Journal of Personality Assessment, *93*(4), 380–389, 2011

ISSN: 0022-3891 print / 1532-7752 online
DOI: 10.1080/00223891.2011.577475

Routledge
Taylor & Francis Group

Computerized Adaptive Assessment of Personality Disorder: Introducing the CAT–PD Project

LEONARD J. SIMMS,[1] LEWIS R. GOLDBERG,[2] JOHN E. ROBERTS,[1] DAVID WATSON,[3] JOHN WELTE,[4] AND JANE H. ROTTERMAN[1]

[1]*Department of Psychology, University at Buffalo, The State University of New York*
[2]*Oregon Research Institute, Eugene, Oregon*
[3]*Department of Psychology, University of Notre Dame*
[4]*Research Institute on Addictions, University at Buffalo, The State University of New York*

Assessment of personality disorders (PD) has been hindered by reliance on the problematic categorical model embodied in the most recent *Diagnostic and Statistical Model of Mental Disorders* (*DSM*), lack of consensus among alternative dimensional models, and inefficient measurement methods. This article describes the rationale for and early results from a multiyear study funded by the National Institute of Mental Health that was designed to develop an integrative and comprehensive model and efficient measure of PD trait dimensions. To accomplish these goals, we are in the midst of a 5-phase project to develop and validate the model and measure. The results of Phase 1 of the project—which was focused on developing the PD traits to be assessed and the initial item pool—resulted in a candidate list of 59 PD traits and an initial item pool of 2,589 items. Data collection and structural analyses in community and patient samples will inform the ultimate structure of the measure, and computerized adaptive testing will permit efficient measurement of the resultant traits. The resultant Computerized Adaptive Test of Personality Disorder (CAT–PD) will be well positioned as a measure of the proposed *DSM–5 PD* traits. Implications for both applied and basic personality research are discussed.

Personality pathology is prevalent in the community and in mental health settings. In recent epidemiological surveys of *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. [*DSM–IV*]; American Psychiatric Association, 1994) and *International Classification of Diseases* (10th ed. [*ICD–10*; World Health Organization, 1992]) personality disorders (PDs), prevalence rates have ranged between 9% and 14% in community samples (e.g., Ekselius, Tillfors, Furmark, & Fredrikson, 2001; Samuels et al., 2002; Torgersen, Kringlen, & Cramer, 2001) and as high as 45% in patient samples (e.g., Zimmerman, Rothschild, & Chelminski, 2005). Moreover, personality pathology is highly comorbid with Axis I disorders (Grant et al., 2005; Zimmerman et al., 2005), and such comorbidity changes or complicates the treatment course of such syndromes. Studies have indicated that personality pathology negatively affects the course and outcome of both psychotherapeutic and pharmacological treatments for Axis I disorders (e.g., Cyranowski et al., 2004; Feske et al., 2004; Reich, 2003) and is associated with higher health-care service utilization and functional impairment in a variety of important life domains (e.g., Bender et al., 2001; Skodol et al., 2005; Smith & Benjamin, 2002).

Thus, personality pathology is an important mental health concern that should be routinely assessed and treated in mental health settings. However, the length and administration time associated with most PD measures, the problems associated with the current categorical classification of PD, and the wide variety of alternative PD models and measures have interfered with the routine assessment of personality pathology in both research and clinical settings. Both interview and questionnaire methods can be costly in terms of the time and resources required to administer, score, and interpret them properly. Unfortunately, these time and staff requirements are difficult to accommodate in most settings (e.g., Piotrowski, 1999; Piotrowski, Belter, & Keller, 1998; Yates & Taub, 2003), especially in the context of current ambiguities about how best to classify and assess personality problems. To remedy these concerns and improve the assessment of PD, new classification models are needed that bridge and improve on existing models. Moreover, measurement methods are needed that increase the efficiency of PD assessment. The development of comprehensive and efficient measures of personality pathology would be an important addition to the toolbox of mental health clinicians and researchers.

CATEGORICAL VS. DIMENSIONAL CLASSIFICATION OF PERSONALITY PATHOLOGY

Traditional nosological systems of personality pathology, such as *DSM–IV* and *ICD–10*, describe PD using a medical model within which pathological syndromes are viewed as being either present or absent. However, although the inclusion of PDs on Axis II as an independent domain in *DSM–III* (3rd ed.; American Psychiatric Association, 1980) was an important advance, the categorical model used by that and subsequent editions of the *DSM* suffers from a number of problems that limit its usefulness, including high rates of diagnostic comorbidity among purportedly distinct PDs (e.g., Clark, Watson, & Reynolds, 1995; Dolan, Evans, & Norton, 1995; Fossati et al., 2000; Oldham et al., 1995), within-disorder heterogeneity (e.g., Clark et al., 1995; Widiger, 1993), an arbitrary boundary between normal and abnormal personality traits (e.g., Clark et al., 1995; Livesley, Jang, & Vernon, 1998; Widiger & Clark, 2000), poor reliability (Dreessen & Arntz, 1998; Pilkonis et al., 1995;

Received August 23, 2010; Revised December 13, 2010.

Address correspondence to Leonard J. Simms, Department of Psychology, University at Buffalo, State University of New York, Park Hall 218, Buffalo, NY 14260; Email: ljsimms@buffalo.edu

Zanarini et al., 2000), and low convergent validity (see Clark, Livesley, & Morey, 1997, for a review). Moreover, categorical systems appear to result in substantial information loss, especially for individuals who manifest clinically significant signs and symptoms that do not quite reach the arbitrary thresholds specified by the *DSM*.

As a result, many have called for a dimensional approach to describing and assessing personality pathology (e.g., Clark, 2007; Livesley & Jackson, 2009; Widiger & Clark, 2000; Widiger & Simonsen, 2005). Two basic approaches have been proposed to dimensionalize Axis II. The first is to maintain the current PD category labels and simply measure them along continua, either by summing the *DSM* criteria, creating measures to tap the relevant aspects of each PD construct, or by developing rational or empirical PD prototypes that can be rated dimensionally in terms of prototype similarity (e.g., Oldham & Skodol, 2000; Shedler & Westen, 2004; Westen, Shedler, & Bradley, 2006). Such methods generally lead to increased stability of measurement (e.g., Zanarini et al., 2000); however, diagnostic overlap and within-class heterogeneity are still problems (Clark, 2007). The second method involves (a) discarding the a priori assumption that personality pathology is adequately defined by the current set of *DSM–IV* PD categories, and (b) identifying and measuring the trait dimensions that underlie phenotypic manifestations of personality pathology.

A number of measures and models have been proposed along these lines. Most notably, the Five-factor model (FFM) has gathered support in recent years as a viable model for personality pathology in general and as a potential basis for describing PD in the next revision of the *DSM* (e.g., Miller et al., 2010; Widiger & Simonsen, 2005; Widiger & Trull, 2007). In addition, the interpersonal circumplex has been proposed as a dimensional framework for understanding personality pathology (e.g., Pincus & Gurtman, 2006). Both of these models represent attempts to explain personality pathology in terms of existing structural models of *normal-range* personality. In contrast, several bottom-up, PD-specific models have been offered (e.g., the Schedule for Nonadaptive and Adaptive Personality–2nd Edition [SNAP–2; Clark, Simms, Wu, & Casillas, in press] and the Dimensional Assessment of Personality Pathology–Basic Questionnaire [DAPP–BQ; Livesley & Jackson, 2009]) that first identify lower order traits relevant to PD and then let the covariance among those traits drive the ultimate structure of the domain.

In either case, the primary descriptive units in dimensional models of this second broad type are the basic personality traits that underlie the domain of personality pathology. Patients are rated on a number of distinct traits relevant to personality dysfunction, rather than being placed in one or more diagnostic categories. The distinction between normal and abnormal functioning then can be determined on the basis of empirical criteria. Statistical infrequency is a common criterion of abnormality in dimensional models, with individuals scoring, say, 1.5 or 2 *SD* above or below the norm considered to be in the "abnormal" or "pathological" range. Of course, the use and location of cut points along dimensions can be arbitrary and generally results in the loss of statistical power, leading some to eschew the use of cutoffs and interpret dimensional scores quantitatively, either relative to norms or to other dimensions within the same clinical profile. Moreover, some have argued that statistical infrequency alone is an inadequate criterion to signal the presence of personality disorder unless it is coupled with concomitant dysfunction or impairment in important areas of functioning (e.g., Livesley & Jang, 2000; Tyrer, 2005). In any case, comprehensive dimensional systems can be used as the foundation for empirically based classification systems (e.g., based on latent class and latent profile analyses) in which diagnostic entities are formed by identifying individuals with similar profiles of personality traits (e.g., Eaton, Krueger, South, Simms, & Clark, in press). Thus, although the primary strength of trait-based systems is their ability to yield relatively homogeneous and distinctive dimensional interpretations, such systems also are flexible enough to yield empirically based categorical information.

Dimensional Models of PD

A variety of dimensional models have been proposed as alternatives to the current categorical PD system (e.g., Clark, 1993; Clark et al., in press; Livesley & Jackson, 2009; Widiger & Simonsen, 2005, 2006; Widiger & Trull, 2007). Widiger and Simonsen (2005, 2006) summarized 18 dimensional PD models and organized the traits included in these models into a pathology-slanted version of the FFM: (a) extraversion versus introversion, (b) antagonism versus compliance, (c) constraint versus impulsivity, (d) emotional dysregulation versus emotional stability, and (e) unconventionality versus closedness to experience. Within this integrative framework, each broad domain is made up of a number of narrower, lower order dimensions that, when factored, should give rise to the five higher order dimensions. Table 1 includes a summary of Widiger and Simonsen's (2005, 2006) assignment of lower order dimensions to each broad domain. Several themes are apparent from this collection of dimensions. First, within each broad domain, there is substantial overlap across similarly named traits (e.g., sociability and social closeness vs. aloofness, detachment, and social avoidance all appear to tap quite similar aspects of interpersonal behavior in the extraversion–introversion domain). Second, some lower order traits are listed across multiple domains (e.g., alienation, entitlement, social closeness, dependency), which likely is due to different conceptualizations of these traits across models.

For these reasons, the lower order dimensions identified by Widiger and Simonsen are too numerous and overlapping in their current form to be of real practical value. Moreover, each of the 18 models that contributed to Table 1 (including such prominent models as the FFM, Livesley's DAPP–BQ, and Clark's SNAP–2) is incomplete in its representation of relevant dimensions of personality pathology. Widiger and Simonsen (2006) summarized their review by noting that "none of the models lacks any limitations that could not at times be well compensated through an integration with another model" (p. 3). Thus, although previous models exist to describe the lower order structure of personality pathology, no single model proposed to date encompasses the full range and breadth of dimensions relevant to personality pathology. Based on these models, Widiger and Simonsen (2006) concluded "that an important goal of future research will be the identification of a common ground among alternative dimensional models of personality disorder" (p. 15). As such, additional work is needed to refine this set of dimensions (i.e., identify the core, nonoverlapping constructs that are relevant to personality pathology) and to generate a measure that efficiently and practically assesses each of them.

Working Toward *DSM–5*

The American Psychiatric Association began the planning process for *DSM–5* in 1999. The proposed revisions were

TABLE 1.—Summary of personality-disorder dimensions identified by Widiger and Simonsen (2005, 2006), organized by the Five-factor model.

Broad Domain	Relevant Lower Order Trait Facets Identified in the Literature
Extraversion vs. introversion	Activity, aloofness, assertiveness, detachment, entitlement, excitement seeking, exhibitionism, exploratory excitability, extravagance, gregariousness, histrionic sexualization, intimacy problems, optimism, positive emotionality, restricted expression, schizoid orientation, shyness, sociability, social avoidance, social closeness, social potency, stimulus seeking, warmth, well-being
Antagonism vs. compliance	Aggression, agreeableness, alienation, altruism, attachment, callousness, compassion, compliance, conduct problems, dependency, diffidence, empathy, entitlement, helpfulness, insecure attachment, interpersonal disesteem, manipulativeness, mistrust, modesty, narcissism, passive oppositionality, psychopathy, pure-hearted, rejection, sentimentality, social acceptance, social closeness, straightforwardness, submissiveness, suspiciousness, tender-mindedness, trust
Constraint vs. impulsivity	Achievement-striving, childishness, competence, compulsivity, conscientiousness, deliberation, disorderliness, dutifulness, eagerness of effort, harm avoidance, impulsivity, irresponsibility, obsessionality, order, perfectionism, propriety, resourcefulness, responsibility, risk taking, self-discipline, traditionalism, workaholism
Emotional dysregulation vs. stability	Affective lability, alienation, angry hostility, anticipatory worry, anxiousness, dependency, depressiveness, dysphoria, emotional dysregulation, fear of uncertainty, hostility, hypochondriasis, identify problems, inferiority, introspection, irritability, negative affect, pessimism, self-acceptance, self-consciousness, self-harm, sensitivity, stress reaction, unhappiness, vulnerability, worthlessness
Unconventionality vs. closedness to experience	Absorption, dissociation, eccentric perceptions, eccentricity, openness to experience, perceptual cognitive distortion, rigidity, spiritual acceptance, thought disorder, transpersonal identification

released to the public on the DSM5.org Web site early in 2010. Based on the issues raised earlier regarding categorical PD models, the Personality and Personality Disorders (PPD) Work Group, which includes a diverse group of psychologists and psychiatrists with PD expertise, has proposed a fairly radical reformulation of the PD domain that includes, in part, a dimensional trait system. Ongoing field trials might lead to changes from what has been proposed, but the current proposal includes four primary components for PD classification: (a) a new general definition of PD focused on deficits in personality functioning and elevated pathological traits, (b) a five-tier dimensional scheme for describing personality-related functioning and impairment, (c) five PD prototypes that are based on *DSM–IV* disorders deemed worthy of retention by the work group, and (d) a trait system characterized by six broad, higher order personality trait domains (Negative Emotionality, Introversion, Antagonism, Disinhibition, Compulsivity, and Schizotypy), into which 37 specific trait facets are organized.

Although the proposed list of traits seems to be reasonably representative of the PD domain, the origins of this particular set of trait facets have not been described in much detail. The rationale for the selected lower order traits, as posted on the DSM5.org Web site, is limited to the following: "The proposed specific trait facets were selected as representative based on existing measures of normal and abnormal personality, as well as recommendations by experts in personality assessment. Nonetheless, the proposed trait set is provisional, and currently is being tested for its structural validity before finalizing the *DSM–V* proposal" (Clark & Krueger, 2010). However, the nature of the experts used to guide this process, as well as how these particular traits relate to the broad PD domain, are not elaborated by Clark and Krueger. Thus, although the field trials likely will result in refinement of the initial trait set, traits not included in the initial set will have no opportunity to enter the model at this later stage. Any omissions from the initial model almost certainly will be maintained in the final scheme.

THE CAT–PD PROJECT

To summarize, accumulating evidence has revealed significant problems with the current categorical framework underlying PD description in *DSM–IV,* and adopting a dimensional model of the traits underlying personality pathology can ameliorate most of these problems. Moreover, the proposed *DSM–5* criteria for PD include an explicitly dimensional system that, although not fully validated as yet, reflects a fundamental shift away from a purely categorical model. However, although reasonable consensus has emerged regarding the higher order trait structure of the personality/PD domains (e.g., Markon, Krueger, & Watson, 2005), no such consensus exists regarding the nature and number of the lower order traits to be included in a comprehensive PD trait system. In addition, most common measures of PD are quite inefficient and must be administered and scored by professional staff. To that end, we designed the Computerized Adaptive Test of Personality Disorder (CAT–PD) project to accomplish two related goals: (a) identify a comprehensive and integrative set of higher and lower order personality traits relevant to personality pathology, and (b) develop a computerized system, based on the principles of adaptive testing to measure the resultant traits efficiently. In addition to these primary goals, it is likely that the resultant CAT–PD model and measure has the potential to improve PD research and clinical work in a number of interesting ways, such as (a) providing a basis for richer etiological and treatment models of PD, (b) improving our understanding of the higher and lower order structure of PD-relevant personality traits, and (c) providing a flexible and comprehensive basis for clinical personality profile analyses.

We currently are in Phase 2 of a five-phase CAT–PD process:

1. Identification of all possible candidate traits to be assessed and development of the initial item pool.
2. Community and patient data collection, followed by scale development and refinement.
3. Calibration of the final item sets using item response theory (IRT) and computerized adaptive testing (CAT) simulation studies to optimize the organization and administration of the CAT–PD.
4. Development of the CAT-PD software.
5. Construct validation of the final CAT–PD test and software.

In the remainder of this article, we provide details from Phase 1 of the project, highlight the major aims of the remaining steps, and describe briefly the potential benefits of IRT and CAT for the CAT–PD project and for personality assessment more generally.

Identification of CAT–PD Candidate Traits

Given the inclusiveness and comprehensiveness of Widiger and Simonsen's (2005, 2006) summary of 18 dimensional models of PD-relevant traits, we based our construct development efforts on tapping each of the lower order dimensions listed in Table 1. To do this, we followed several steps. First, we sorted and combined these traits based on obvious redundancies in the list. Second, we reviewed the literature related to each candidate trait dimension, and each existing model and PD measure, so as to hone the list further and develop operational definitions for each trait. In general, we strived to develop operational definitions that tapped all central features of each trait (i.e., building content validity into the definitions). This process resulted in an initial list of 53 candidate traits, organized into five broad domains similar to those proposed by Widiger and Simonsen: (a) negative emotionality, (b) positive emotionality, (c) antagonism, (d) (dis)constraint, and (e) oddity.

Although the research team had considerable expertise to inform the selection and conceptualization of the traits to this point, we next opened the process up to external expert review to ensure that the candidate trait dimensions were broadly representative of the PD domain and that the results were not unduly influenced by the particular biases of the research team. To that end, we solicited feedback from a diverse sample of 28 personality and PD experts, who were selected on the basis of their published contributions to either the personality or PD literatures. Experts were drawn from academic psychology (79%) and psychiatry (21%) settings, and all held either PhD (93%) or MD (7%) degrees. They were drawn primarily from the United States (92%) and described themselves as mostly White (93%) and male (89%). Experts reported a mean of 18 years since their terminal degree ($SD = 14$), which suggests that they had ample experience to inform their work on this task. The experts expressed a range of opinions regarding the way PDs should be described in the next *DSM,* with 75% arguing for a trait dimensional system of some kind and 18% arguing for a system resembling the current *DSM–IV* model. Experts were recruited via telephone, e-mail, or both; visited a password-protected Web site that guided them through the task; and were paid $100 as compensation. Specifically, experts were asked to (a) rate the *relevance* and *representativeness* (i.e., content validity; Haynes, Richard, & Kubany, 1995) of each candidate trait with respect to personality pathology, (b) identify deficiencies in the set of candidate traits, and (c) evaluate the accuracy of the operational definitions that we developed for each trait.

The numerical ratings and open-ended feedback provided by the experts were then considered and integrated by the research team, a process that resulted in a modified set of 59 candidate dimensions and operational definitions. This revised set of candidate traits appears in Table 2, along with a rational mapping of our candidate CAT–PD traits onto the 37 traits proposed by the PPD Work Group. Interestingly, our set of candidate traits appears to cover all of the dimensions in the proposed *DSM–5* scheme. Notably, the CAT–PD set includes multiple traits relevant to 13 *DSM–5* traits, as well as six candidate traits not modeled in the *DSM–5* list. Thus, the CAT–PD candidate traits appear to be more differentiated and somewhat broader in scope than those proposed by the PPD Work Group. Of course, structural analyses following data collection in Phase 2 of the project will ultimately determine the exact nature and number of traits,

TABLE 2.—Summary of 59 CAT–PD candidate traits and conceptually relevant *DSM–5* personality traits.

Domain	CAT–PD Candidate Traits	Relevant *DSM–5* Traits
Negative emotionality	Anxious apprehension	Anxiousness
	Fearfulness	Anxiousness
	Depressive dysphoria	Depressivity
	Affective lability	Emotional lability
	Stress reactivity	Emotional lability
	Shame/guilt	Guilt/shame
	Low self-esteem	Low self-esteem
	Self-harm	Self-harm
	Suicidality	Self-harm
	Rejection sensitivity	Separation insecurity
	Submissiveness	Submissiveness
	Exploitability	Submissiveness
	Hypochondriasis	—
	Jealousy	—
Positive emotionality	Anhedonia	Anhedonia
	Exhibitionism	Histrionism
	Seductiveness	Histrionism
	Dramaticism	Histrionism
	Entitlement	Narcissism
	Arrogance	Narcissism
	Optimism/pessimism	Pessimism
	Emotional detachment	Restricted affectivity
	Social aloofness	Social detachment
	Social avoidance	Social withdrawal
	Romantic disinterest	Intimacy avoidance
	Lack of activity/energy	—
Antagonism	Aggression	Aggression
	Anger/irritability	Aggression
	Hostility	Hostility
	Callousness	Callousness
	Depravity	Callousness
	Social insensitivity	Callousness
	Deceitfulness	Deceitfulness
	Manipulativeness	Manipulativeness
	Conduct problems	Irresponsibility
	Oppositionality	Oppositionality
	Selfishness	Narcissism
	Blame externalization	—
	Domineering	—
(Dis)constraint	Urgency	Impulsivity
	Lack of premeditation	Impulsivity
	Undependability	Irresponsibility
	Rebellious nonconformity	Oppositionality
	Orderliness	Orderliness
	Perfectionism	Perfectionism
	Excessive achievement striving	Perfectionism
	Lack of perseverance	Perseveration
	Risk-taking/recklessness	Recklessness, risk aversion (–)
	Rigid propriety	Rigidity
	Lack of concern for consequences	Risk aversion (–)
	Excitement seeking	—
Oddity	Cognitive dysregulation	Cognitive dysregulation
	Absorption	Dissociation proneness
	Obliviousness	Distractibility
	Peculiarity/oddity	Eccentricity
	Suspiciousness	Suspiciousness
	Cynicism	Suspiciousness
	Magical thinking	Unusual beliefs
	Perceptual aberrations	Unusual perceptions

Note. CAT–PD = Computerized Adaptive Test of Personality Disorder; *DSM–5* = *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.).

as well as the overall structure of the higher and lower order dimensions.

Development of the Initial Item Pool

Item-pool development was guided by the principles of *substantive validity* as elaborated by Loevinger (1957) and rearticulated by others (e.g., Clark & Watson, 1995; Simms & Watson, 2007). Most notably, the initial pool was developed to be overinclusive and representative of the operational definitions written for each candidate trait. We began with the International Personality Item Pool (IPIP; Goldberg, 1999; Goldberg et al., 2006)—a broad, public-domain collection of 2,413 personality items—as the foundation on which to base the CAT–PD item pool. IPIP items tap a wide variety of constructs, and scales have been developed to serve as proxies for a number of well-known structural models (e.g., Big Three, Big Five, and Big Seven) and for a wide assortment of measures of personality traits, PDs, and other forms of psychopathology. Its large size and public-domain status made the IPIP an ideal starting point for our pool. Although not developed explicitly to measure abnormal-range personality features, the IPIP includes numerous scales tapping traits with high or low ends that are relevant to PD. However, because most IPIP items were written to tap normal-range variation in personality, we knew a priori that we would need to develop additional items for our pool to tap PD-relevant traits that are underrepresented in the IPIP, as well as to extend the measurement range to include items reflective of the pathological extremes of each candidate trait.

We implemented an iterative rational strategy to place IPIP items into the 59 CAT–PD candidate trait "bins." Eleven trained graduate and undergraduate research assistants (RAs) completed each stage of the sorting process. In the first stage, the 2,413 IPIP items were sorted into the five broad CAT–PD domains. RAs were trained on the definitions for each domain and used a computer spreadsheet to sort each item into one or more domains. Items that were sorted into a domain by at least five RAs were provisionally assigned to that domain. At this stage, items could be assigned to multiple domains. In the second sorting stage, similar procedures were used to sort items within each domain into the lower order trait bins relevant to that domain, as listed in Table 2. This stage was completed over a series of 5 weeks, each of which was devoted to a single domain. For each domain, RAs were trained on the definitions for the traits and were given 7 days to sort the assigned items into the relevant lower order trait bins. Additional sorting stages were implemented to identify additional relevant items from established IPIP scales, review all unsorted (i.e., "leftover") IPIP items for potentially useful additions, and eliminate item overlap across trait bins and domains. This iterative process resulted in 1,570 IPIP items being selected for the CAT–PD initial item pool.

The research team also developed new items for each trait bin to tap underrepresented content in the IPIP and to broaden the measurement range of each dimension to include the extreme poles reflective of personality pathology. For this task, the IPIP items within each trait bin were carefully studied for gaps in coverage, and members of the research team generated new items to fill those gaps. This process resulted in more than 2,000 new items being written. The first author (L. J. Simms) then edited the team-created items to improve their readability, eliminate redundancy, and select the final set of new items to be included in the CAT-PD pool. At the end of this process, 1,019 team-created items were added to the CAT–PD initial item pool.

Taken together, the CAT–PD initial item pool includes 2,589 items to be used in the first round of data collection and scale development. To maintain reasonable consistency with previous studies using the IPIP, we adopted a modified 5-point response format for the CAT–PD ranging from *very untrue of me* to *very true of me*. A preliminary study indicated that the CAT–PD modified response format is psychometrically parallel to the original IPIP format (i.e., yields equivalent descriptive statistics, reliability, and validity). Given the large size of the item pool, a balanced incomplete block design (BIBD) was developed to facilitate data collection. A BIBD is a planned-missingness design in which each participant completes only a portion of the items. These designs come in many shapes and sizes, depending on the particular needs of a given study (Cochran & Cox, 1957). Our BIBD approach was selected to optimize the pairwise sample size for conceptually similar traits and includes several important features. First, traits and items were assigned to a series of nine blocks such that conceptually similar traits appeared in the same block (i.e., to facilitate within-block structural analyses following data collection). Second, blocks were assigned to 12 "booklets" in a completely balanced manner: Each booklet included exactly three blocks, and each block was assigned to exactly four booklets. A summary of this design is presented in Table 3.

Data Collection and Scale Development Progress and Plans

We currently are in the middle of Phase 2 of the project, the focus of which is on collecting responses to the initial CAT–PD pool from more than 1,000 community-dwelling adults and 600 current or recent psychiatric patients, and later to perform struc-

TABLE 3.—Summary of the balanced incomplete block design used in the first round of data collection.

		Blocks of Traits/Items								
Booklet	Items	1	2	3	4	5	6	7	8	9
A	886	x	x	x						
B	828	x			x	x				
C	860	x					x	x		
D	831	x							x	x
E	892		x		x		x			
F	878		x			x			x	
G	899		x					x		x
H	851			x	x					x
I	866			x		x		x		
J	862			x			x		x	
K	840				x			x	x	
L	863					x	x			x

Note. 1 = Affective lability, stress reactivity, depressive dysphoria, low self-esteem, shame/guilt, self-harm, suicidality; 2 = domineering, dramaticism, exhibitionism, exploitability, jealousy, rejection sensitivity, seductiveness, submissiveness; 3 = arrogance, callousness, deceitfulness, entitlement, manipulativeness, selfishness; 4 = absorption, anxious apprehension, fearfulness, hypochondriasis, magical thinking, obliviousness, perceptual aberrations; 5 = cognitive dysregulation, conduct problems, cynicism, depravity, oppositionality, peculiarity/oddity, rebellious nonconformity, suspiciousness; 6 = aggression, anger/irritability, blame externalization, hostility, optimism vs. pessimism, social insensitivity; 7 = excessive achievement striving, orderliness, perfectionism, rigid propriety, undependability; 8 = activity/energy, anhedonia, emotional detachment, romantic disinterest, social aloofness, social avoidance; 9 = excitement seeking, lack of concern for consequences, lack of perseverance, lack of premeditation, risk-taking/recklessness, urgency.

tural analyses of these responses to guide scale development and refinement. Community participants are being recruited via random-digit dialing and computer-assisted telephone interview procedures aimed at recruiting a representative sample of participants from the Buffalo-Niagara region of New York (matched roughly on age, gender, and ethnicity to U.S. Census data). The community study is nearly finished as of this writing. Interim structural analyses—made up mostly of within-block targeted IRT and factor analyses—will be conducted following the community study to hone the item pool and develop provisional scales. The resulting reduced item pool will be used for the patient study and will serve as the basis for scale cross-validation analyses. More information about these analyses and scale development procedures will be presented in future publications. The remaining phases of the project will be focused on building cross-platform CAT–PD software, completing real-data and Monte Carlo CAT/IRT simulation studies to guide that development process, and finally conducting a construct validation study of the CAT–PD in a new sample of 300 psychiatric patients.

The exact form of the final CAT–PD model and measure will be influenced by the analyses conducted following Phase 2 of the project. In broad strokes, however, the intent of the CAT–PD development team is to develop a computerized tool that can be useful in both applied assessment and treatment settings with patients as well as research settings in which comprehensive trait coverage of the PD domain is desired. The software will be developed to work on multiple operating systems and could ultimately be delivered via the Internet via a Web browser, depending on the complexity of the IRT model selected to calibrate the items and guide the CAT. We anticipate permitting clinicians and researchers to customize the CAT–PD software for their particular needs (e.g., setting it up to administer only certain scales, etc.), and computerized interpretive reports can be included after sufficient data have been gathered to inform their development. Finally, any costs associated with use of the software will be minimal and limited to those necessary to support the software and keep administration servers running smoothly.

Computerized Adaptive Testing

Over the past three decades, computers have been used increasingly to automate the administration, scoring, and interpretation of a wide variety of psychological measures, including tests of ability and academic achievement (e.g., Mills, 1999), neuropsychological status (e.g., Russell, 2000), vocational interests (e.g., Hansen, Neuman, Haverkamp, & Lubinski, 1997), and personality traits (e.g., Simms & Clark, 2005; Vispoel, Boo, & Bleiler, 2001). Computers provide an efficient and reliable means for delivering assessment services to clients and research participants (e.g., Butcher, 1987; Gosling, Vazire, Srivastava, & John, 2004). A number of personality and measurement researchers (e.g., Reise & Henson, 2000; Simms & Clark, 2005; Waller & Reise, 1989) have discussed how a specific form of computerized assessment—CAT—might be applied to personality tests. CAT methods originally were developed in the ability testing literature and have been implemented successfully in a number of high-stakes testing programs.

In the most basic sense, CAT permits the selection and administration of items that are individually tailored to the latent trait level of a given examinee, which can lead to substantial time savings with little or no loss of reliability or validity (Sands, Waters, & McBride, 1997; Wainer, 2000; Weiss, 1985). A typical CAT selects and administers only those items that provide the most psychometric information for each individual at a given ability or trait level, eliminating the need to present items with very low or very high endorsement probabilities given a particular examinee's trait level. For example, in a CAT version of a general arithmetic test, the computer would not administer easy items (e.g., simple addition or subtraction) once it was clear from the examinee's responses that his or her ability level exceeded that level of arithmetic skill (e.g., she or he was correctly answering trigonometry or calculus items). Applied to personality measurement, a CAT to measure, for example, trait aggression would not administer items reflecting low or normative levels of anger (e.g., "I am a person who gets angry sometimes") once the examinee endorses items reflecting higher trait aggression (e.g., "I'm the type of person who gets into lots of fistfights").

A typical CAT includes three basic elements: (a) a procedure for estimating the examinee's latent trait level, (b) a procedure for selecting items from the pool, and (c) a termination rule to determine when testing may be discontinued. In practice, CAT begins with the administration of an item representative of the median trait level. The computer then scores that item, calculates a trait level estimate, and determines whether the termination rule has been satisfied. If not, the computer administers a new item that provides maximum information at the newly calculated trait level, scores the item, reestimates the trait level, and determines whether the termination rule has been satisfied. This iterative cycle continues until the termination rule has been satisfied.

How does the computer know how much psychometric information a given item provides at different levels of a trait? The ability of a CAT application to work efficiently depends on its capacity to calibrate items properly. To do this, CATs typically are built on a foundation laid by IRT, which includes a variety of related psychometric models that characterize each test item by one or more parameters (Hambleton & Swaminathan, 1985; Lord, 1980). Although a complete treatment of IRT is well beyond the scope of this article, we present some basic details here to aid readers in understanding the method and its application to CAT. Interested readers who want more details are referred to contemporary texts on IRT that are geared toward psychologists (e.g., Embretson & Reise, 2000).

An important strength of IRT for CAT applications is that item characteristics can be combined into a single index—item information—that describes how precisely an item measures the trait at various points along the trait continuum. Item information is a function of an item's discrimination ability (i.e., the "a" parameter; akin to the item's factor loading) and its difficulty or severity (i.e., the "b" parameter), and it typically is represented graphically in an *item information curve* (IIC). On an IIC, item information is plotted as a function of trait level (theta or θ). An IIC has its peak at the difficulty or severity level of an item, and the relative height of its peak is related to the item's discrimination ability. Given these properties, CATs use information to administer only those items that provide maximum information (i.e., are most precise markers of the trait) given the currently estimated level of the trait (Weiss, 1985).

Figure 1 shows the information provided by four hypothetical test items on a particular trait dimension. Consider a CAT in which the current trait estimate is $\theta = 1.0$ (this value can be interpreted similar to a *z* score). In a typical *maximum-information*

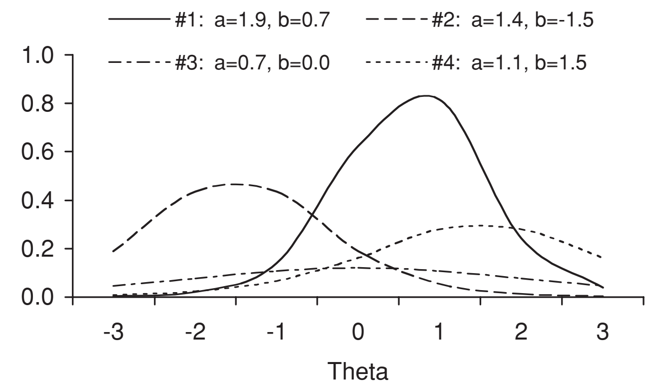

FIGURE 1.—Several prototypical item information curves. *Note.* a = discrimination; b = difficulty/severity.

item-selection strategy, assuming the trait estimate does not change markedly after each item response, the computer would present the items in the order 1, 4, 3, 2, based on the height of each IIC at the point where $\theta = 1.0$. If, however, the trait estimate was located at $\theta = -1.0$, the items would be presented in the order 2, 1, 3, 4. As mentioned earlier, another important concept in CAT is the *termination rule.* One such rule might be to stop presenting new items when reasonably informative items no longer exist in the pool. For example, Figure 1 suggests that Items 2 and 4, respectively, might not have been administered in the first and second examples earlier, because they offer negligible amounts of psychometric information at those levels of θ. Other types of termination rules are possible, but in general termination rules limit the number of items administered; marked CAT efficiency gains are possible if the item pool for a particular scale is sufficiently broad and large.

Traditional IRT models were limited to dichotomous items (e.g., true–false) from unidimensional scales. Fortunately for the CAT–PD and the field of personality assessment more generally, IRT has evolved to include (a) models for which patients are asked to rate items along a gradient of severity, frequency, or agreement (e.g., Muraki, 1990; Samejima, 1969); and (b) constructs that deviate from strict unidimensionality (e.g., Gibbons et al., 2007; Reckase, 1997).

Use of CAT for Personality and Psychopathology Assessment

Despite the widespread use of IRT and, to a lesser extent, CAT in the ability testing literature, relatively few applications of IRT-based CAT have appeared in the personality and mental health literatures, likely due to a number of factors, including (a) the greater statistical complexity of IRT/CAT compared to tests built using classical test theory, (b) the lack of user-friendly IRT/CAT software packages, and (c) the lack of consistent training in IRT/CAT methods in clinical assessment curricula. Of the limited attempts to apply IRT-based CAT in the personality literature, most of those have been based on post-hoc simulations using previously collected response data, rather than tests with live participants. For example, Waller and Reise (1989) simulated a CAT version of the Absorption scale of the Multidimensional Personality Questionnaire. Real-data CAT simulations, based on responses from 1,000 participants who previously had completed the Absorption scale in the traditional paper-and-pencil format, yielded item savings ranging from 50% to 75%, depending on the termination rule utilized. In a similar demonstration, Kamakura and Balasubramanian (1989) found item savings ranging from 60% to 66% on the Socialization scale of the California Psychological Inventory. Reise and Henson (2000) extended the personality CAT literature to multiscale batteries, conducting real-data simulations on the 30 facet scales of the Revised NEO Personality Inventory (NEO PI–R). Using a polytomous IRT model to account for the 5-point Likert scale used to rate NEO PI–R items, they achieved average item savings of 50% per facet.

Although such simulation studies have been useful in establishing that CAT methodology can be applied effectively to personality and psychopathology constructs, live-testing studies also are important to establish the ecological validity of the technique. Simms and Clark (2005) were the first to develop a prototype CAT of personality and personality pathology that was examined in a live-testing study. Based on Clark's (1993) SNAP, Simms and Clark demonstrated that CAT methods could be effectively used for a PD assessment. In their study, the SNAP–CAT yielded significant time savings (ranging from 58%–60%) over the traditional SNAP administered using paper and pencil or computer; importantly, descriptive statistics, test–retest stability, internal factor structure, and validity patterns largely were comparable across administration modes. Moreover, participants preferred the computerized version to the paper-and-pencil version.

Unfortunately, in both the simulation studies described earlier and the SNAP–CAT study, efficiency gains were achieved at the expense of small but statistically significant losses in reliability and validity, which is an inevitable result of using an existing traditional personality measures with scales that do not include adequate information at all levels of the underlying traits. Notably, traditional scale development statistical procedures (e.g., factor analysis) favor items with moderate endorsement rates; thus, extreme items (i.e., those indicative of the pathological poles of personality traits) often are inadvertently tossed out because they tend to yield much weaker factor loadings. As such, CATs for personality pathology are needed that are built from the ground up to include much broader and larger item pools, such that all relevant levels of each dimension to be measured are represented adequately. When item pools are sufficiently broad and large, CATs can yield equivalent or better reliability and validity compared to traditional tests, with fewer items, by focusing test administration only on those items that are relevant to a given patient. Thus, the IRT-based CAT methods underlying the CAT–PD project are poised to improve the efficiency of PD measurement without any loss of reliability or validity.

Indeed, because of these features of CAT, several groups funded by the National Institutes of Health (NIH) currently are working to build CAT measures for use in various physical and mental health domains. In one such application, Gibbons et al. (2008) have been working to build a comprehensive CAT measure of depressive symptomatology. Likewise, in a large, multisite, NIH-funded effort, the PROMIS group (i.e., Patient Reported Outcomes Measurement Information System) have been developing CAT item banks for a variety of common health outcomes, such as emotional distress, chronic pain, sleep disturbance, and arthritis (Cella et al., 2010). Thus, important work is ongoing to use IRT and CAT techniques in health care settings; the CAT–PD project is an example of this growing trend.

Summary and Conclusions

PDs are prevalent in the population and associated with significant functional impairment and a complicated course of treatment in psychiatric settings. However, numerous concerns—such as psychometric problems associated with the current *DSM–IV* PDs as well as the time and resources needed to administer, score, and interpret most current measures of PD—have led to a decrease in structured personality assessment in resource-limited research and applied settings. Various trait-based dimensional models and measures have been proposed in recent decades as alternatives to the current categorical approach to PD description and assessment. Such models have been shown to have numerous advantages over category or prototype-based approaches. However, no single model encompasses all prominent PD-related traits, and there remains a lack of consensus as to the lower order structure of PD-related traits. The CAT–PD project was designed to solve these interrelated concerns by (a) developing a comprehensive and integrative model of PD-related trait dimensions, and (b) creating an efficient method for measuring those traits. We assumed that a comprehensive model of PD traits would yield too large and unwieldy an instrument to be of much practical value in applied or research settings. Thus, we have elected to adopt CAT as a measurement method, which can be expected to result in significant efficiency gains over traditional questionnaire and interview methods.

The results of Phase 1 of the project have revealed 59 candidate traits organized into five broad factors corresponding to the Big Four plus oddity. These traits and the organizational scheme are consistent with the integrative work of Widiger and Simonsen (2005, 2006) and also have been influenced by feedback solicited from a large group of personality and PD experts. Moreover, the CAT–PD candidate traits are well positioned to tap the trait system recently proposed by the *DSM–5* PPD Work Group, as the CAT–PD list includes all of the dimensions on the provisional *DSM–5* list. However, a possible limitation of the CAT–PD project is its reliance on self-report methodology in the development and implementation of the new measure. Given previous work showing blind spots for certain aspects of PD (e.g., Oltmanns & Turkheimer, 2006), it will be important to extend the initial CAT–PD project to develop an informant version once the basic CAT–PD model and measure are finished and validated.

Regardless, the use of CAT as a basis for mental and physical health measurement has shown promise as a method to efficiently measure a broad range of symptoms, features, or traits with little or no loss to measurement precision or validity. The CAT–PD will continue the trend toward sophisticated measurement systems for patient-reported problems.

Acknowledgments

We thank William Calabrese, Monica Rudick, Wern How Yam, Tom Yufik, and Kerry Zelazny for their help in coordinating the expert review study and the development of the initial item pool. In addition, we thank the experts who provided feedback on the candidate traits as well as the research assistants who sorted items into candidate trait bins. Finally, we thank Lee Anna Clark for her consultation during the early stages of the project.

The studies presented in this article were supported by a research grant to the L. J. Simms from the National Institute of Mental Health (No. R01MH080086).

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.) Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlashan, T. H., . . . Gunderson, J. G. (2001). Treatment utilization by patients with personality disorders. *American Journal of Psychiatry, 158,* 295–302.

Butcher, J. N. (1987). The use of computers in psychological assessment: An overview of practices and issues. In J. N. Butcher (Ed.), *Computerized psychological assessment: A practitioner's guide* (pp. 292–324). New York, NY: Basic Books.

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., . . . Hays, R. D., on behalf of the PROMIS Cooperative Group. (2010). Initial item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS) network: 2005–2008. *Journal of Clinical Epidemiology, 63*, 1179–1194.

Clark, L. A. (1993). *Schedule for nonadaptive and adaptive personality (SNAP): Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology, 58*, 227–257.

Clark, L. A., & Krueger, R. F. (2010). *Rationale for a six-domain trait dimensional diagnostic system for personality disorder*. Retrieved from http://www.dsm5.org/ProposedRevisions/Pages/RationaleforaSix-DomainTraitDimensionalDiagnosticSystemforPersonalityDisorder.aspx

Clark, L. A., Livesley, W. J., & Morey, L. (1997). Personality disorder assessment: The challenge of construct validity. *Journal of Personality Disorders, 11*, 205–231.

Clark, L. A., Simms, L. J., Wu, K. D., & Casillas, A. (in press). *Schedule for nonadaptive and adaptive personality-second ed. (SNAP–2): Manual for administration, scoring, and interpretation*.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319.

Clark, L. A., Watson, D., & Reynolds, S. (1995). Diagnosis and classification of psychopathology: Challenges to the current system and future directions. *Annual Review of Psychology, 46*, 121–153.

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs* (2nd ed.). New York, NY: Wiley.

Cyranowski, J. M., Frank, E., Winter, E., Rucci, P., Novick, D., Pilkonis, P., . . . Kupfer, D. J. (2004). Personality pathology and outcome in recurrently depressed women over 2 years of maintenance interpersonal psychotherapy. *Psychological Medicine, 34*, 659–669.

Dolan, B., Evans, C., & Norton, K. (1995). Multiple Axis-II diagnoses of personality disorder. *British Journal of Psychiatry, 166*, 107–112.

Dreessen, L., & Arntz, A. (1998). Short-interval test–retest interrater reliability of the Structured Clinical Interview for *DSM–III–R* Personality Disorders (SCID–II) in outpatients. *Journal of Personality Disorders, 12*, 138–148.

Eaton, N. R., Krueger, R. F., South, S. C., Simms, L. J., & Clark, L. A. (in press). Contrasting prototypes and dimensions in the classification of personality pathology: Evidence that dimensions, but not prototypes, are robust. *Psychological Medicine.*

Ekselius, L., Tillfors, M., Furmark, T., & Fredrikson, M. (2001). Personality disorders in the general population: *DSM–IV* and *ICD–10* defined prevalence as related to sociodemographic profile. *Personality and Individual Differences, 30*, 311–320.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Feske, U., Mulsant, B. H., Pilkonis, P. A., Soloff, P., Dolata, D., Sackeim, H. A., & Haskett, R. F. (2004). Clinical outcome of ECT in patients with major depression and comorbid borderline personality disorder. *American Journal of Psychiatry, 161*, 2073–2080.

Fossati, A., Maffei, C., Bagnato, M., Battaglia, M., Donati, D., Donini, M., . . . Prolo, F. (2000). Patterns of covariation of *DSM–IV* personality disorders in a mixed psychiatric sample. *Comprehensive Psychiatry, 41*, 206–215.

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., . . . Stover, A. (2007). Full-information bi-factor analysis of graded response data. *Applied Psychological Measurement, 31*, 4–19.

Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., . . . Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59*, 361–368.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? *American Psychologist, 59*, 93–104.

Grant, B. F., Hasin, D. S., Stinson, F. S., Dawson, D. A., Chou, S. P., Ruan, W. J., & Huang, B. (2005). Co-occurrence of 12-month mood and anxiety disorders in the US: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Psychiatric Research, 39*, 1–9.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff.

Hansen, J. C., Neuman, J. L., Haverkamp, B. E., & Lubinski, B. R. (1997). Comparison of user reaction to two methods of Strong Interest Inventory administration and report feedback. *Measurement & Evaluation in Counseling & Development, 30*, 115–127.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238–247.

Kamakura, W. A., & Balasubramanian, S. K. (1989). Tailored interviewing: An application of item response theory for personality measurement. *Journal of Personality Assessment, 53*, 502–519.

Livesley, W. J., & Jackson, D. N. (2009). *Dimensional assessment of personality pathology—Basic questionnaire: Technical manual*. Port Huron, MI: Sigma Assessment Systems.

Livesley, W. J., & Jang, K. L. (2000). Toward an empirically based classification of personality disorder. *Journal of Personality Disorders, 14*, 137–151.

Livesley, W. J., Jang, K. L., & Vernon, P. A. (1998). The phenotypic and genetic structure of traits delineating personality disorder. *Archives of General Psychiatry, 55*, 941–948.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635–694.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*, 139–157.

Miller, J. D., Maples, J., Few, L. R., Morse, J. Q., Yaggi, K. E., & Pilkonis, P. A. (2010). Using clinician-rated five-factor model data to score the *DSM–IV* personality disorders. *Journal of Personality Assessment, 92*, 296–305.

Mills, C. N. (1999). Development and introduction of a computer adaptive Graduate Record Examinations General Test. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 117–135). Mahwah, NJ: Erlbaum.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14*, 59–71.

Oldham, J. M., & Skodol, A. E. (2000). Charting the future of Axis II. *Journal of Personality Disorders, 14*, 17–29.

Oldham, J. M., Skodol, A. E., Kellman, H. D., Hyler, S. E., Doidge, N., Rosnick, L., & Gallaher, P. E. (1995). Comorbidity of Axis I and Axis II disorders. *American Journal of Psychiatry, 152*, 571–578.

Oltmanns, T. F., & Turkheimer, E. (2006). Perceptions of self and others regarding pathological personality traits. In R. F. Krueger & J. L. Tackett (Eds.), *Personality and psychopathology* (pp. 71–111). New York, NY: Guilford.

Pilkonis, P. A., Heape, C. L., Proietti, J. M., Clark, S. W., McDavid, J. D., & Pitts, T. E. (1995). The reliability and validity of two structured diagnostic interviews for personality disorders. *Archives of General Psychiatry, 52*, 1025–1033.

Pincus, A., & Gurtman, M. B. (2006). The interpersonal circumplex and the interpersonal theory: Perspectives of personality and its pathology. In S. Strack (Ed.), *Differentiating normal and abnormal personality* (2nd ed., pp. 83–112). New York, NY: Springer.

Piotrowski, C. (1999). Assessment practices in the era of managed care: Current status and future directions. *Journal of Clinical Psychology, 55*, 787–796.

Piotrowski, C., Belter, R. W., & Keller, J. W. (1998). The impact of "managed care" on the practice of psychological testing: Preliminary findings. *Journal of Personality Assessment, 70*, 441–447.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*, 25–36.

Reich, J. (2003). The effects of Axis II disorders on the outcome of treatment of anxiety and unipolar depressive disorders: A review. *Journal of Personality Disorders, 17*, 387–405.

Reise, S. P., & Henson, J. M. (2000). Computerization and adaptive administration of the NEO-PI-R. *Assessment, 7*, 347–364.

Russell, E. W. (2000). The application of computerized scoring programs to neuropsychological assessment. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (pp. 483–515). Mahwah, NJ: Erlbaum.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monographs, 34*(4, Pt. 2, Whole No. 17).

Samuels, J., Eaton, W. W., Bienvenu, O. J., Brown, C. H., Costa, P. T., Jr., & Nestadt, G. (2002). Prevalence and correlates of personality disorders in a community sample. *British Journal of Psychiatry, 180*, 536–542.

Sands, W. A., Waters, B. K., & McBride, J. R. (1997). *Computerized adaptive testing: From inquiry to operation*. Washington, DC: American Psychological Association.

Shedler, J., & Westen, D. (2004). Refining personality disorder diagnosis: Integrating science and practice. *American Journal of Psychiatry, 161*, 1350–1365.

Simms, L. J., & Clark, L. A. (2005). Validation of a computerized adaptive version of the Schedule for Nonadaptive and Adaptive Personality. *Psychological Assessment, 17*, 28–43.

Simms, L. J., & Watson, D. (2007). The construct validation approach to personality scale construction. In R. Robins, C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 240–258). New York, NY: Guilford

Skodol, A. E., Oldham, J. M., Bender, D. S., Dyck, I. R., Stout, R. L., Morey, L. C., . . . Gunderson, J. G. (2005). Dimensional representations of *DSM–IV* personality disorders: Relationships to functional impairment. *American Journal of Psychiatry, 162*, 1919–1925.

Smith, T. L., & Benjamin, L. S. (2002). The functional impairment associated with personality disorders. *Current Opinion in Psychiatry, 15*, 135–141.

Torgersen, S., Kringlen, E., & Cramer, V. (2001). The prevalence of personality disorders in a community sample. *Archives of General Psychiatry, 58*, 590–596.

Tyrer, P. (2005). The problem of severity in the classification of personality disorders. *Journal of Personality Disorders, 19*, 309–314.

Vispoel, W. P., Boo, J., & Bleiler, T. (2001). Computerized and paper-and-pencil versions of the Rosenberg Self-Esteem Scale: A comparison of psychometric features and respondent preferences. *Educational and Psychological Measurement, 61*, 461–474.

Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Erlbaum.

Waller, N. G., & Reise, S. P. (1989). Computerized adaptive personality assessment: An illustration with the Absorption scale. *Journal of Personality and Social Psychology, 57*, 1051–1058.

Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology, 53*, 774–789.

Westen, D., Shedler, J., & Bradley, R. (2006). A prototype approach to personality disorder diagnosis. *American Journal of Psychiatry, 163*, 846–856.

Widiger, T. A. (1993). The *DSM–III–R* categorical personality disorder diagnoses: A critique and an alternative. *Psychological Inquiry, 4*, 75–90.

Widiger, T. A., & Clark, L. A. (2000). Toward *DSM–V* and the classification of psychopathology. *Psychological Bulletin, 126*, 946–963.

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders, 19*, 110–130.

Widiger, T. A., & Simonsen, E. (2006). Alternative dimensional models of personality disorder: Finding a common ground. In T. A. Widiger, E. Simonsen, P. J., Sirovatka, & D. A. Regier (Eds.), *Dimensional models of personality disorders: Refining the research agenda for DSM–V* (pp. 1–21). Washington, DC: American Psychiatric Association.

Widiger, T. A., & Trull, T. J. (2007). Plate techtonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62*, 71–83.

World Health Organization (1992). *ICD–10 Classifications of Mental and Behavioural Disorder: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: Author.

Yates, B. T., & Taub, J. (2003). Assessing the costs, benefits, cost-effectiveness, and cost-benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment, 15*, 478–495.

Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., . . . Gunderson, J. G. (2000). The Collaborative Longitudinal Personality Disorders Study: Reliability of Axis I and II diagnoses. *Journal of Personality Disorders, 14*, 291–299.

Zimmerman, M., Rothschild, L., & Chelminski, I. (2005). The prevalence of *DSM–IV* personality disorders in psychiatric outpatients. *American Journal of Psychiatry, 162*, 1911–1918.