.

Training requirements and interviewer qualifications are similar to that of the Axis I SCID. There is no clinician version of the SCID-II. The psychometric properties of the SCID-II are strong, and the interested reader is referred to First
and Gibbon (2004) for a comprehensive review. Given the extensive coverage of the personality disorders, modular approach, and strong operating characteristics, the SCID-II should remain a popular and effective tool for personality disorder assessment. The SCID-II Web site is the same as for the SCID Axis I version, www.scid4.org.

SPECIALIZED STRUCTURED AND SEMISTRUCTURED INTERVIEWS

In this section, we describe briefly a host of specialized interviews that are commonly used in clinical or research arenas.

STRUCTURED INTERVIEW OF REPORTED SYMPTOMS

The Structured Interview of Reported Symptoms (SIRS; Rogers, 1992) is a fully structured interview designed specifically for the assessment of feigning of mental disorders and related response styles. The SIRS by itself is an effective measure of feigning, although it is commonly paired with the Minnesota Multiphasic Personality Inventory (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) as a supplement. The SIRS may also be used when a suspected malingerer has declined to complete the MMPI or other psychological testing (Rogers, 2001). The SIRS has 172 items, and the output includes eight primary and five supplementary scales. Three primary scales examine very unusual symptom presentation in terms of rare symptoms, preposterous content, and atypical symptom pairs. Four primary scales look at the range and severity of symptom endorsements, including blatant symptoms of a major mental disorder, subtle symptoms more likely to be associated with minor psychological problems, indiscriminant reporting of symptoms, and overendorsement of symptoms with extreme severity. The final primary scale evaluates differences between self-report and observation during the interview (Rogers, 2001). The primary purpose of the SIRS is to precisely classify feigning and nonfeigning respondents, and it has its strongest application in forensic settings. Rogers (2001) recommends that the SIRS be used whenever the feigning of mental disorders is suspected.

EVALUATION OF COMPETENCY TO STAND TRIAL—REVISED

The Evaluation of Competency to Stand Trial—Revised (ECST-R; Rogers, Tillbrook, & Sewell, 2004) is a semistructured interview for assessing the underlying dimensions of competency to stand trial. It is also designed to screen
systematically for malingering and feigned incompetence. The ECST-R contains three competency scales that correspond to standards in the field: Consult with Counsel, Factual Understanding of Courtroom Proceedings, and Rational Understanding of Courtroom Proceedings. The interview was developed as a tool for the growing number of psychologists who work with the legal system. The ECST-R contains scales called the Atypical Presentation (ATP) scales, which are used to screen for malingering. There are five subscales of the ATP: ATP-P (psychotic subscale), ATP-N (nonpsychotic subscale), ATP-B (both the ATP-P and ATP-N subscales), ATP-I (impairment subscale), and ATP-R (realistic subscale). The ATP scales were found to be useful in the screening of malingering, and they provide ancillary data about feigned incompetence in cases where malingering has been established (Rogers et al., 2004).

**STRUCTURED INTERVIEW FOR THE FIVE-FACTOR MODEL OF PERSONALITY**

The Structured Interview for the Five-Factor Model of Personality (SIFFM; Trull et al., 1998) is a semistructured interview used to gauge the major dimensions of the famous Five-Factor Model (FFM) of personality: neuroticism, extraversion, openness, agreeableness, and conscientiousness. The SIFFM also provides assessment of the facets for each major dimension that are particularly important regarding differentiation among personality disorders. The SIFFM was modeled after the self-report NEO Personality Inventory Revised (NEO-PI-R) and comprises 120 items related to the FFM. The SIFFM is considered to be of use to clinical researchers whose interests are in assessing personality and personality disorders (Trull et al.).

**DIAGNOSTIC INTERVIEW FOR BORDERLINES—REVISED**

The Diagnostic Interview for Borderlines—Revised (DIB-R; Zanarini, Gunderson, Frankenburg, & Chauncey, 1989) is a semistructured interview used for evaluating and diagnosing Borderline Personality Disorder. Notably, the DIB-R was designed to enhance discrimination between clinically diagnosed borderline patients and the other personality disorders. The DIB-R has four distinct sections, each evaluating a core aspect of borderline personality: affect, cognition, impulse action patterns, and interpersonal relationships. In total, the interview includes 186 questions that make inquiries into 108 scored areas. In the DIB-R, all questions are geared to the preceding two years of the respondent's life. The measure has been found successful in discriminating those with Borderline Personality Disorder from those without (Zanarini et al.) and is a popular measure for in-depth assessment of the borderline construct.
The Diagnostic Interview Schedule—Hamilton Rating Scale for Depression (DIS-HRSD; Whisman et al., 1989) is a fully structured interview designed to measure depression severity and is used primarily in research settings. As the name implies, the DIS-HRSD is based on the widely used Hamilton Rating Scale for Depression (Hamilton, 1960), which is a clinician-administered rating scale for depressive symptoms. A drawback to the original Hamilton Rating Scale, however, is that Hamilton provided only general guidelines for administration and scoring, which can negatively impact the reliability of the instrument. In contrast, the fully structured DIS-HRSD clearly specifies administration and scoring and has the further advantage that it does not need to be administered by an experienced clinician.

The Acute Stress Disorder Interview (ASDI; Bryant, Harvey, Dang, & Sackville, 1998) is a structured interview designed to diagnose Acute Stress Disorder. The ASDI contains 19 items that are dichotomously scored based on criteria from the *DSM-IV*. The dichotomous scoring system is used for the following three reasons: to formally identify cases of Acute Stress Disorder, to allow the ASDI to be given after a disaster, and to allow nonclinicians to administer the ASDI. The ASDI appears to be a useful tool to identify those individuals who suffer from Acute Stress Disorder and who are at risk for long-term Post-Traumatic Stress Disorder. The measure is successful at discriminating those diagnosed with Acute Stress Disorder from those who do not meet the criteria for the disorder (Bryant et al.).

The Interview for the Diagnosis of Eating Disorders—IV (IDED-IV; Kutlesic, Williamson, Gleaves, Barbin, & Murphy-Eberenz, 1998). The IDED-IV is a semi-structured interview developed to adhere strictly to *DSM-IV* diagnostic criteria for anorexia nervosa, bulimia nervosa, binge eating disorder, and eating disorder not otherwise specified. Compared to earlier versions of the interview, the IDED-IV includes more inquiries about demographic information and eating disorder history, expanded instructions for administration of the interview, and condensed ratings from a seven- to a five-point rating scale (Kutlesic et al.).

The Eating Disorder Examination (EDE; Cooper & Fairburn, 1987) is a semi-structured interview constructed to evaluate the broad range of eating disorder
psychopathology, including assessment of people who have concerns about their body shape and weight. The interview consists of 62 items and requires brief training for administration. The EDE has been widely used in studies of eating disorder psychopathology as well in outcome studies in the area (Cooper & Fairburn).

ANXIETY DISORDERS INTERVIEW SCHEDULE FOR DSM-IV

The Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV; Brown, Di Nardo, & Barlow, 1994) is a semistructured interview that focuses primarily on the anxiety disorders but also provides assessment of some mood, substance, and somatoform disorders. The ADIS-IV focuses on current diagnoses only. A modification called the ADIS-IV Lifetime version (Di Nardo, Brown, & Barlow, 1994) provides assessment of current and lifetime disorders as well as a diagnostic timeline that promotes accurate determination of the onset, remission, and temporal sequence of the disorders. A strength of the ADIS-IV is that it provides both dimensional and binary evaluation of the disorders. Major segments of the ADIS-IV and ADIS-IV-L consist of demographics, inquiries about the presenting complaint and stressors, diagnostic sections for DSM-IV anxiety disorders, mood disorders, somatoform disorders, substance-related disorders, a section of screening questions for other major disorders, questions regarding family history of psychological disorders, medical history, and psychosocial treatment history, the Hamilton rating scales, and a diagnostic timeline (in the ADIS-IV-L only).

SUMMARY AND CONCLUSIONS

This chapter highlights the fact that structured and semistructured interviews have greatly facilitated the task of differential diagnosis and problem clarification in diverse clinical and research settings. The reliability of diagnosis is much improved with the use of structured interviews, compared to the unstandardized approach common in clinical practice, and improved reliability provides the foundation for enhanced validity or meaningfulness of diagnosis. Given the field's recent emphasis on empirically supported psychotherapeutic interventions and processes (Beutler & Castonguay, in press; Kendall, 1998), we hope that a concomitant focus on clinically relevant, standardized, and validated assessment procedures will be realized as well. Structured and semistructured interviews can and should play an important role in the advancement of the science of clinical psychology.
REFERENCES


INTRODUCTION

It has been around 20 years since the last comprehensive chapter was written that focused exclusively on self-monitoring (Bornstein, Hamilton, \& Bornstein, 1986). Since that chapter was published in 1986, there have been many new applications and technological developments about this behavioral assessment procedure. Over the years, numerous terms have been used to describe the process of systematically observing and recording one's own ongoing behavior (i.e., self-assessment, self-monitoring, self-observation, self-recording). Some authors have suggested that the term self-monitoring be replaced with the term self-observation in order to distinguish it from the personality construct of self-monitoring (Cone, 1999; Jackson, 1999). For historical continuity, however, the terms self-monitoring and self-assessment will be used interchangeably in this chapter.

Self-assessment continues to be one of the most commonly used assessment methods in clinical research and practice. Since the early 1960s, research has provided considerable empirical support for the continued utilization of this assessment method. Self-assessment has been used to help individuals monitor the frequency, antecedents, and consequences of problematic behaviors (i.e., behaviors that need to decrease or increase). This type of information greatly contributes to the understanding and formulation of the functional analysis approach to problematic behaviors (Haynes, 2000). In some of the earlier studies that focused on this assessment technique, researchers found an added benefit as well as a potential problem with the self-monitoring process; individuals reacted to the process by decreasing the baseline frequency of behaviors just by observing...
their own behavior (e.g., Ciminero, Nelson, & Lipinski, 1977). Since this finding, self-monitoring has become an integral component of most empirically supported treatments, with the explicit purpose of increasing sensitivity to behavior occurrence, increasing awareness of the frequency of the behavior, and then providing techniques to decrease/increase target behaviors even further (Korotitsch & Nelson-Gray, 1999).

An increase in self-assessment research and use coincided with the ascent of the behavior therapy movement in the 1960s and 1970s. Because of behavior therapy's commitment to objective assessment, research began focusing on assessment techniques that relied less on inference. It is interesting to note that this quest for more objective assessment procedures paralleled early attempts in psychology to teach individuals how to be better introspectionists (Kazdin, 1974). Early research on behavioral assessment reflected its roots in behavior therapy; examining behavior change in the client's environment, focusing on self-control in addition to external control, and an increasing interest in assessing private events (e.g., thoughts, urges, images, fantasies; Ciminero et al., 1977). Innovative techniques were devised with an emphasis on an idiographic approach to assessment that did not appear to be subject to evaluation by classic psychometric methods.

Because of behavior therapy's focus on the situational specificity of behavior and less reliance on indirect and inferential methods of assessment, behavioral assessment techniques were not evaluated with regard to classic psychometric principles (Goldstein & Hersen, 1990). Goldstein and Hersen remarked that such lack of attention to psychometrics in the past resulted in the "baby thrown out with the bath water" (p. 11) but noted that more recently the "baby is being returned from the discarded bath water" (p. 12). In contrast to early researchers, who proposed that behavioral assessment techniques should only be evaluated functionally (e.g., Nelson, 1983), recent articles have revisited the applicability of psychometrics to self-monitoring (Cone, 1999; Jackson, 1999).

Several authors have conceptualized self-monitoring, in addition to other behavioral assessment measures, as falling along continuums of directedness (Cone, 1999) and convenience (Korotitsch & Nelson-Gray, 1999). When compared to methods of self-report (e.g., questionnaires, interviews) that fall toward the indirect end of the directness continuum, self-monitoring falls closer to the direct end (Cone). For example, in self-monitoring, an individual observes and records a target behavior in a certain place within close proximity in time to its occurrence. In contrast, on self-report measures, an individual reports behaviors that occurred at another time and another place. Korotitsch and Nelson-Gray also note that behavioral assessment measures can be placed along a continuum of convenience. In terms of convenience, the foregoing behavioral assessment methods appear to switch places along the continuum. Indirect measures (e.g., self-report on questionnaires and interviews) are found at the most convenient end of the spectrum because they are easy to administer and can be completed relatively quickly. On the other hand, self-monitoring methods can fall at
different points of the convenience spectrum, depending on the complexity and expense of the method utilized.

In order to self-monitor, an individual engages in a multicomponent process (e.g., Cone, 1999; Korotitsch & Nelson-Gray, 1999). First, the individual must be aware that the behavior occurred (i.e., the client discriminates whether or not a response has occurred; Bornstein et al., 1986). For example, if an individual will be monitoring the frequency of binge-eating episodes, then he or she should be clear about what constitutes a binge-eating episode (i.e., objective versus subjective binge). In the second part of the process, the individual records occurrences of the behavior, which could also include other relevant information (e.g., antecedents, consequences, contextual variables). Korotitsch and Nelson-Gray propose that both aspects of the process are needed for accurate data collection. In some contexts, it may also be helpful for the individual to analyze the collected data and display the results (Cone).

**POTENTIAL USES OF SELF-ASSESSMENT**

Several reasons contribute to the widespread use of self-assessment in clinical research and practice. First, self-monitoring represents a relatively inexpensive and efficient way to gather information on target behaviors (Korotitsch & Nelson-Gray, 1999). In its simplest form, self-monitoring may only require paper and pencil, making it one of the most portable, practical, and cost efficient means of assessment (e.g., Baird & Nelson-Gray, 1999; Bornstein et al., 1986). Because of the nature of some target behaviors, it may be impractical (e.g., target behavior may be obsessions, sexual fantasies) or too costly to employ direct observation methods by independent observers. Although self-monitoring can have reactive effects (discussed in a later section), it eliminates external observer bias. In addition, some target behaviors may be observable only by the client (e.g., thoughts, images).

Second, self-monitoring can be used for clients with a wide range of psychological and medical issues, such as eating disorders (e.g., Latner & Wilson, 2002; Wilson & Vitousek, 1999), HIV-related sexual behavior (Weinhardt, Forsyth, Carey, Jaworski, & Durant, 1998), frontal lobe dysfunction (Alderman, Fry, & Youngson, 1995), chronic pain (e.g., Kerns, Finn, & Haythornthwaite, 1988), insomnia (e.g., Savard & Morin, 2002), anxiety disorders (e.g., Campbell & Brown, 2002; Craske & Tsao, 1999; Rapee, Craske, & Barlow, 1990), and substance use (e.g., Tucker, Vuchinich, & Murphy, 2002). In fact, Korotitsch and Nelson-Gray (1999) assert that “currently, self-monitoring procedures are described and recommended within most empirically supported treatments” (p. 415).

Third, self-monitoring allows clients to become more aware of their behaviors (e.g., Haynes, 1978; Wilson & Vitousek, 1999). Through the self-monitoring process, a client can receive continuous and immediate feedback
about target behaviors that occur outside of the therapeutic setting. Similarly, such data can reveal behavioral patterns that may not be readily apparent to therapists and clients. For example, an individual with binge-eating disorder may not be aware of the antecedents that precede a binge-eating episode. However, after continuously monitoring food intake for several weeks, the client and therapist can review the monitoring forms to recognize common patterns (e.g., the client always binge-eats after work on days when she does not eat breakfast).

Fourth, self-monitoring represents the only way to access private events (e.g., thoughts, emotions, perceptions) and behaviors that do not typically occur in observable conditions (e.g., drug use, sexual activity) and that are inaccessible to outside observers (Baird & Nelson-Gray, 1999; Foster, Laverty-Finch, Gizzo, & Osantowski, 1999). For example, one component of cognitive-behavioral therapy for depression involves having a client record antecedents (e.g., employer does not speak to the client) and consequences of a problematic situation (e.g., negative mood, fails to get a task done on time) in which negative, automatic thoughts are occurring (e.g., "my boss must hate me," "I am such a loser"). The recording of these types of situations and thoughts can help the client and therapist focus on recurring cognitive schemas that can be challenged during treatment.

Finally, in an age when managed care requires accountability, self-monitoring provides a valuable method for keeping track of client progress (Korotitsch & Nelson-Gray, 1999; Strosahl & Robinson, 2004). Surveys have estimated that self-monitoring is used between 38% (Bornstein, Bridgwater, Hickey, & Sweeney, 1980) and 44% of the time (Elliott, Miltenberger, Kaster-Bundgaard, & Lumley, 1996) in clinical research. It is not known, however, how often self-monitoring methods are used by practitioners to gather assessment data. Although use of self-assessment methods continues to proliferate, research on psychometric applicability and its clinical utility have not experienced a paralleled increase (Cone, 1999).

FUNCTIONS OF SELF-ASSESSMENT


EDUCATION/DESCRIPTION

Because of its diverse applicability and ease of use, self-monitoring has been used to gain a better understanding of target behaviors, antecedents, and consequences (e.g., Cone, 1998; Haynes, 1978; Jackson, 1999). For example, 97 patients diagnosed with Panic Disorder with Agoraphobia (PDA) self-monitored
their panic attacks for a period of 6–12 weeks in order to gain detailed information about precipitating circumstances and attack symptomatology (Garssen, de Beurs, Buikhuizen, van Balkom, Lange, & van Dyck, 1996). Using self-monitoring techniques, the researchers were able to learn more about the variety and frequency of symptoms that individuals diagnosed with PDA experience both before and during panic attacks. This type of information contributes to a better understanding of the impact that behaviors may have across settings, behavior frequency and disruption, and intensity of the responses.

**DIAGNOSTIC CLARIFICATION**

Self-monitoring can be used to clarify that a diagnosis is warranted for a particular individual (Korotitsch & Nelson-Gray, 1999). In the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR*, American Psychiatric Association [APA], 2000) psychological disorders (e.g., Bipolar II Disorder, Generalized Anxiety Disorder, Premenstrual Dysphoric Disorder) have specific criteria that must be present (or absent) for a diagnosis to be made. Making an accurate diagnosis can be difficult for numerous reasons (e.g., the client may have little awareness of problematic behaviors, it is unclear whether or not the presenting symptoms occur for a medical condition or another Axis I disorder). In such cases, self-monitoring may be useful to help determine the most accurate diagnosis. For example, to diagnose an individual with Premenstrual Dysphoric Disorder (PMDD), the *DSM-IV* requires that the presenting symptoms not be an exacerbation of symptoms for another disorder and also that the cyclical pattern of symptoms occur for two consecutive months (APA). More importantly, the *DSM-IV* requires that the presence of symptoms be confirmed by daily self-monitoring for two consecutive months (APA).

Because panic attacks can occur in the context of other anxiety disorders (e.g., Social Phobia, Specific Phobia), it is important to clarify the context in which the attacks occur. Verifying the functional reason for the panic attacks by self-monitoring could be useful in clarifying the relations between the panic attacks and the context in which they occur in order to make a more accurate diagnosis. For example, after the client self-monitors for a certain period of time, the therapist and client may discover that the client only experiences panic attacks when he or she speaks in public. In this situation, a diagnosis of Social Phobia would be more accurate and might lead to a more effective treatment. Self-monitoring data may also be helpful when clients present with vague symptoms or goals (e.g., Korotitsch & Nelson-Gray, 1999). In general, self-monitoring appears to be a useful technique for clarifying diagnoses.

**PRETREATMENT ASSESSMENT**

Overall, self-assessment can be useful in identifying target behaviors, their antecedents, and their controlling variables. Self-monitoring is often used in
behavioral assessment to obtain baseline information about target behaviors prior to an intervention (Ciminero et al., 1977; Korotitsch & Nelson-Gray, 1999). Clinicians utilize self-monitoring as part of a thorough functional analysis, with the aim of determining the frequency and intensity of identified target behaviors (Haynes, 1978). In addition, self-monitoring helps identify antecedents and consequences that functionally relate to the target behaviors (Ciminero et al., 1977). For example, an individual with trichotillimania might be asked to monitor frequency of hair pulling and also indicate the situation, time, and mood state when hair pulling occurs. The purpose of the self-assessment would be to determine which events are functionally related to hair pulling. In this example, it may be that the client engages in more frequent hair pulling when she is watching television and reports feeling bored.

Information gained through self-monitoring represents a valuable addition to information that might be obtained through clinical interviews or self-report questionnaires. Instead of relying solely on the client’s ability to accurately remember and retrospectively report how often behaviors took place and the relevant contextual variables, self-monitoring allows the clinician and client to recognize the severity and impact of the client’s problems as they take place in real time (Cone, 1999). In addition, self-monitoring may help determine which behaviors covary and their controlling variables in order to test out hypotheses about their relations (Haynes, Leisen, & Blaine, 1997). During preintervention assessment, self-monitoring illustrates the pragmatic principles grounded in behavioral assessment; the behaviors and controlling variables identified help guide the formulation of a treatment plan (Korotitsch & Nelson-Gray, 1999).

**TREATMENT EVALUATION**

Self-monitoring is used in behavioral assessment to monitor the effects of ongoing treatment (Ciminero et al., 1977). In particular, a client will continue to self-monitor throughout treatment so that the information collected during the treatment phase can be compared to the client’s baseline data. Thus, self-monitoring provides an opportunity for the therapist and client to continuously assess the client’s progress instead of waiting until the end of therapy to assess whether or not the treatment was effective (Korotitsch & Nelson-Gray, 1999). The therapist (along with the client) can use the information collected to guide decisions regarding the utility of the treatment strategies that are selected.

For example, if the client appears to be making progress, the therapist can continue with the existing treatment (e.g., cognitive behavioral treatments [CBT] for bulimia nervosa). If the client is making few, if any, gains, then the therapist might decide to modify the current treatment (e.g., rely more on the behavioral components of CBT). However, if the client’s symptoms appear to be worsening or if the treatment appears to have no effect, then the therapist might decide to discontinue the current treatment and implement a new treatment approach (e.g., stop CBT and begin interpersonal psychotherapy).
The information collected during self-monitoring can also guide the focus of individual treatment sessions. In a review of self-monitoring and eating disorders, Wilson and Vitousek (1999) acknowledge that self-monitoring represents the most important component of CBT when treating individuals diagnosed with an eating disorder. The authors mention that “each session begins with a review of the patient’s self-monitoring records. The agenda for each session is based on each patient’s response to the previous therapy session” (p. 484). Taken as a whole, continuous self-monitoring during the treatment phase provides essential information to monitor and evaluate the effects of ongoing treatment.

**SELF-ASSESSMENT AS AN INTERVENTION**

Self-monitoring can be used to facilitate changes in an individual’s behavior (Haynes, 1978). A change in a target behavior’s frequency, after an individual begins self-monitoring, is defined as the reactive effects of self-monitoring (Baird & Nelson-Gray, 1999). In general, the change in behavior frequency usually occurs in a therapeutically beneficial direction (i.e., there is a decrease in problematic behaviors and/or an increase in more appropriate behaviors; Haynes). This phenomenon has been found in numerous studies looking at clinically relevant behaviors, such as cigarette smoking (Frederiksen, Epstein, & Kosevsky, 1975), suicidal ideation (Clum & Curtin, 1993), and tics (Thomas, Abrams, & Johnson, 1971). For example, 30 women diagnosed with either bulimia nervosa or Binge Eating Disorder self-monitored their daily food intake for a period of 6–18 days. Results indicated a statistically significant reduction in the frequency of binge-eating episodes over the self-assessment period (Latner & Wilson, 2002). In fact, results indicated that the frequency of binge-eating episodes recorded during the self-monitoring phase decreased by more than half (44%) as compared to those reported during the initial interview (Latner & Wilson).

Despite initial benefits in treatment, some clinicians and researchers have found reactive effects in self-monitoring to be short-lived (e.g., Critchfield & Vargas, 1991; Febbraro & Clum, 1998). Self-monitoring may, however, represent a useful strategy to implement at the beginning of treatment to facilitate early change in the client’s behavior. This information could increase the probability that a client will continue treatment; however, additional treatment approaches should be implemented throughout therapy so that the client can continue to improve (Korotitsch & Nelson-Gray, 1999). For example, individuals who consistently self-monitored food consumption during a long-term weight-loss program (People at Risk) lost more weight than individuals who did not continue to self-monitor (Baker & Kirschenbaum, 1993). This type of finding has led weight-loss researchers to conclude that successful weight-loss programs must have consistent and continuous self-monitoring as a central component (e.g., Wadden et al., 1997).

Although the potential reactive side effects of self-monitoring are typically advantageous for treatment purposes, these effects may pose a problem for assess-
ment of pretreatment target behaviors (Jackson, 1999). Researchers interested in obtaining data for descriptive purposes (e.g., comparing the number of negative automatic thoughts between individuals diagnosed with depression and a comparison group) may have individuals self-monitor the targeted behavior in an attempt to make comparisons between groups. However, self-monitoring, due to its reactive effects, may actually minimize the differences found between the groups. Similarly, reactive effects of self-monitoring may decrease the differences found when comparing an individual's baseline data with post-treatment data. Due to reactivity effects, information collected at baseline may not accurately reflect the frequency of the client's behavior before self-monitoring began. In summary, self-monitoring can sometimes facilitate behavior changes in an individual's behavior, and, depending on the clinician's or researcher's goals, the changes may be considered favorable.

METHODOLOGICAL CONSIDERATIONS

When deciding to use self-assessment methods, a number of issues needs to be considered before choosing a particular technique. In general, the therapist or researcher needs to determine the purpose for self-monitoring, the nature of the target behavior, the method that will be used to gather information, and the type of monitoring device to be used. The following sections will describe various methodological components involved in self-monitoring.

ASPECTS OF THE TARGET BEHAVIOR

Frequency Counts

Frequency counts are commonly used when a therapist or researcher is interested in the rate of a target behavior and asks the client to record each occurrence of the behavior (Ciminero et al., 1977). For example, a client may be asked to record the number of cigarettes that he or she smokes each day. Frequency counts are useful when the self-monitored behavior occurs at a low-to-medium frequency, when the behavior is a distinct event, and when it is important to understand the number of times the individual engages in the behavior (Foster et al., 1999).

When self-monitoring frequency counts, it is also important for the behavior to have a clear beginning and end so that separate instances of the behavior can be recorded (Foster et al., 1999). Similarly, it would be difficult (and potentially misleading) to monitor frequency counts for ongoing behaviors such as talking and smiling. For example, if a client with social skills deficits engages in a conversation for 30 seconds with one individual and then with another individual for 20 minutes, these occurrences would only be counted as two instances of social interaction. However, each conversation lasted for a different duration, so impor-
tant information would be lost if the clinician or researcher relied solely on frequency counts. In addition, the client may become frustrated with the amount of time spent trying to self-monitor high-frequency behaviors (e.g., recording the number of negative thoughts in one day).

There are several advantages for self-monitoring frequency counts. First, frequency counts are relatively simple to record and instructions for the process are typically easy to understand (Foster et al., 1999). Second, frequency conveys a direct measure regarding the amount of times a behavior occurs, and many behavioral treatment goals attempt to increase (or decrease) certain behaviors (e.g., Bellack & Schwartz, 1976). Third, frequency measures reflect changes over time, providing an opportunity for the clinician and client to have continuous, ongoing feedback regarding the client’s progress.

**Duration**

Duration represents another dimension of a target behavior that an individual can self-monitor (Baird & Nelson-Gray, 1999). Duration methods require an individual to self-monitor the amount of time a target behavior occurs (Foster et al., 1999). For example, an individual who engages in compulsive behaviors can self-monitor the amount of time spent cleaning his or her house. Duration is a particularly useful method for measuring continuous, ongoing responses that can vary in length (Ciminero et al., 1977). Therefore, it is important to clearly define the start point and endpoint of a response. Duration is also beneficial when the length of time spent engaging in a behavior is more important than the number of times the behavior takes place (Foster et al.).

For example, an individual diagnosed with Obsessive-Compulsive Disorder may engage in excessive, lengthy hand-washing rituals that limit his or her ability to engage in other activities. Therapeutically, the amount of time this individual spends hand washing (e.g., five hours per day) is more of a concern than the rate of hand washing (e.g., three times a day). In a similar vein, researchers compared the amount of time individuals diagnosed with Generalized Anxiety Disorder (GAD) and non-GAD controls spent worrying (Dupuy, Beaudoin, Rhéaume, Ladouceur, & Dugas, 2001). Each participant recorded the number of minutes spent worrying daily, for two weeks, in a self-monitoring notebook. The researchers concluded that self-monitoring the duration of time spent worrying was a valid and helpful measure of worry, especially given that the results significantly correlated with the current gold standard measure for worry, the *Penn State Worry Questionnaire* (Dupuy et al.).

**Latency**

Latency, the amount of time it takes an individual to begin the target behavior, represents a third aspect of behavior commonly recorded with self-monitoring methods (Foster et al., 1999). Latency can be conceptualized as the amount of time elapsed between some beginning point or antecedent and the target behavior (Foster et al.). For example, latency would be very relevant for an individual who
is self-monitoring aspects of his or her insomnia. The beginning point would be when the client lays down in bed, and the endpoint or target behavior would be when the client falls asleep. Similar to duration, latency needs to have a clearly defined beginning point and endpoint.

Intensity

A fourth dimension of behavior recorded with self-monitoring methods is intensity (Cone, 1999). Measures of intensity require the individual to record the magnitude or strength of a behavior, often using a global subjective rating (Foster et al., 1999). For example, an individual could self-monitor the intensity of his or her depressed mood using a 10-point Likert scale ranging from 1 ("not at all") to 10 ("extremely depressed"). Foster and colleagues suggest that it is important to work with the client to develop unambiguous anchors (e.g., past experiences) for each point on the scale to maintain client consistency over time. In another example, it may be useful to monitor the intensity of an anger outburst; one very intense anger outburst (e.g., screaming, hitting others, and breaking things) could have more impact on an individual's social and occupational relations (e.g., friends, relatives, coworkers) than more relatively minor outbursts (e.g., slamming a door, walking off). In this situation, the magnitude of the target behavior may be more important than the number of times the behavior takes place.

SELF-ASSESSMENT RECORDING FORMATS

After determining which dimension of the target behavior will be the focus of self-assessment, the next step involves a decision regarding the best way to record the information. Various recording formats have been used to obtain data for self-monitoring and are presented next.

Event Sampling

Event sampling represents a procedure used to capture each time the behavior occurs throughout a specified time period (e.g., five hours, one day, two weeks). This format is typically helpful when an individual needs to monitor only one low-frequency behavior (Foster et al., 1999). For example, an individual records each instance of self-abuse (e.g., cutting) for a period of one week.

Interval Recording

Interval recording represents another format used to capture data for self-monitoring (Cone, 1999). With this strategy, an individual self-monitors his or her target behaviors for a single block of time that is further divided into a series of smaller intervals (Foster et al., 1999). During each interval, the individual records whether or not the target behavior occurred. It is important to note that the individual only records whether or not the behavior has occurred; if there are multiple instances of the target behavior within the same interval, they are not recorded separately. Chronic mood states, anxiety symptoms, muscle tension
ratings, and intrusive thoughts are examples of behaviors that may be recorded by intervals.

**Momentary Time Sampling**

Self-monitoring of target behaviors can also be recorded with a format known as momentary time sampling (Cone, 1999). Momentary time sampling involves having an individual record a behavior each time he or she receives a signal or cue (e.g., tone, vibration; Foster et al., 1999). For example, electronic signals are periodically sent to a client on a handheld computer as a cue to record his or her level of social anxiety at that moment in time. Time-sampling formats have the benefit of varying different aspects of the procedure, such as frequency of signals, duration of the time frame, and signal patterns (Foster et al.).

Overall, multiple formats (e.g., event sampling, interval recording, time sampling) can be used to gather self-monitoring information. The clinician or researcher interested in using self-monitoring should take into consideration the same issues involved in determining the aspect of the target behavior being monitored.

**SELF-ASSESSMENT DEVICES**

The final methodological consideration involves the selection of a self-monitoring device. According to Ciminero and colleagues (1977), clinicians and researchers should consider the following device characteristics before making a final decision: portability (i.e., manageable, handy), simplicity (i.e., user-friendly), cost (i.e., economical), and obtrusiveness (i.e., noticeable to individual, not others). Ideally, the authors also state, the target behavior should be recorded in real time (Ciminero et al.). The following devices are routinely used for self-monitoring of target behaviors.

**Paper and Pencil**

Paper-and-pencil techniques allow an individual to record information about target behavior rates and relevant controlling variables (Ciminero et al., 1977). For example, an individual diagnosed with Binge Eating Disorder can record the number of binge-eating episodes in addition to the situations that preceded the binge and the consequences that occurred following the binge episode. Diaries are commonly used paper-and-pencil devices that allow individuals to provide more detailed information regarding the target behavior and its controlling variables (Baird & Nelson-Gray, 1999). In general, paper-and-pencil devices are cost effective and manageable (Foster et al., 1999).

**Mechanical Devices**

Mechanical devices have also been used to gather information regarding frequency of the target behavior (Ciminero et al., 1977). Golf-stroke counters and belt/wrist counters are examples of mechanical devices. An individual diagnosed
with agoraphobia might self-monitor the number of panic attacks experienced in a particular situation (e.g., a mall) using a mechanical device. In this situation, the individual would only click the counter each time he or she experienced a panic attack. Typically, mechanical counters are user-friendly, inexpensive, and suitably noticeable to only the individual self-monitoring (Foster et al., 1999).

Timing devices are often used in self-monitoring to collect information about the duration of a single target behavior (Ciminero et al., 1977). For example, an individual can use a stopwatch or kitchen timer to measure the amount of time spent worrying. However, this method does not allow for information to be recorded about potential controlling variables (Foster et al., 1999). Despite this limitation, timing devices are generally cost effective, simple to use, and unobtrusive (Haynes, 1978).

**Electronic Devices**

Electronic devices are becoming a popular method for self-monitoring. Tape recorders, video cameras, cellular phones, and pagers (e.g., Johnson & Larson, 1982) have been used for self-monitoring purposes. Johnson and Larson used the experience-sampling method (ESM) to compare 15 normal-weight individuals diagnosed with bulimia nervosa to 24 controls. In this study, participants carried a pager for one week and completed self-monitoring forms when they were signaled by a pager at least once within a 2-hour period between 8:00 AM and 10:00 PM throughout the day. Individuals from both groups completed a total 2789 forms (bulimia nervosa group: 627 total, 44.8 per person; control group: 2117 total, 46.5 per person). Results from the study indicate that individuals diagnosed with bulimia nervosa experienced more negative moods, spent greater time alone, and were more preoccupied with thinking about, preparing, or eating food. The researchers concluded that ESM was a useful procedure for obtaining information about the binge–purge cycle and also the behaviors, thoughts, and feelings experienced by individuals diagnosed with bulimia nervosa.

Over the past 15 years, other instruments have expanded the scope of behaviors that can be used in self-assessment. Pedometers are often used to assess how many steps have been taken. Actigraphs have been utilized to look at activity patterns during wake-and-sleep cycles. In behavioral medicine, ambulatory monitoring of physiological activity, blood pressure monitoring, blood glucose monitoring, and peak air flow and respiration monitoring have greatly increased the awareness of the relation between environmental events and internal responses (Haynes, 1998).

**Computers**

Recently, the use of computers in behavioral assessment has greatly increased. As a result, more opportunities for self-assessment are available. Computer-assisted data collection is only as limited as the individual’s expertise and creativity in devising a method. These methods can be cost effective if individuals self-monitor behaviors on personal digital assistants (PDAs) (e.g., Norton,
Wonderlich, Myers, Mitchell, & Crosby, 2003) or handheld computers. Such computers can be programmed to cue an individual when it is time to record (e.g., time sampling), or frequency counts can be recorded. These types of data can be downloaded and summarized for the therapist and client. Another advantage of handheld computers is that they will likely reduce the error variance associated with other self-monitoring devices (e.g., Haynes, 1998).

Prospective data can be recorded and a more accurate sample of behavioral events in real time can be captured with computer recordings. For example, Taylor, Fried, and Kenardy (1990) had individuals with Panic Disorder record the frequency of panic attacks, the setting, and the thoughts they experienced surrounding the attacks. Similarly, Agras, Taylor, Feldman, Losch, and Burnett (1990) had obese clients monitor the amount of food intake, their weight, and the goals and amount of exercise. These devices are portable and appear to be less obtrusive in the client's environment (e.g., high numbers of individuals currently carry cell phones and pagers). Receiving a beep on a device to record may be less noticeable now than it might have been 20 years ago.

In summary, numerous devices are available for recording self-monitoring information. Although written almost 30 years ago, suggestions for choosing self-monitoring devices are still relevant (Ciminero et al., 1977). First, the device needs to be relatively manageable. Second, the device needs to be simple to use. Third, the device needs to be relatively cost effective. Fourth, the device should be sufficiently noticeable to the individual engaging in the self-monitoring, however barely discernible to surrounding individuals.

**REACTIVITY**

When self-monitoring is used for assessment purposes, therapists and researchers may want to minimize reactivity in order to get more accurate baseline information (Bornstein et al., 1986). When self-monitoring is used as an intervention, the researcher and clinician may want to maximize the reactive effects. In order to gain a better understanding of what effects occur under what circumstances for which subjects and with what type of self-monitoring methods, more research needs to be conducted (e.g., Korotitsch & Nelson-Gray, 1999). The following variables have been identified that can contribute to reactivity in self-monitoring: salience and cue value of the discriminative stimulus for recording; awareness of checks on accuracy of self-monitoring; training to conduct self-monitoring; frequency and timing of self-monitoring; motivation and expectations of the client; the nature of the target behavior (positive and nonverbal behaviors may be more reactive); valence of the target behavior; concurrent demands on the client in addition to self-monitoring; consequences for compliance with self-monitoring; and maintenance of reactive effects (Korotitsch & Nelson-Gray). Thus, maximizing or minimizing reactivity effects will be guided by the specific assessment and intervention goals of self-monitoring.
Whether or not classic psychometric principles can and should be used to evaluate behavioral assessment remains a debated issue (e.g., Cone, 1998; Jackson, 1999; Nelson, 1983). Haynes (1990) proposed that behavioral assessment procedures be evaluated on "applicability and utility, reliability, validity, and sources of error variance" (p. 442). However, the author also notes that there are difficulties with evaluating behavioral assessment with the principles underlying psychometric theory. Behavioral assessment is based on the notion that behaviors are not stable; they are situation specific. Historically these methods, in particular self-monitoring, have only been evaluated with regard to face validity and accuracy.

Jackson (1999) presents a number of issues that should be considered if psychometrics are applied to behavioral assessment procedures. According to the author, assessment procedures should be evaluated with regard to whether or not the procedure fulfills its intended purpose. In the case of self-monitoring, a decision must first be made about what is being measured. For example, it is important to distinguish conceptually what is being assessed. Are clients self-monitoring behaviors or response classes (whether observable or private) in real time? Or are clients being asked about hypothetical constructs or abilities/skills that are less likely to be affected by time (e.g., trait anxiety, coping skills, problem-solving skills)?

The concepts of accuracy and content validity may be more applicable to the evaluation of information obtained by self-monitoring of behaviors and response classes in real time. For hypothetical constructs or ability/skill monitoring, the psychometric concepts of reliability (e.g., stability over time), content validity, discriminant validity, convergent validity, concurrent validity, and predictive validity are more relevant for evaluation (Jackson, 1999). However, it is difficult to evaluate the accuracy of self-monitored private events when these events are not accessible to independent observers or to external means of recording these behaviors. For a more in-depth discussion of the applicability of psychometric principles to self-monitoring, readers are encouraged to consult the following excellent resources: Cone, 1995, 1998, 1999; Haynes, 1998; and Jackson. The following section focuses on behaviors that are observable by independent means with regard to accuracy.

**ACCURACY**

Principles of behavioral assessment are guided by a functional approach (i.e., how well the measure guides decision-making processes and treatment goals). Thus, the role of self-monitoring is not just to describe behavior but also to provide a means to assess the functions of the behavior and its maintaining variables. In addition, like other measures used in behavioral assessment, self-monitoring methods need to be supported by empirically validated data to ensure
SELF-ASSESSMENT

that what it intends to measure approximates the real behavior (Jackson, 1999). Therefore accuracy (i.e., the extent to which data reflect the true properties of the behavior) is an important psychometric concept for evaluating self-monitoring procedures (Baird & Nelson-Gray, 1999; Ciminero et al., 1977).

According to Korotitsch and Nelson-Gray (1999), three methods are commonly used to assess accuracy in self-monitoring. First, agreement comparisons can be made between the individual who is self-monitoring and an independent external observer. For example, high agreement was obtained between external observers and individuals self-monitoring of swimming laps (e.g., McKenzie & Rushall, 1974), number of cigarettes smoked (e.g., Frederiksen et al., 1975), and face-touching episodes (Nelson, Lipinski, & Black, 1976). However, in some studies, external observers may also contribute to self-monitoring reactivity. For example, Green (1978) compared temporal monitoring patterns (e.g., conducting monitoring before or after eating) of obese women to external observers. Results indicated that although temporal monitoring patterns did not lead to significant weight reduction alone, reactivity to an observer did contribute significantly to weight loss. In general, to achieve maximum accuracy, an observer should use the same recording procedure to record the identical target behavior (i.e., same operational definition) that the individual is self-monitoring. In addition, an independent observer should be recording information while the individual is self-monitoring (Jackson, 1999).

Second, data collected with self-monitoring can be compared to data collected by mechanical devices. In this situation, a mechanical device would be used to independently record each behavior occurrence, and the information obtained would be compared to the data collected from the individual's self-monitoring. For example, individuals with insomnia were asked to keep sleep logs (e.g., self-record sleep latency, sleep time, number of wakening episodes during the night), while polysomnography recordings were obtained in a laboratory setting (Coates, Killen, George et al., 1982). Results indicated that the self-recorded sleep log information positively correlated (although not perfectly) with the objective measures of total sleep time and latency to sleep obtained by polysomnography. Although self-assessment methods may not always be as accurate as more objective measures (e.g., mechanical devices), therapists and researchers often cite cost effectiveness, ease of use, and convenience as primary reasons for the continued use of self-monitoring methods (e.g., Barton, Blanchard, & Veazey, 1999).

Third, self-monitoring data can be compared to naturally occurring consequences of the behavior. For example, an individual who is self-monitoring alcohol consumption can take a blood test to compare the number of self-reported drinks consumed with blood alcohol levels. However, it should be noted that the natural by-products are not considered "ideal" incontrovertible indices (i.e., one-to-one correspondence between measure and behavior) because a third factor (e.g., confounding variables, measurement limitations) may be causing the results (Jackson, 1999). Unfortunately, none of the foregoing techniques can be used to assess accuracy when self-monitoring private events or covert behaviors given
these behaviors are only accessible to the individual who is self-monitoring (e.g., Ciminero et al., 1977; Korotitsch & Nelson-Gray, 1999).

Researchers have identified numerous variables that appear to have an effect on the accuracy of self-monitored data (Ciminero et al., 1977; Jackson, 1999; Korotitsch & Nelson-Gray, 1999). For example, providing training in self-monitoring has been shown to increase accuracy (Foster et al., 1999). According to Foster and colleagues, some training is better than no training; the training should include clear descriptions of the behavior being targeted and explicit directions regarding when to monitor and how to use the device. During training, it may be beneficial to have the individual repeat back all information and model correct ways to self-monitor.

Accuracy in self-monitoring also appears to increase when the individual is told that he or she is being checked for accuracy (Korotitsch & Nelson-Gray, 1999; Ciminero et al., 1977). In addition, research has shown that individuals tend to self-monitor more accurately when they are provided reinforcement for accurate monitoring (Baird & Nelson-Gray, 1999). Conversely, researchers have shown that accuracy decreases when an individual self-monitors while simultaneously engaging in other tasks (Korotitsch & Nelson-Gray). Hence, individuals should only monitor one behavior until they become comfortable, at which time additional target behaviors could be monitored.

Taken together, it is important to consider accuracy when using self-monitoring in behavioral assessment. The following methods are used to assess accuracy in self-monitoring: comparison to an external observer, comparison to a mechanical device, and comparison to natural by-products. Accuracy appears to be influenced by several factors, such as the type of target behavior, the number of target behaviors being recorded at the same time, awareness of accuracy checks, and whether or not reinforcement is included. Numerous researchers assert that training an individual on how and what to self-monitor increases accuracy (Ciminero et al., 1977; Foster et al., 1999; Jackson, 1999).

**CONTENT VALIDITY**

Content validity represents another psychometric concept that may be useful in evaluating the meaning and usefulness of data collected from self-monitoring procedures (Haynes, Richard, & Kubany, 1995; Jackson, 1999). Self-monitoring involves an assumption that multiple samples of the target behavior need to be sampled across time and situations. Thus, one potential value of using self-monitoring methods is greater access to many samples of behavior. In some cases, a therapist or researcher may access the literature and ascertain how many samples might be sufficient in order to have enough information for assessment purposes. In general, more stable behaviors of interest may need fewer samples for sufficient representation (Jackson).

Explicit behavioral definitions may be useful in ensuring that the monitored behavior adequately reflects and represents the conceptualized or intended target.
For example, an individual who smokes is asked to monitor each time and to specify the conditions under which he or she smokes a cigarette. The definition of smoking, however, must specify exactly what constitutes smoking—from one puff only to smoking the entire cigarette. In evaluating the content validity of self-monitoring, Jackson (1999) suggests that therapists and researchers carefully consider the purpose and meaning of the data.

**ISSUES FOR THE 21ST CENTURY**

With the advent of managed care and its concern with documenting outcomes, indirect methods of assessment have proliferated. Given that indirect methods of assessment (e.g., self-report questionnaires, interviews) are quick and easy to administer, research on these methods has risen dramatically, whereas research on self-monitoring has declined (Korotitsch & Nelson-Gray, 1999). Clinicians, however, continue to value the more direct method of self-monitoring and find it useful in their practice (e.g., Elliott et al., 1996). Korotitsch and Nelson-Gray (1999) note that the “climate of managed care also demands high-quality assessment that can enhance formulation of effective intervention as well as sensitive measures that can detect changes in the behaviors of interest” (p. 416). This type of climate may be particularly conducive to research on self-monitoring methods that can effectively evaluate treatment effectiveness.

**FUTURE RESEARCH**

Many questions remain unanswered regarding the validity and utility of data collected from self-monitoring procedures. Given that a lot of the earlier work on reactivity focused on nonclinical populations, this research needs to be replicated with clinical samples. Sensitivity of self-monitoring compared to other assessment methods also needs to be investigated. Potential biases associated with self-monitoring, the impact of self-monitoring on the quality of treatment, potential negative effects of self-monitoring (e.g., increased rumination or self-focus), and psychometric principles should be examined in order to evaluate the validity and utility of self-monitoring procedures.

**SUMMARY**

Self-monitoring is a commonly used technique in behavioral assessment that involves having an individual monitor and record his or her own behaviors. Behavioral assessors value self-monitoring because it is a practical method that emphasizes the situational specificity of target behaviors as they occur in the environment. In addition, self-monitoring is considered one of the only means for
assessing covert behaviors. Within behavioral assessment, self-monitoring has five primary purposes: education/description, diagnostic clarification, pretreatment assessment, treatment evaluation, and facilitation of behavior change. Self-monitoring represents a flexible technique that can capture different dimensions of behavior (e.g., frequency, intensity), using various recording formats (e.g., momentary time sampling, interval recording) and devices (e.g., paper-and-pencil, PDAs). Given current demands for competent and quality assessment (Cone, 1997), self-monitoring data need to be continuously and rigorously evaluated to ensure they are useful and meaningful for their intended purpose (Jackson, 1999).

REFERENCES


Introduction

Comprehensive assessment of patients with psychological problems requires gathering information across multiple domains, including information that can be self-reported as well as directly or indirectly observed. This fundamental approach to assessment has long been advocated by behavioral assessment authorities, including the classic tripartite assessment of anxiety (i.e., behavioral, self-reported, physiological) established by Lang (1968) and the extension of this approach into the behavioral assessment grid described by Cone (1978). Although this chapter focuses on psychophysiological assessment, the reader is reminded that a comprehensive assessment of any psychological problem involves obtaining observational data as well as documentation of self-reported phenomena.

In contrast to other assessment strategies, psychophysiological measurement involves neither subjective appraisals of functioning nor direct observation of easily detectable responses. Rather, psychophysiological assessment employs sophisticated electronic monitoring and recording of physiological parameters that can provide information regarding a particular patient's presenting problem or response to treatment. Although countless physiological parameters can be measured, psychophysiological assessment typically focuses on parameters known to be affected by nervous system activity, including both components of the central nervous system as well as both branches of the peripheral nervous system (i.e., the somatic nervous system and the autonomic nervous system). Recording apparatuses have become more sophisticated, so psychophysiological assessment now includes measures of nervous system functioning that could have only been dreamt about a decade ago (e.g., cortical activation via brain imaging).
Not surprisingly, early efforts to employ psychophysiological methods of assessment focused on gross changes in fairly easily detectible physiological phenomena such as heart rate and sweat gland activity. For example, Carl Jung (1910) employed a measure of skin resistance, which he termed the psychogalvanic response, on patients in his therapeutic practice. By simply noting words that elicited the most substantial skin resistance responses during word association tasks, he was able to quickly identify unconscious conflicts to target in analysis. Instrumentation in these early applications, however, was very crude and not sensitive enough to measure the minute physiological alterations (e.g., micromhos) that occur in response to presentation of mentally challenging or threatening stimuli. With the development of the polygraph, additional physiological parameters could be detected and measured systematically and were found to reflect changes in nervous system activity, including the brain (electroencephalography), muscles (electromyography), the heart (electrocardiography), eye movements (electrooculography), and the digestive system (electrogastrography). These devices also enabled signal amplification so that previously undetectable physiological changes could be easily observed and recorded. Over the decades, computerized signal acquisition systems were developed that permitted even further refinements in psychophysiological measurement. For example, the electrical activity of the heart measured via computerized electrocardiography yielded direct measures of important components of the signal itself, such as systolic time intervals, T-wave amplitude, and indices of contractile force, all measures that could not easily be determined using the paper record of the polygraph. Most recently, these computerized systems have been rebuilt as microprocessors, enabling development of tiny data-acquisition systems that can be worn by patients as they encounter activities that occur during normal daily life.

When employing psychophysiological assessment for almost any physiological parameter, an important distinction is made between tonic and phasic measures. Tonic measures of psychophysiology refer to physiological measures obtained during periods of time in which no specific stimulus has been presented to the individual (e.g., resting or baseline psychophysiological measures). For example, measures of elevated blood pressure observed among patients with essential hypertension and elevated electrical activity in the frontalis muscles of patients with tension muscle headaches represent measures of tonic physiological states.

In contrast, phasic psychophysiological measures reflect the magnitude of alterations that occur in response to known environmental stimuli. For example, heart rate is known to increase during speaking conditions, and the magnitude of this reaction is much greater among socially anxious patients in comparison to low socially anxious controls. Phasic psychophysiological measures are often expressed in terms of simple change scores, in which resting values are subtracted from values obtained during presentation of a particular stimulus context.

Although the distinction between tonic and phasic measures is important in basic psychophysiological research, it becomes less obvious in applications of
FIGURE 8.1 Organ systems and associated psychophysiological measurement parameters. Solid lines depict primary neural connections; dotted lines depict physiological parameters commonly measured in psychophysiological studies. Key: EEG = electroencephalography; ERPs = event related potentials; EMG = electromyography; EOG = electrooculography; GI = gastrointestinal; EGG = electrogastrography; EDA = electrodermal activity; Resp = respiration; PetCO₂ = pulmonary end tidal carbon dioxide.

Psychophysiological assessment methods when employed with clinical populations, where threatening stimuli of internal origin are often present. When assessing a phobic patient, for example, obtaining presumably tonic psychophysiological measures may be influenced unknowingly by a failure to account for the patient's internal anticipation of exposure to a phobic stimuli. In other words, "true" rest periods may be difficult to achieve among certain psychiatric conditions.

The primary purpose of this chapter is to summarize briefly the vast quantity of empirical work pertaining to the various parameters examined in psychophysiological investigations. Although a great deal of psychophysiological investigation has been conducted aimed at examining normal, adaptive responses to various stimulus presentations, the focus of this chapter is on using psychophysiological assessment with populations exhibiting various psychiatric disorders, specifically studies on experimental psychopathology. For additional information regarding the effect of human behavior on psychophysiological measures, the reader is referred to Stern, Ray, and Quigley (2001).

Figure 8.1 illustrates the organization of various psychophysiological parameters that have been examined with respect to a wide range of psychiatric
disorders. Parameters include those that largely reflect central nervous system functioning as well as those that reflect somatic and autonomic nervous system functioning. In order to provide a clear organization of the various measures of psychophysiological functioning, this schematic diagram grossly oversimplifies the complexity of relations among organ systems and their associated measurement parameters. In fact, all interrelations between organ systems and the negative feedback systems involved in regulating these organ systems have been omitted from this figure, for purposes of clarity.

CENTRAL NERVOUS SYSTEM FUNCTIONING

Although the central nervous system (CNS) comprises both the brain and the spinal cord, psychophysiological measurement of CNS activity has focused primarily on cortical functioning. Obviously, direct measures of neural activity of the brain are not possible in patients for whom the cortex is not typically exposed. Rather, psychophysiological parameters, which necessarily rely on surface electrodes or various imaging strategies for detection, can only provide indirect estimates of cortical activity. As depicted in Figure 8.1, electrical neural functioning occurring in the CNS can be detected using electroencephalography and associated measures of event-related electrical potentials. Measures of cortical functioning have also more recently been conducted using various brain-imaging strategies.

ELECTROENCEPHALOGRAPHY (EEG)

Electrical signals that emanate from neural activity in the brain can be detected using surface electrodes located on the scalp. In particular, measurement of these signals has been quite important in assessing general levels of cortical alertness and sleepiness. During states of alert wakefulness and mental or physical activation, the brain typically elicits a high-frequency, low-amplitude electrical activity called beta activity. During periods of relaxed wakefulness with eyes closed, however, the frequency of these electrical waves decreases and their amplitude increases slightly, resulting in recognizable alpha activity. Brain waves of even lower-frequency, theta and delta electrical brain wave activity, characterize various stages of deep sleep. Interestingly, as an individual enters paradoxical, or rapid-eye-movement (REM), sleep, the low-frequency electrical activity associated with deep sleep gives way to brain waves that are indistinguishable from beta wave activity. Therefore, the brain is just as awake during episodes of dreaming as it is during everyday problem solving.

Although measurement of EEG using early recording devices required the simultaneous detection and plotting of dozens of electrode configurations as well as a technician with a trained eye, computerized quantification of EEG signals has permitted more immediate and sophisticated analyses. For example, spectral
analyses of the EEG signal has enabled technicians very quickly to ascertain the predominant electrical wave form associated with a given time period. Based on the connection between cortical arousal and various psychopathologic disorders, EEG has been examined in several studies of experimental psychopathology. For example, studies have shown that alpha activity is decreased among patients with both anxiety disorders (e.g., Sachs, Anderer, Dantendorfer, & Saletu, 2004) and Attention-Deficit Hyperactivity Disorder (e.g., Barry, Clarke, & Johnstone, 2003). Additionally, patients with depressive disorders have been shown to display an EEG asymmetry characterized by increased alpha activity in the left frontal cortical regions (Gotlib, Ranganath, & Rosenfeld, 1998).

Although measures of EEG have provided important information regarding tonic levels of cortical arousal, they have not proven as useful in measuring phasic changes in cortical activity in response to specific stimuli. No doubt, this is a result of the fact that EEG represents a gross measure of the total electrical activity of the millions of neurons in the cortex. Alterations in the electrical activity of a single neural pathway in response to a specific stimuli would hardly impact a measure of total cortical arousal. However, surface electrodes employed to measure EEG can be used to provide reliable indices of phasic cortical response through the assessment of what have been termed event-related potentials.

**EVENT-RELATED POTENTIALS (ERPs)**

Electrical cortical responses to a specific stimulus can be measured as long as the stimulus can be presented repeatedly during data acquisition. Using multiple stimulus presentations, a computer averages measures of electrical activity in a designated area of the brain; through this averaging process, random electrical activity of the cortex fades into the background, leaving an observable waveform associated with the presentation of the stimulus. Within milliseconds (ms) following presentation of a stimulus, a series of characteristic waveforms can be detected that are tied to hypothesized components of information processing. A positive wave that occurs approximately 300 ms following stimulus presentation (known as the P300 wave) is associated with cognitive processing of the stimulus, including attention, stimulus discrimination, and decision making. The P300 is most notably detected using what has been called the “oddball” paradigm; using this paradigm, one stimulus is repeatedly presented, followed by a different stimulus (i.e., the “oddball”). The magnitude of the P300 wave is clearly enhanced following presentation of the “oddball,” lending support to the hypothesis that this component of the ERP wave is associated with awareness that a different stimulus has just been detected.

Because information processing is significantly disrupted in patients with thought disorders, patients with schizophrenia have been examined for decades using ERP research, including the “oddball” paradigm. Results of these studies have shown that P300 waves are noticeably smaller among patients with...
schizophrenia (e.g., Steinhauer & Zubin, 1982). Interestingly, the magnitude of the P300 wave returns to normal following treatment with neuroleptic medication, suggesting that the disruption in information processing detected among patients with schizophrenia is unique to their condition during a psychotic state.

ERP research has also been conducted on patients with other psychiatric disorders known to be associated with problems in information processing. For example, diminished P300 waves have been observed among children diagnosed with Attention-Deficit Disorder (Jonkman et al., 2000) and elderly patients diagnosed with dementia (Frodl et al., 2002). Recently, Pollak and Tolley-Schell (2003) reported enhanced P300 waves among physically abused children exposed to angry faces, indicating that increased attentional engagement was detected using this paradigm. In brief, the P300 component of the ERP waveform can be reliably used to ascertain both enhanced and reduced attentional processing associated with a number of diagnostic conditions.

Additional components of cortical electrical activity have been observed using ERP research methods. Slow potentials, such as the contingent negative variation (CNV), reflect gradual shifts in electrical potential that occur over several hundred milliseconds. CNV, for example, is observed when a participant is provided a warning stimulus before stimulus delivery. During this brief period when the participant is presumably waiting and anticipating stimulus delivery, a negative potential gradually emerges. CNV, therefore, has been hypothesized to correspond to expectancy. Like the more rapid responses of the ERP, slow potentials have been shown to exhibit characteristic dampened profiles among various types of psychopathology, including schizophrenia (McNeely, West, Christensen, & Alain, 2003), depression, and anxiety disorders (Tecce & Cattanach, 1987).

BRAIN IMAGING

With the advent of radiographic imaging strategies (e.g., positron emission tomography [PET]) and the more recent functional magnetic resonance imaging (MRI) technologies, research on experimental psychopathology has been able to examine more closely brain regions associated with various diagnostic groups. In general, these functional brain-imaging methods estimate cortical function in various regions of the brain by detecting the use of oxygen and/or glucose. Presumably, increased localized use of oxygen or glucose translates into increased neural activity in that brain region. Both tonic and phasic measures of cortical functioning obtained through brain-imaging methods can be useful to experimental psychopathologists.

Functional brain-imaging strategies have been conducted using patients with a wide array of psychiatric conditions. For example, prefrontal cortical activation that normally occurs during inhibition of a learned task has been shown to be impaired among patients with schizophrenia (MacDonald & Carter, 2003), lending support to hypotheses that schizophrenia is associated with frontal lobe dysfunction. In another functional neuroimaging study, limbic system activation
was shown to be higher among patients diagnosed with Major Depression when exposed to negative feedback during a cognitive task (Tucker, Luu, Frishkoff, Quiring, & Poulsen, 2003), indicating that hypersensitivity of the limbic system functioning plays a role in affective disorders. Brain-imaging research has also elucidated the importance of limbic system activation in Post-Traumatic Stress Disorder (Bremner et al., 1999) and enhanced activation of the orbitofrontal regions of the cortex in Obsessive-Compulsive Disorder (Anderson & Savage, 2004).

Brain-imaging strategies have also proven useful in detecting changes in cortical functioning that occur in response to treatment. For example, reductions in brain activity in the caudate nucleus, detected via PET scan, have been observed following treatment for Obsessive-Compulsive Disorder with either pharmacotherapy or behavior therapy (Baxter et al., 1992). In a recent intervention study, changes in brain activity were shown to differ between pharmacologic and cognitive-behavioral treatments for Major Depression (Goldapple et al., 2004); pharmacologic treatment was associated with increased prefrontal activation and hippocampal hypactivity, whereas cognitive-behavioral treatment was associated with the opposite physiological pattern. Clearly, by using brain-imaging technology, we are learning a great deal about how various successful treatment strategies work.

**SOMATIC NERVOUS SYSTEM FUNCTIONING**

The somatic nervous system refers to components of the central and peripheral nervous systems that regulate sensorimotor activity. Although many neural fibers of the somatic nervous system lie outside the CNS, several regions of the CNS are also involved in sensorimotor functioning, including the spinal cord and several cranial nerves that regulate motor activity of the muscles on the face and head. For example, three cranial nerves, the oculomotor (III), the trochlear (IV), and the abducens (VI) nerves, are responsible for eye movements. Signals for muscle activity for the rest of the body are transmitted from the brain to peripheral regions via the spinal cord.

**EYE MOVEMENTS**

Eye movements can be detected through gross visual observations or through movement sensors placed adjacent to the eyes, instrumentation referred to as electrooculography (EOG). Sleep researchers rely heavily on these measures of eye movements to depict onset and cessation of REM sleep. Although there are a variety of recognizable types of eye movements during the awakened state, considerable amount of research has focused on smooth-pursuit eye movements, a relatively slow eye movement used to track moving objects in the environment. These studies have confirmed that patients with schizophrenia exhibit notable
dysfunctions of smooth-pursuit eye movement (Sponheim, Iacono, Thuras, Nugent, & Beiser, 2003). Interestingly, first-degree relatives of patients with schizophrenia also exhibit identical dysfunctional smooth-pursuit eye movements, suggesting that this physiological phenomenon represents a marker for vulnerability for schizophrenia.

One type of eye movement that has received recent attention in experimental psychopathology is the magnitude of the eye blink response that occurs as a part of the startle response. Very simply, when an individual is exposed to an unpredictable burst of white noise, a startle reflex occurs, accompanied by an eye blink. Lang (1995) demonstrated that the magnitude of the eye blink startle response was enhanced under conditions of arousal of negative emotions, such as anger, fear, and sadness. Interestingly, the magnitude of the eye blink startle response can be attenuated if a prepulse stimulus (i.e., warning stimulus) is delivered beforehand, a phenomenon termed prepulse inhibition. The magnitude of this prepulse inhibition effect has been shown to be influenced by the valence of the emotion experienced among anxious or fearful individuals, so startle responses are attenuated during pleasant emotions but enhanced during unpleasant emotions (Cook, 1999). The distinction between the magnitudes of prepulse inhibition effects for pleasant and unpleasant emotions is not observed among patients with antisocial traits, individuals who presumably do not exhibit normal fear-based learning (Levenston, Patrick, Bradley, & Lang, 2000). Prepulse inhibition has been investigated in several patient samples; for example, patients with a variety of anxiety disorders, including Panic Disorder (e.g., Ludewig, Ludewig, Geyer, Hell, & Vollenweider, 2002) and Obsessive-Compulsive Disorder (e.g., Swerdlow, Benbow, Zisook, Geyer, & Braff, 1993), exhibit attenuated prepulse inhibition effects.

**MUSCLE ACTIVITY**

Electrical muscular activity, detected via electromyography (EMG), is measured using sets of electrodes similar to those used for measuring EEG or EOG. Increased electrical activity occurs in a particular muscle group during muscle tension, and reductions in electrical activity in the muscle group occur under relaxed states. Although EMG can be measured from any muscle group in the body, the resulting signal tends to represent a global measure of bodily muscle tension, because muscle groups quite distal from the electrode site can influence the signal. Certainly, movements affect EMG, but so does underlying muscle tension. In fact, EMG has served as a very helpful indicator of muscle tension in assessing and evaluating treatments for patients with muscle tension headaches (e.g., Ong, Nicholson, & Gramling, 2003). Furthermore, measures of EMG have been useful in assessing psychophysiological responses among patients with Post-Traumatic Stress Disorder (e.g., Carlson, Singelis, & Chemtob, 1997), Generalized Anxiety Disorder (e.g., Hazlett, McLeod, & Hoehn-Saric, 1994), and Panic Disorder (e.g., Beck & Scott, 1987).
AUTONOMIC NERVOUS SYSTEM
FUNCTIONING

The autonomic nervous system comprises two branches: the sympathetic nervous system, which is responsible for the fight-flight activation system affecting peripheral organ systems, and the parasympathetic nervous system, which is responsible for the relaxation response affecting many of the same organ systems. In brief, the fight-flight response, which is commonly triggered by the presence of a threatening stimulus, involves a redistribution of blood flow and energy resources from the body's primary organs to fuel the muscle and brain activity needed to confront or escape from the source of threat. In contrast, the relaxation response returns blood flow to homeostatic life support systems (e.g., the gastrointestinal system). Although initially thought to operate reciprocally, the sympathetic and parasympathetic branches of the autonomic nervous system have been shown to operate independently (Berntson, Cacioppo, & Quigley, 1991). In fact, both sympathetic and parasympathetic activity can be enhanced in response to a given stimulus. As depicted in Figure 8.1, many psychophysiological measures can be used to assess autonomic nervous system activity.

PUPILLOMETRY

Pupillary dilation and constriction represent easily observable physiological responses that reflect sympathetic and parasympathetic nervous system activation, respectively. In contrast to the electrodes used to measure several other psychophysiological parameters, pupillometry involves measurement of pupil diameter via a video recording camera. Although pupil diameter seems fairly easy to measure, measurement is complicated by numerous extraneous factors, including eye movement, a number of eye reflexes, and the amount of light present in the environment. When psychophysiologists have controlled these sources of variability, they have learned a great deal about the relation between pupil diameter and CNS activity; in brief, the magnitude of pupillary responses to specific stimuli has been shown to be associated with cognitive processing occurring in the CNS (Granholm & Steinhauer, 2004). As such, pupillary responses to selected stimuli can be used to examine indirect parameters of cognitive processing among various pathological conditions. For example, the magnitude of the pupillary response to a given stimulus has been shown to be significantly reduced among patients with schizophrenia (Steinhauer & Zubin, 1982) and Alzheimer's dementia (Potiou, Fountoulakis, Tsolaki, Goulas, & Palakaris, 2000), conditions both hypothesized to be associated with dysfunctional cognitive processing. Additionally, because psychoactive drugs are known to alter pupillometric responses, this psychophysiological assessment strategy has proven useful in detecting external drug use among patients in substance abuse treatment programs (Murillo, Crucilla, Schmittner, Hotchkiss, & Pickworth, 2004).
GASTROINTESTINAL MOTILITY

The gastrointestinal system is influenced by both sympathetic and parasympathetic nervous systems; sympathetic activation results in decreased blood flow to the gastrointestinal system, while parasympathetic activity results in increased blood flow for purposes of digestion and motility. The electrogastrogram (EGG) is a device that employs electrodes positioned over the abdomen that monitor the electrical activity and contractility of the gastrointestinal system. Because stress is associated with increased gastric dysrhythmia (Stern, Vasey, Hu, & Koch, 1991), investigators have turned to examining EGG recordings among patients with disorders that affect the gastrointestinal system. In these studies, abnormal EGG responses have been observed in patients with gastroesophageal reflux disease (Leahy, Besherdas, Clayman, Mason, & Epstein, 2001), functional dyspepsia, and irritable bowel syndrome (Leahy, Besherdas, Clayman, Mason, & Epstein, 1999). In an interesting application of EGG technology, Gianaros, Stern, Morrow, and Hickok (2001) demonstrated that pretreatment episodes of gastric tachyarrhythmia predicted incidence of chemotherapy-induced nausea among cancer patients.

ELECTRODERMAL ACTIVITY

Electrodermal activity (EDA) stands out as a measure of autonomic nervous system functioning that is solely mediated by sympathetic nervous system without parasympathetic influence. Unusually, this tract of the autonomic nervous system employs acetylcholine as a neurotransmitter, rather than the more commonly observed noradrenergic sympathetic nervous system neurotransmission. In contrast to other sweat glands of the body, the eccrine glands on the palms of the hands and the soles of the feet are responsive to mental activation and emotion, and the degree of sweat gland activity can be measured easily using surface electrodes attached to the hand or fingers. Several measures of EDA can be obtained, the most common being skin conductance level (SCL) and skin conductance response (SCR). In both cases a low-current electrical charge is transmitted across the surface of the skin. Because sweat is an excellent conductor, the charge is conducted more quickly when sweat activity increases and more slowly when sweat activity decreases. SCL represents a tonic measure of EDA, and SCR represents a phasic measure of EDA in response to a given stimulus. Some authors have reported EDA findings using the terms skin resistance level and skin resistance response; these values are simply inverse measures of SCL and SCR, respectively. Because the skin possesses its own intrinsic electrical activity, this can also be measured without exposing the skin to an external electrical source, a measure termed skin potential. Most experimental work on EDA, however, reports findings in terms of SCL or SCR.

During exposure to stressful stimuli, sweat glands exhibit increased rates of secretion that result in increased SCR; when the stressful stimuli are removed,
measures of skin conductance gradually return to prestress levels. Therefore, both measures of magnitude of SCR and recovery rate can be examined. Numerous investigations of SCR and recovery rate have been conducted on patient populations. In general, patients with anxiety disorders exhibit higher SCLs (e.g., Bond, James, & Lader, 1974) and greater SCRs during exposure to feared stimuli and slower recovery rates than nonanxious controls (e.g., Wessel & Merckelbach, 1998), and patients with antisocial personality characteristics or disturbances of conduct exhibit lower SCLs and smaller SCRs to stress than controls (Lorber, 2004). Therefore, studies employing measures of EDA have confirmed that patients with anxiety disorders are overaroused, while patients with antisocial traits are more commonly underaroused.

EDA has been examined in patient populations other than those with anxiety or antisocial/conduct disorders. For example, SCRs of depressed patients are typically dampened, in contrast to nondepressed controls (e.g., Iacano, 1984). In contrast, patients with schizophrenia typically exhibit increased SCL as well as frequent nonspecific SCRs (Dawson & Schell, 2002), and patients with binge-eating disorders exhibit heightened SCRs during exposure to their favorite binge food (Vögele & Florin, 1997).

EDA has been particularly valuable in assessing physiological components of the anxiety response of patients undergoing treatment. In his classic study on systematic desensitization, Paul (1966) demonstrated the superiority of the desensitization procedure for treating phobia by using SCR as an outcome measure. Boulougoulis, Marks, and Marset (1971) demonstrated that significantly greater reductions in SCR occurred among participants receiving exposure therapy, in contrast to those being treated with desensitization. Based on the importance of assessing EDA in these early-intervention studies, monitoring of sympathetic arousal via EDA continues to be employed as an indicator of treatment progress in many clinic settings.

RESPIRATION

Respiration is another easily detectable psychophysiological measure. Although it can be thought of as a measure of autonomic nervous system activity, it is also regulated from brain stem mechanisms in the CNS as well as through somatic diaphragm muscle activity. Although many sophisticated measures of respiratory parameters can be obtained (e.g., end tidal volume, peak inspiratory flow), researchers have typically focused on a simple measure of respiratory rate or amplitude. It is well known, for example, that rate of respiration increases under conditions of anxiety and stress but decreases under conditions of relaxation. Despite its ease of measurement, it is not used frequently in studies on experimental psychopathology, except among relatively recent studies examining the role of hyperventilation in the etiology of Panic Disorder (see Wilhelm & Roth, 2001). In brief, due to the significance of respiratory symptoms among many patients diagnosed with Panic Disorder, researchers have examined whether
the report of these symptoms was based on actual differences in respiratory functioning among patients with Panic Disorder. Indeed, Hegel and Ferguson (1997) reported that patients with Panic Disorder exhibited lower pulmonary end-tidal CO₂ levels than Generalized Anxiety Disorder patients or nonanxious controls. These findings support clinical observations that chronic hyperventilation is involved in the etiology of panic disorder.

**CARDIOVASCULAR ACTIVITY**

Measures of cardiovascular functioning represent the most widely used psychophysiological measures of autonomic functioning, due to the relative ease with which heart rate (HR) signals and blood pressure (BP) determinations can be made. Although there is a relation between HR and BP, it is not a simple or direct one. BP in the circulatory system is influenced by both cardiac and vascular factors. Both increases in HR and amount of blood ejected from the heart during each beat (i.e., stroke volume) represent cardiac influences that lead to increased BP through increased cardiac output. The degree of vasoconstriction and vasodilation in various segments of the circulatory system represent vascular influences on BP by altering peripheral resistance to blood flow. Because systolic blood pressure reflects BP during cardiac pump action, it is commonly assumed to represent a better index of the cardiac influence on BP than vascular influences; in contrast, diastolic blood pressure is often assumed to be a better index of vascular influences on BP. Although these assumptions tend to be true, more accurate measures of cardiac and vascular influences on BP can be obtained using impedance cardiography (see Sherwood et al., 1990).

Cardiac parameters, such as HR, reflect the joint innervation of the sympathetic and parasympathetic nervous systems, while vascular parameters, such as forearm blood flow, reflect sympathetic nervous system activation; firing of the β-adrenergic branch results in vasodilation in the forearm, and firing of the α-adrenergic branch results in vasoconstriction. Although HR is one of the more commonly used psychophysiological parameters, it is often difficult to interpret due to its dual innervation by sympathetic and parasympathetic systems. Increases in HR in response to a given stimulus could represent increased sympathetic arousal, decreased parasympathetic activation, or some combination of the two. Some additional cardiovascular measures have proven useful in assisting researchers in determining which neural system is responsible for the observed cardiac effects. Pre-ejection period (PEP) is a systolic time interval of the cardiac cycle that has been shown to reflect primarily β-adrenergic activation. Heart rate variability (HRV) and respiratory sinus arrhythmia (RSA), the degree to which HR is influenced by the respiratory cycle, have been shown to reflect vagal parasympathetic influences on the heart. Thus, shortening of the PEP accompanying a faster HR suggests that the increased cardiac arousal is mediated by the sympathetic nervous system, and decreased HRV or RSA accompa-
nying a faster HR suggests that the increased cardiac arousal is mediated by the reduced parasympathetic nervous system tone.

Similar to EDA, both tonic and phasic measures of HR are commonly elevated among patients with anxiety disorders, in contrast to nonanxious individuals (e.g., Bond et al., 1974), and persons with antisocial traits exhibit lower HR and HR responses to stress than controls (Lorber, 2004). To examine whether these observed differences in HR can be attributed to sympathetic or parasympathetic influences, some investigators have measured PEP or HRV in patient groups. Both Generalized Anxiety Disorder and Panic Disorder, for example, are associated with decreased HRV (Friedman & Thayer, 1998; Thayer, Friedman, & Borkovec, 1996), indicating a parasympathetic nervous system dysfunction in these anxiety disorders. Comparable parasympathetic dysfunction has been observed among depressed patients (e.g., Gorman & Sloan, 2000). In contrast, reduced PEP response has been observed among male adolescents with conduct disorders (Beauchaine, Katkin, Strassberg, & Snarr, 2001), suggesting that sympathetic nervous system dysfunction may be responsible for the commonly observed lower HR among persons with antisocial traits. Comparable findings have been reported among patients with other impulse control problems, such as bulimia nervosa (Koo-Loeb, Pedersen, & Girdler, 1998).

Measures of BP and forearm blood flow have also been shown to distinguish patients with anxiety disorders from controls. Increased BP responses (e.g., Buckley & Kaloupek, 2001) and increased forearm blood flow (i.e., vasodilation; e.g., Kelly & Walter, 1969) have been observed among patients with a variety of anxiety disorders. Reductions in blood flow to the extremities (i.e., vasoconstriction) among anxious patients have also been reported (e.g., Ackner, 1956), indicating that the distribution of blood flow in the circulation in patients with anxiety disorders mimics blood flow observed during the classic fight-flight response. Finally, measures of blood volume in genital regions have proven instrumental in assessing patients with sexual disorders or sexual dysfunctions (Geer, O’Donohue, & Schorman, 1986).

NEUROENDOCRINE SYSTEM FUNCTIONING

Measures of neuroendocrine system activity typically involve blood draws or urine screens, although recent assays have been developed to assess various elements of the neuroendocrine system from salivary samples. Some investigators do not consider measures of neuroendocrine activity as psychophysiological parameters, for they typically require invasive assessment methods. However, for the purposes of this chapter and because of the ease with which blood, urine, and salivary samples can be obtained, they can be considered here. Additionally, because the neuroendocrine system is quite responsive to stress, both tonic and phasic measures can be obtained from patient samples exposed to clinically relevant stimuli. The neuroendocrine system operates much like the nervous system,
except it employs the bloodstream, rather than neural pathways, for signal transmission. For example, the pituitary gland, in response to a brain signal, emits chemical-releasing or-stimulating factors into the bloodstream that travel through the circulatory system and trigger various endocrine glands to release hormones. Catecholamines (norepinephrine and epinephrine) and cortisol are released from the adrenal gland, thyroid hormone is released from the thyroid glands, and various sex hormones (e.g., estrogen, testosterone) are released from the gonads. Obviously, blood concentrations of all of these hormones and releasing/stimulating factors can be measured through a simple blood draw.

**HYPOTHALAMIC-PITUITARY-RELEASING FACTORS**

Although several pituitary-releasing or-stimulating factors have been identified and measured, only a few have been systematically examined with respect to various psychiatric disorders. For example, corticotropin-releasing factor (CRF), the hormone released by the pituitary that triggers the release of cortisol by the adrenal cortex, has been shown to be elevated among persons with both anxiety and mood disorders (Ströhle & Holsboer, 2003). Thyroid-stimulating hormone (TSH), which similarly leads to the release of thyroid hormone by the thyroid glands, has been reported to be decreased among patients with anxiety or mood disorders (Rao, Vartzopoulos, & Fels, 1989). Because mood and anxiety disorders are both hypothesized to be associated with altered hypothalamic-pituitary function, it is not surprising that several studies have explored hormone levels associated with altered releasing- or stimulating-factor levels among a variety of psychopathological conditions.

**STRESS HORMONES**

Although many hormones can be measured, most experimental psychopathologists have concentrated on the so-called stress hormones: cortisol, epinephrine, and norepinephrine. Cortisol is released into circulation from the adrenal cortex and has been shown to be elevated in a number of patient samples, including patients with anxiety disorders, such as Panic Disorder (Bandelow et al., 2000) and Obsessive-Compulsive Disorder (Monteleone, Catapano, Tortorella, & Maj, 1997), and mood disorders (Ströhle & Holsboer, 2003). Comparable to psychophysiological parameters of the autonomic nervous system, cortisol levels have been shown to be lower among patients with antisocial personality characteristics (Bergman & Brismar, 1994).

Catecholamines are released from the adrenal medulla and have direct excitatory effects on the cardiovascular system as well as the autonomic nervous system. In particular, norepinephrine has been shown to be elevated in patients with anxiety and mood disorders (e.g., Sevy, Papadimitriou, Surmont, Goldman, & Mendlewicz, 1989) but reduced among patients with antisocial personality traits (e.g., Lidberg, Levander, Schalling, & Lidberg, 1978).
Levels of thyroid hormone have also been investigated among various patient groups. Rao et al. (1989), for example, reported lower levels of thyroid hormones among both anxious and depressed women. Interestingly, recent reports have indicated that elevated thyroid hormone levels are associated with antisocial behaviors (Soderstrom & Forsman, 2004).

In summary, distinct neuroendocrine profiles have emerged among anxiety/mood disorders and antisocial/conduct disorders. In general, these profiles are consistent with findings derived from studies of the autonomic nervous system that support Gray's (1975) hypothesis that anxiety disorders are associated with hyperresponsive behavioral inhibition systems and that antisocial behavior disorders are associated with hyporesponsive systems.

FACETS OF PSYCHOPHYSIOLOGICAL ASSESSMENT

It is apparent that researchers have numerous physiological parameters to consider when electing to employ psychophysiological assessment in studies on experimental psychopathology. Foremost among these considerations is selecting the appropriate measurement parameter(s). Although one certainly needs to consider the availability of recording instrumentation, that should not represent the primary consideration in making this decision. Of utmost importance in selecting the measurement parameter is considering the appropriateness of specific parameters and their conceptual relevance to the experimental question being asked. For example, studies examining the efficacy of progressive muscle relaxation training certainly should employ EMG for purposes of evaluating treatment outcome. Many researchers have fallen prey to a tendency to select psychophysiological assessment measures that are relatively easy to assess (e.g., HR) when other parameters may have been more appropriate. Although changes in HR, for example, may reflect sympathetic nervous system activation, they may also reflect alterations in parasympathetic influence, rendering interpretations of the neural systems involved in changes in HR difficult to make.

In general, it is a good rule of thumb to employ multiple psychophysiological measures whenever possible. According to Fowles (1986), HR is a better index of the behavioral activation system and EDA is a better measure of the behavioral inhibition system. Based on this premise, HR would be more appropriate for measuring interventions for impulse control disorders and EDA would be more appropriate for measuring interventions for anxiety disorders. Obviously, if both parameters were assessed, stronger conclusions could be drawn regarding the specific neural systems affected by the intervention.

Once appropriate parameters have been selected, researchers should carefully consider their selection of stimuli used during assessment and the mode of stimulus presentation. It is well known that both type of task selected and the mode of presentation will influence the psychophysiological response profile, a
phenomenon termed *stimulus response specificity*. For example, tasks involving processes of sensory intake (e.g., watching a film segment) will elicit quite distinct physiological responses, in comparison with tasks involving sensory rejection (e.g., recalling a fearful incident; Lacey, Kagan, Lacey, & Moss, 1963). In this regard, it is essential to include a nonpatient comparison group to elucidate the expected physiological response profile for the specific task chosen. Like selecting physiological responses to measure, multiple stimuli are optimal, including standardized “control” stimuli to evaluate overall physiological responsiveness. Additionally, with the advent of ambulatory recording technology, exposure to real-life threatening stimuli can be considered, rather than relying entirely on contrived laboratory situations or standardized laboratory stressors, which both have limited generalizability to real-life settings.

Although the appropriateness of selected stimuli is important, it is also essential to obtain measures of physiological functioning proximal to stimulus presentation to serve as resting or baseline levels. Of considerable importance is allowing the patient to accommodate to the novel surroundings of the experimental environment, a process termed *adaptation*. Depending on the nature of the experimental procedure, the duration of this adaptation phase may differ. For example, in studies employing cannula placement for purposes of neuroendocrine or arterial pressure assessment, patients often need a much longer adaptation period to “calm down” after insertion of the cannula. In many cases (e.g., highly anxious patients), it may be more difficult to obtain reliable tonic measures of physiological functioning during these resting periods than to obtain the phasic response levels during stimulus presentation. It is also important to obtain resting levels that are proximal to the stimulus presentations, for many factors (e.g., time of day, temperature, substance use, menstrual phase) are known to alter resting physiological state and could make findings difficult to interpret if the patient is instructed to return at a different day or time for completion of the assessment.

Physiological responses to repeated presentations of the same stimuli tend to decrease over time, a process called *habituation*. In some patient samples, however, habituation may be delayed. For example, SCRs among phobic patients exposed to feared stimuli may not habituate as quickly as SCRs of nonphobic controls, possibly due to the typically abbreviated exposure times that phobic patients can endure. Habituation rates among patients normalize following interventions aimed at increasing the duration of exposure to feared stimuli (e.g., exposure therapy, implosive therapy). However, because habituation is a universal phenomenon, experimental psychopathologists are obligated to employ appropriate control conditions in order to interpret physiological responses to repeated stimulus presentations.

Once data have been collected, the experimenter needs to consider how such psychophysiological data will be reduced and analyzed. Although raw scores of physiological functioning can be simply recorded and analyzed, psychophysiologists are frequently interested in analyzing the difference between the resting
state and during stimulus presentation. As stated earlier, one strategy for handling psychophysiological data is through calculation of change scores, by simply subtracting average values of a rest period from those obtained during presentation of a designated stimulus (e.g., phobic object). Although this strategy is widely used, it can be criticized because there is often an association between resting levels and the magnitude of the physiological response, a phenomenon called the law of initial values. Elevated resting levels of a given parameter may be associated with reductions in the magnitude of physiological responding due to a ceiling effect or with increased physiological responding due to a priming effect. In order to control for the association between resting and stress-induced physiological levels, psychophysiological researchers have employed analyses of covariance or residualized change scores to remove the portion of “reactive” variance that is explained by resting levels of that parameter.

Psychophysiological parameters, of course, should not represent sole dependent variables in studies of experimental psychopathology; rather, they should be combined with measures of affect, cognition, and behavior obtained via self-report and direct behavioral observations. However, inclusion of several measurement domains can be problematic for researchers, because correlations among these various response channels are typically quite low. Wilhelm and Roth (2001) argue convincingly that concordance among physiological and self-reported measurement parameters can be improved if more attention is paid to making sure assessment domains are measuring a common construct. For example, physiological respiratory parameters are more likely to be associated with self-reported ratings of shortness of breath than with overall anxiety symptoms. Thus, experimental psychopathologists need to select the nonphysiological parameters of interest as carefully as the physiological parameters.

SUMMARY

Psychophysiological assessment provides researchers and clinicians important clues about cognitive and emotional functioning that are not easily observed with the naked eye. Through the careful selection of response parameters and application of recording apparatuses, these tiny physiological responses can be detected and used to explore underlying physiological pathologies of psychological disorders as well as effective strategies for monitoring treatment progress and outcome. Given the careful attention to the selection of physiological parameters and eliciting stimuli, important knowledge regarding the physiological foundations for a variety of psychological disorders has been acquired. For example, based on sound psychophysiological research, distinct autonomic nervous system and neuroendocrine profiles have been observed among patients with anxiety/mood disorders and those with disorders of conduct. In addition, certain psychophysiological measures have become useful markers for psychopathology, such as smooth-pursuit eye movement dysfunctions among persons at risk for
developing schizophrenia. As distinct physiological systems at the root of various psychiatric disorders are identified, interventions aimed specifically at those underlying physiological disturbances can be developed and tested both in the laboratory and during daily life. With continued collaboration between psychophysiologists and experimental psychopathologists, the role of psychophysiological assessment holds much promise in assessing and treating numerous psychological disorders.

REFERENCES


