JOURNAL OF PERSONALITY ASSESSMENT, 1992, 59(2), 223-238

Reliability and Validity of the Coolidge Axis II Inventory: A New Inventory for the Assessment of Personality Disorders

Frederick L. Coolidge and Michelle M. Merwin
University of Colorado at Colorado Springs

This group of studies describes the development of a 200 item, self-report, 4-point true–false inventory (Coolidge Axis II Inventory [CATI]) to assess personality disorders according to the criteria established in the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev. [*DSM–III–R*]; American Psychiatric Association, 1987). The 13 personality disorder scales of the CATI had a mean test–retest reliability of .90 and a median internal consistency (Cronbach's alpha) of .76. There was a 50% concordance rate with clinician's diagnosis for 24 personality disordered patients. The median concurrent validity (raw score sums) between the CATI and the Millon Clinical Multiaxial Inventory–II for the 13 personality disorder scales was .58. Preliminary studies also support the reliability and validity of Depression, Anxiety, and Brain Dysfunction scales.

With the change to a multiaxial classification system in the American Psychiatric Association's (1980) *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed. [*DSM–III*]), the category of personality disorders received renewed attention. Although clinical syndromes were placed on Axis I, the personality disorders were placed on a separate axis, Axis II. Unlike the previous edition of the diagnostic manual, the personality disorders in the *DSM–III* also have specific criteria listed for each disorder, and the patients have to meet a specified number of standards to meet the criterion for a disorder.

Millon (1977) created one of the first self-report inventories for the assessment of personality disorders, the Millon Clinical Multiaxial Inventory (MCMI). Widiger, Williams, Spitzer, and Francis (1985) criticized the MCMI on a number of grounds including the issue of whether the questions from the MCMI adhered to the diagnostic criteria in the *DSM–III*. Millon (1985) responded to these

criticisms with the publication of the MCMI–II (Millon, 1987). The relatively high cost of the MCMI–II and the initial lack of hand-scoring mechanisms apparently reduced the MCMI–II's widespread acceptance. In addition, the MCMI–II is inappropriate when the patient is not suspected of having a personality disorder; for example, the MCMI–II is not useful in cases of differential diagnoses between adjustment reactions and personality disorders.

The American Psychiatric Association revised the diagnostic classification system again in 1987 with the publication of the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev. [*DSM–III–R*]). In addition to the original 11 personality disorders, 2 new personality disorders were added into the appendix under the heading, "Proposed Diagnostic Categories Needing Further Study" (p. 367). They are the sadistic and self-defeating personality disorders.

Hyler et al. (1988) created the Personality Diagnostic Questionnaire (PDQ), a self-report personality disorder inventory based on the *DSM–III*, which contains 163 true–false items. In a preliminary study with 552 patients with and without suspected personality disorders, Hyler claimed that a single derived score could be used to differentiate those highly likely to have a personality disorder from those less likely to have a personality disorder.

Morey, Waugh, and Blashfield (1985) derived 11 personality disorder scales from the MMPI. They chose the items for the scales through clinical judgment and subsequent reliability analyses. There were no validity studies. Streiner and Miller (1988) analyzed these 11 scales using the MCMI on 74 psychiatric patients. They found that the correlations were modest at the very best, and in many cases there were significant negative correlations among the corresponding scales. They concluded that, although the MCMI should not be considered a "gold standard," the MMPI personality scales should not be used to determine personality disorders without validation. In addition, Streiner and Miller concluded that the MMPI personality disorder scales may measure entities other than what they purport to measure.

In this article, we report on the development of a new self-report questionnaire, the Coolidge Axis II Inventory (CATI; Coolidge, 1984). The CATI was originally designed to assess the 11 personality disorders on Axis II of the *DSM–III*, and it has been revised to conform with the changes in criteria on Axis II in the *DSM–III–R*. The earlier version of the CATI has been used in clinical populations, such as the morbidly obese (Grana, Coolidge, & Merwin, 1989), and in senile dementia (Coolidge, Bracken, Taylor, Smith, & Peters, 1985). The purpose of this research was to establish the reliability and validity of the current version of the CATI.

METHOD

Description of the CATI

The 13 personality disorder scales. The CATI contains 200 questions that are answered on a 4-point true–false scale ranging from *strongly false* (1) to *strongly*

true (4). It takes approximately 45 min to 1 hr to complete. The questions were created almost directly from the 117 unique criteria from the 11 personality disorders on Axis II of the *DSM–III–R* plus the 2 personality disorders from Appendix A of the *DSM–III–R*. Thus, each criterion of the 117 is represented by at least one question on the CATI. Additional questions were created when a *DSM–III–R* criterion covered more than one issue or when the criterion was ambiguous. For example, Criterion 8 of the histrionic personality disorder states, "has a style of speech that is excessively impressionistic and lacking in detail." Thus, the CATI contains the following two statements: "I have been told that my style of speech is strange or vague," and "People do not understand what I am trying to say." In order to control for continuous response denial, 38 questions are scored as pathological if they are answered in the false direction. The questions have undergone numerous revisions to ensure their face validity with respect to the *DSM–III–R* criteria.

Each of the 13 scales on the CATI for the 13 personality disorders also may contain questions from the main descriptive features section and from the associated features section in *DSM–III–R* for each of the disorders. Generally, these questions reflect or reinforce one of the specific criteria listed. The number of questions per scale varies from 45 questions that cover 22 criteria on the Antisocial scale to 16 questions covering 8 criteria on the Avoidant, Sadistic, and Self-Defeating scales. Questions that appear on more than one scale were kept to a minimum. Thus, 92 of the 185 personality disorder scale questions appear on one scale, 56 questions appear on two scales, 26 questions appear on three scales, 7 questions appear on four scales, and 4 questions appear on five scales.

Critical Items

Four statements were written to assess drug or alcohol abuse problems, current suicidal ideation, and thinking problems; and they are as follows: "Someone I know thinks I have a drug or alcohol problem," "I have gotten into trouble because of a drinking or drug problem," "I think that there is something wrong with my mind," and "Recently I have felt like killing myself."

Validity Scales

There are two validity scales on the CATI. There is a three-item validity scale designed to assess whether the person read and understood the questions or whether they were randomly responding. On these questions, it is highly unlikely that anyone would not answer a strongly false. They are as follows: "I played quarterback for the Denver Broncos," "I was a member of the French Foreign Legion," and "I swam the English Channel."

A second validity scale of 21 items was empirically derived from the 200 items through a series of studies (Mitton, 1991) to discriminate between those who are

intentionally trying to present themselves in a good light and deny emotional difficulties from those who are intentionally trying to look bad or pathological.

The CATI also contains a 71-item scale (called the Adjustment Index) that purports to measure a patient's overall psychopathology. The items were chosen to represent all 13 personality disorder scales plus the Depression and Anxiety scales.

Axis I Scales

Three Axis I scales were also developed: Depression, and Anxiety, and Brain Dysfunction. The Depression and Anxiety scales were included because they are the two most commonly diagnosed psychopathological syndromes. In addition, the 117 unique criteria of *DSM–III–R* for the 13 personality disorder scales already contain a substantial proportion of the essence of these two scales. The Brain Dysfunction scale was included because of the increasing role that neuropsychological development and brain injury play in the ultimate expression of the personality, both as a cause and as a result.

Brain Dysfunction scale. Fifteen questions were created on a theoretical basis from the clinical neuropsychology literature. Some questions were chosen because the respective symptom commonly occurs after head injury (Binder, 1986) or general cerebral dysfunction (Lezak, 1983). Examples of these questions are: "I think my memory has gotten worse in the past few years," "I tend to forget things I am supposed to do," and "I have trouble tying to remember the names of common objects." Other questions were chosen based on their lower sensitivity to brain dysfunction but greater specificity. Examples of the questions are: "I have problems with my balance," and "I have noticed a change in my sense of taste or smell." With respect to this latter question, Costanzo and Becker (1986) estimated that about 20% to 25% of head injured patients will have an alteration of their sense of taste or smell; however, it remains a somewhat rare symptom in inorganic psychological problems.

Depression scale. The Depression scale was created using 22 items derived by Lucero (1989). A larger set (than the 22 items) of questions was originally chosen from the 200 questions based on the questions' theoretical consistency with the concept of depression. One question of the 22 items was created specifically to test for current suicidal ideation, for example, "Recently I have felt like killing myself."

Anxiety scale. The Anxiety scale consisting of 27 items was empirically derived from the 200 questions in Hosman's (1989) study. As is true of the depression scale, a larger set of items was initially chosen based on their theoretical consistency with the general concept of anxiety.

Test Instructions

The written instructions in the test booklet say that the questions are to help "describe yourself as you see yourself. Some statements will seem strongly false and some statements will seem strongly true. Other statements will seem somewhere between the extremes and you are to choose whether they are more false than true or more true than false."

The subjects or patients are encouraged not to omit any answers and to do their best to find the answers that most appropriately describe them, even if a statement does not exactly apply to them.

RESULTS

Test–Retest Reliability

In a previous study (Merwin & Coolidge, 1987), 39 college students (M age = 21; all were White) from an introductory psychology course were recruited and took the CATI. One week later, they were recontacted and asked to take the test again. They were instructed to take the test under the same instructions that they took it initially (i.e., try to answer honestly) and to try not to second guess the experiments. For the personality disorder scales, the mean scale reliability was .90 (range = .78 for the Obsessive Compulsive scale to .97 for the Schizoid scale).

Personality Disorder Scale Reliabilities

Cronbach's (1951) alpha was derived for each 13 personality disorder scales in a study of 609 purportedly normal subjects. Subjects were recruited by college students and consisted of their friends, relatives, and acquaintances. The college students were asked to recruit people whom they deemed normal or psychologically healthy. Specifically they were told not to test anyone who had previously been in jail or prison, who had previously been committed to a mental hospital, or who was mentally retarded. The mean age of the sample was 26.7 years, and the subjects ranged in age from 16 to 78. There were 226 males and 375 women (8 subjects were excluded). Approximately 99% of the sample reported a high school education or greater, 73% had some college education, and 8% had bachelor degrees or greater. Approximately 58% of the sample were single, 37% of the sample were married or living with a significant other, and 5% were separated or divorced. Approximately 89% of the sample were White, 6% were Hispanic, and 2% were Black.

The median reliability (Cronbach's alpha) for the 13 personality disorder scales was .76, and the reliabilities for each scale are as follows: Dependent, .87;

Antisocial, .86; Avoidant, .80; Borderline, .80; Paranoid, .79; Passive Aggressive, .78; Histrionic, .76; Narcissistic, .74; Schizotypal, .73; Schizoid, .73; Sadistic, .69; Obsessive Compulsive, .68; and Self-Defeating, .66.

Scale Reliabilities of the Axis I Scales

The scale reliabilities for the three Axis I clinical scales, obtained on the same sample as previously noted, are as follows: Brain Dysfunction, .83; Depression, .88; and Anxiety, .89.

Factor Analysis of the Personality Scales

A principal components factor analysis was performed upon the 13 personality disorder scales with varimax rotation of the resulting factors. Three factors were extracted with eigenvalues above 1.00 (5.75, 2.37, and 1.60 for the three factors respectively). Factor 1 accounted for 44.3% of the total-variance, whereas Factors 2 and 3 accounted for 18.3% and 12.3%, respectively. Table 1 presents the factor loadings for the 13 scales on the three factors.

Factor 1 appears to have its most substantial loading from the Avoidant scale (.89). There were no substantial negative loadings. The primary theme of Factor 1 appears to be avoidant behavior, general maladjustment, and anxious features. In a factor analysis of the MCMI-I with 151 psychiatric inpatients, Piersma (1986) found a five-factor structure with the first factor accounting for 50% of the total variance, and the Avoidant scale had the highest loading with Factor 1. Millon (1987) also reported a general maladjustment first factor in an eight-factor structure on a general psychiatric sample on the MCMI-II, and the

TABLE 1
Principal Components Factor Analysis With Varimax Rotation of the 13 Personality Disorder Scales' Sums of the CATI: Factor Loadings

Scale	*Factor 1*	*Factor 2*	*Factor 3*
Antisocial	−.07	.90	.20
Avoidant	.89	.04	.01
Borderline	.30	.50	.62
Dependent	.72	.01	.53
Histrionic	−.18	.09	.91
Narcissistic	.28	.38	.64
Obsessive Compulsive	.75	.02	−.16
Paranoid	.56	.60	.15
Passive Aggressive	.59	.43	.45
Sadistic	.03	.85	.13
Self-Defeating	.62	.53	.19
Schizoid	.00	−.03	−.94
Schizotypal	.55	.63	−.06

first factor accounted for 31% of the total variance. However, only four of Millon's eight factors dealt with the 13 personality disorders. The other four factors appeared to deal with his nine *DSM–III–R* Axis I scales. Similar to our factor analysis results, Millon also found that the highest loadings for the Avoidant, Passive Aggressive, and Self-Defeating scales were on Factor 1. Also, the Avoidant scale had the single highest loading of all possible scales on Factor 1 in both studies (.89 in the study and .92 in Millon's study). This suggests that Factor 1 on the CATI and the MCMI–II has prominent avoidant features such as an active withdrawal from personal relationships and a conscious and anxious estrangement.

Factor 2 had its highest loadings from the Antisocial scale (.90) and sadistic scale (.85); thus, Factor 2 appears to have a strong antisocial/sadistic component. The other loadings may suggest a paranoid or suspicious theme. Piersma's (1986) study had no clear counterpart for this factor, although his Factor 3, accounting for 13% of the total variance, was labeled Impulsivity/Negativism. Millon (1987) also found that the Antisocial and Sadistic scales had their highest loadings on his Factor 2. He concluded that the factor represented acting-out and self-indulgent behaviors.

In our study, Factor 3 had its highest loadings with the Histrionic (.91) and schizoid (−.94) scales; thus, Factor 3 is interpreted as a pathological continuum of introversion and extroversion. Piersma's (1986) study had no counterpart to our Factor 3. The Schizoid and Histrionic scales in his study loaded the highest (positive and negative, respectively) on his Factor 1. There was no clear parallel factor for our Factor 3 in Millon's (1987) study.

Construct and Convergent Validity of the Personality Scales

Eleven licensed clinical psychologists (4 men and 7 women; M length of time in private practice = 12.8 years) from the local community were contacted and agreed to participate in the validation of the CATI with the MCMI–II. They were asked to select their patients whom they suspected of having personality disorders. They were required to have met with the patient for at least five sessions. They were told not to select any patients with psychotic syndromes including organic mental disorders, schizophrenia, and bipolar disorders. Informed consent was obtained from the clinicians and their patients, and the patients were administered the CATI and MCMI–II by their clinicians. The clinicians also filled out a personality disorder checklist, which allowed the clinician to note which of the 13 personality disorders they thought their patients had, and they filled out a separate symptom checklist, which contained a randomized list of the 117 unique criteria for the personality disorders from Axis II and from the appendix of the *DSM–III–R*.

The initial analysis contained 24 patients whose mean age was 38.0. There

were 3 men and 21 women. Eleven were married or living with a significant other, 8 were divorced or separated, and 5 were single. All of the patients had at least graduated from high school, and 16 had some college experience (range = high school to master's degree). According to the clinicians, personality diagnoses for the 24 patients were 2 borderline, 4 dependent, 4 histrionic, 4 passive aggressive, 2 avoidant, 1 obsessive compulsive, 1 narcissistic, and 1 patient had no primary diagnosis but had a diagnoses of obsessive compulsive and self-defeating traits. There were five patients with a suspected diagnosis of two personality disorders—borderline–dependent, borderline–narcissistic, and avoidant–dependent—and two obsessive compulsive–passive aggressive patients. The patients' primary Axis I diagnoses were as follows: 18 depressions, 4 anxiety disorders, and 2 adjustment reactions.

For the CATI, a criterion of 1 *SD* above the mean of the normative sample was chosen to mark the presence of a personality disorder. When this standard was applied to the clinical sample and matched to the diagnoses of the clinicians, a 50% concordance rate was obtained (12 of 24 patients). For the MCMI–II, at a base rate of 75 to mark the presence of a personality disorder, a 63% concordance rate was obtained (15 of 24 patients). However, the higher concordance rate of the MCMI–II was apparently attained at the cost of a higher false-positive rate. According to the clinicians, the patients averaged 1.2 primary diagnoses per person. The MCMI–II averaged 3.4 primary diagnoses (greater than or equal to a base rate of 75) per patient, whereas the CATI averaged 2.8 diagnoses ($\geq$ 1 *SD* above the mean) per patient. For the 15 patients who were diagnosed as having the same disorder by the clinicians and the MCMI–II, the average primary diagnoses per patient was 4.7 on the MCMI–II.

The raw score sums on the MCMI–II were correlated with the raw score sums on the CATI. The median convergent validity correlation was .58 for the 13 scales. The correlations were as follows: Borderline, .87; Passive Aggressive, .86; Avoidant, .80; Histrionic, .72; Self-defeating, .67; Schizotypal, .65; Paranoid, .58; Antisocial, .57; Dependent, .43; Sadistic, .40; Narcissistic, .38; Schizoid, .22; and Obsessive Compulsive, .10.

In view of the extremely low correlation between the CATI and the MCMI–II of the obsessive compulsive personality disorder on the raw score sums, a criteria analysis was performed to determine the nature of the low correlation. The CATI has 32 questions on the Obsessive Compulsive scale, with a minimum of 3 questions per criterion for the nine criteria with the exception of Criterion 8 and Criteria 9, which have two questions and one question, respectively. The MCMI–II has 38 questions. The raw scale sum is obtained through a weighing system wherein important questions are weighted by a factor of 3 and the other questions are rated by either a factor of 2 or 1. The MCMI–II manual lists ten 3-point questions on the Obsessive Compulsive scale. On the basis of their face validity, it was determined that only five of the nine criteria are covered by the ten 3-point questions, and some of the 3-point questions are written for the same

DSM–III–R criterion. Thus, a preliminary analysis suggests that the criteria coverage of the MCMI–II is incomplete. This finding is not surprising in view of the fact that Millon (1985) previously argued that the *DSM–III–R* criteria should not be considered sacrosanct and that a valid test instrument may be obtained without strict adherence to the criteria.

The base rate scores on the MCMI–II were also correlated with the raw scores on the CATI. The median convergent validity correlation was .41 for the 13 scales. The correlations were as follows: Passive Aggressive, .74; Histrionic, .73; Avoidant, .69; Obsessive Compulsive, .56; Antisocial, .53; Borderline, .46; Schizotypal, .41; Dependent, .38; Paranoid, .30; Schizoid, .26; Self-Defeating, .14; Narcissistic, −.05; and Sadistic, −.12.

For each of the 24 patients, a Spearman's rank-order correlation coefficient was obtained between the rankings of the 13 scales on the CATI and the rankings of the 13 scales on the MCMI–II. The median convergent validity correlation was .42. The correlations ranged from a high of .82 to a low of −.07 for the 24 patients.

An additional analysis was made in order to determine whether the CATI 71-item Adjustment Index would be able to differentiate between those who have personality disorders and those who do not. We found that the mean sum (140.7) on the Adjustment Index for the 24 clinical patients was significantly greater than the mean sum (119.0) for an age, gender, and marital status matched control group, $t(46) = 3.17$, $p < .003$. The control group was chosen from a large sample ($N = 609$) that was used to calculate personality disorder scale reliabilities. Theoretically, a higher sum on the adjustment index indicates the endorsement of broad and various aspects of psychopathology. Note, however, that the large sample may have contained members who did have personality disorders, although studies suggest that the rate may be 6% to 10% of a sample of convenience (Merikangas & Weissman, 1986).

The CATI identified 8 of the 24 clinician-diagnosed patients as having no personality disorders (i.e., no scale values were 1 *SD* above the mean). The MCMI–II identified 3 patients as having no personality disorders. When this same criterion was applied to the 24 matched control subjects using the 13 scales, 12 subjects had 1 or more elevated scales, and 6 subjects had 2 or more elevated scales. These findings suggest at least two overlapping possibilities for the CATI: (a) the criterion of 1 *SD* above the mean is too lenient, resulting in an unacceptably high false-positive rate; or (b) the presence of personality disorders was greater than might be expected for the control sample and for the large sample.

The results of the clinician's checklist were somewhat disappointing and troublesome. First, the average number of symptoms checked by the clinician's was 16.5 per patient. This was not enough information to provide for an official diagnosis in 13 of the 24 cases. Second, and perhaps not surprising, there was far less than 100% concordance in the remaining cases between the clinician's

CATI diagnosis and the clinician's diagnosis according to the checklist. However, in at least 15 of the 24 cases, the CATI diagnosis (at 1 *SD*) agreed with the clinician's checklist diagnosis (including a diagnosis of no personality disorder). Thus, despite the clinicians' judgments that these patients had personality disorders, some may not have had them, at least according to the CATI and the clinicians' checklists.

Convergent Validity of the Axis I Scales

Depression scale. A series of pilot studies with the CATI lead to the selection of 21 items that had good face validity with the concept of depression. A sample of convenience volunteered (97 college students: 46 men, 48 women, and 3 missing values; M age = 20.2) to take the CATI, Beck's Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), and the Minnesota Multiphasic Personality Inventory (MMPI) Depression scale. The Depression scale (sum score) of the CATI correlated .68 with the *T* score of the MMPI Depression scale and correlated .69 with the BDI Depression scale. The BDI and the MMPI Depression scales correlated .60 with each other. All three correlations were significant at $p < .01$ (two-tailed tests).

To test for construct validity, the subjects were divided into two groups (depressed and nondepressed) for three separate analyses on the basis of their BDI score, their MMPI score, and their combined BDI and MMPI scores. Independent *t* tests between the two groups revealed that the mean depression score sum on the CATI for each of the depression groups was significantly greater than the nondepressed group's mean.

The Depression scale of the CATI was also validated on the 24 personality disordered patients used in the concurrent validation of the CATI with the MCMI–II because 75% of the clinical sample had an Axis I disorder involving some type of depressive symptomatology. Their mean sum (53.0) on the depression scale of the CATI was significantly greater than the mean (37.0) of the 24-person control group, $t(46) = 5.291$, $p < .001$.

Anxiety scale. Pilot studies with the CATI also led to the selection of 27 items that had good face validity with the concept of anxiety. Another sample of convenience volunteered (92 college students: 32 men and 60 women; M age = 22.2) to take the CATI, the Spielberger State–Trait Anxiety Inventory (Spielberger, 1983), and the MMPI anxiety scale.

The CATI Anxiety scale sum correlated .63 with the Spielberger State Anxiety scale, .86 with the Spielberger Trait Anxiety scale, and .83 with the MMPI Anxiety scale. All three correlations were significant at $p < .01$ (two-tailed tests).

To test for construct validity, the subjects were divided into two groups (anxious and nonanxious) on the basis of their MMPI anxiety *T* scores,

Spielberger State Anxiety scores, and Spielberger Trait Anxiety scores. Independent *t* tests between the two groups revealed that the mean depression score sums on the CATI for the anxious groups was significantly greater than the nonanxious groups' means in all three analyses.

Brain Dysfunction scale. In a discriminant validity study (Coolidge & Merwin, 1989; Coolidge, Merwin, & Philbrick, 1989), 17 closed head injured (CHI) patients (M age = 36 years) volunteered to take the CATI. They were all enrolled in a cognitive rehabilitation program for less than 1 year. There were 13 White women and 4 White men, and their mean Full Scale IQ score was 91. All but two had high school educations or greater. None of the patients had prior mental hospitalizations. Thirteen of the patients had been in car accidents, and 11 had no known period of unconsciousness. All 17 patients showed significant or borderline impairment on a neuropsychological test battery.

The mean sum (38.9) on the Brain dysfunction scale of the CATI for the CHI group was significantly higher than the mean (25.5) of a control group, $t(32) = 4.75$, $p < .001$. Trial and error revealed that if a mean cutoff score of 30 was used to identify brain dysfunction in these two groups, 15 of the 17 (88%) CHI patients would have been correctly identified, whereas 14 of (82%) of the control group would have fallen below the cutoff.

A factor analysis (Merwin & Coolidge, 1989) with a varimax rotation of the Brain Dysfunction scale on the initial sample of 609 subjects revealed a three-factor structure, which factors for 32%, 7%, and 7% of the total variance, respectively. In the rotated factor matrix, Factor 1 measured concentration and new learning, Factor 2 involved expressive and receptive speech, and Factor 3 measured global memory abilities. A multivariate analysis of variance on the 15 questions of the Brain Dysfunction scale between the CHI and control groups was significant, $F(1, 17) = 3.32$, $p < .01$. The question, "I tend to forget things I am supposed to do," had the highest discriminatory power among the 15 questions. Questions about concentration, memorization, reading, and balance had the next highest discriminatory ability, presented here in rank order. Of the 15 questions, 10 were significant at .019 or less. Of the 5 nonsignificant questions, 3 of them had been designed to detect aphasia, and this is consistent with the finding that specific aphasic syndromes are uncommon following CHIs in adults (Ewing-Cobbs, Levin, Eisenberg, & Fletcher, 1987; Heilman, Saffran, & Geschwind, 1971).

DISCUSSION

These results provide preliminary support for the continued research of the CATI in the assessment of personality disorders. The stability over time and moderately high internal consistency of the personality disorder scales attests to

the CATI's test reliability. The initial validity analyses also lend support to the CATI's concurrent and discriminant validity, although the latter results must be deemed preliminary due to the small sample size ($N = 24$ personality disordered patients).

The reliability and validity studies of the Axis I scales also lend support for the CATI as a measure of anxiety and depression. The Brain Dysfunction scale also received preliminary support as a measure of self-perception of cognitive change in adults with CHIs. Again, the Brain Dysfunction scale should be validated on a larger sample and on additional types of brain disease and injury (e.g., early or initial onset of dementia). Although traditional neuropsychological assessment employs more elaborate cognitively based testing than emotional testing (Nelson et al. 1989), the contributions of CHI and disease to various patterns of emotional responding is well-documented (Gasparrini, Satz, Heilman, & Coolidge, 1978; Lezak, 1983, 1987). The advantages of a standardized inventory such as the CATI in the assessment of the self-perception of emotional and cognitive functioning should be fairly obvious. In addition, with the development of the significant-other form of the CATI, the patient's perception of their own functioning can be compared to a person who may be familiar with the patient. This comparison may be useful to those with brain injuries or disease when the patient denies problems (e.g., McGlynn & Schacter, 1989).

Our study had a number of limitations, not the least of which was the small sample sizes in the personality disorder scale validity study. With respect to the reliability studies, note that the questions of the CATI have been designed to reflect directly the criteria from the *DSM–III–R*. We did not have the luxury of eliminating criteria that hindered overall reliability. In fact, in an earlier version of the CATI, high Cronbach alphas were obtained for most of the 13 scales; however, questions were eliminated without regard for the criteria they represented. Thus, high reliability was obtained at the expense of content validity. The current version of the CATI balances the need for content validity with acceptable scale reliability.

The CATI test–retest reliabilities of the 13 personality disorder scales are high ($M\ r = .90$), although our study employed a small number of purportedly normal subjects ($N = 39$). In addition, the test–retest interval may have been too short (1 week) and should be extended to at least 1 month or longer.

The normal sample ($N = 609$) was also obtained largely through convenience. Some experimental control was provided to ensure that this sample did not contain former mental hospital patients or severely CHI patients, although no interviews were conducted in ensure that the sample did not contain personality disordered people. Thus, the purportedly normal sample probably contained a percentage of people with personality disorders. Therefore, the analyses comparing the 24 personality disordered patients with a 24 person matched control group may have been hampered by the actual presence of personality disorders in the control group. Further studies should be conducted

to determine whether the CATI can be used to differentiate between personality disordered patients and normal controls and between personality disorders and various Axis I disorders. Despite these limitations, the 71-item Adjustment Index of the CATI was significantly higher for the personality disordered patients than the matched control group. At a criterion of 1 *SD* above the mean, the CATI also identified 16 of the 24 personality disordered patients (according to clinicians' judgment) having at least 1 elevated scale, whereas the CATI identified 12 persons with at least 1 elevated scale of the 24 in the matched control group. Despite this possible high false-positive rate in the matched sample, there were significantly more elevated personality disorder scales (at 1 *SD*) for the personality disordered group than in the matched sample, $t(46) = 2.21$, $p < .05$.

Another major difficulty faced in our study was the attempt at establishing discriminant and concurrent validity of the CATI with the clinician's judgment and with the MCMI–II. Hyler et al. (1988) noted the controversy surrounding the validity of solely using a clinician's judgment to determine the presence and identity of a personality disorder, and studies have shown that the interrater reliability of the diagnosis of a personality disorder ranges from poor to modest (Mellsop, Varghese, Joshua, & Hicks, 1982; Spitzer, Forman, & Nee, 1979). Therefore, we chose to compare the CATI to the clinician's diagnosis and the MCMI–II. With respect to the latter test, however, Millon (1985) noted that the *DSM–III–R* criteria are not sacrosanct and not infallible. He proposed that questions reflecting different conceptual orientations may still provide validity. However, in the case of the obsessive compulsive personality disorder, the MCMI–II provides coverage for only half of the stated criteria with its ten 3-point questions on the scale; thus, the argument that the criteria are not sacrosanct becomes specious. Why should the MCMI–II ignore the *DSM–III–R* criteria? Can Millon offer theoretical or empirical justifications for not including specific criteria?

Note that although overlapping of questions on the CATI scales was minimized, the factor analysis of the personality disorder scales may have been contaminated to some extent by question overlap. Because there is some *DSM–III–R* criteria overlap (e.g., Avoidant, Schizoid, and Schizotypal scales), question overlap is nearly unavoidable. However, additional factor analytic studies may address this thorny problem.

The issue of personality disorder scale validity is predicated on many assumptions. We assumed that the personality disorders themselves are valid entities, that the *DSM–III–R* criteria are reliable and valid measures of these entities, and that people can reliably and validly assess their own behavior. Finally, a test instrument that measures people's assessment should be reliable and valid. To a greater and lesser degree, all test constructors share these assumptions. Given the tenuousness of the relationships among these assumptions, the need for adherence to the *DSM–III–R* criteria cannot be overemphasized. In the concur-

rent validity study between the CATI and the MCMI–II, some of the low validity correlations were obtained due to a lack of adherence of the MCMI–II to *DSM–III–R* criteria. When adherence was high, the correlations were stronger between the CATI and the MCMI–II.

In summary, our findings suggest that the CATI has potential in the assessment of personality disorders. Further research is required for the CATI's test–retest reliability and discriminant validity. However, it appears that CATI shows some promise as a measure of personality disorders, anxiety, depression, and the self-perception of cognitive brain dysfunction. In addition, an earlier study (Coolidge et al., 1985) used a significant-other form of the CATI to evaluate personality change in dementia patients, and studies are currently underway to establish the reliability and validity of this form.

ACKNOWLEDGMENTS

Reprints and/or copies of the CATI and a computer (IBM) scoring program are available for research purposes. Address requests to Frederick L. Coolidge, PhD, Psychology Department, P.O. Box 7150, University of Colorado, Colorado Springs, CO 80933–7150.

REFERENCES

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

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Binder, L. M. (1986). Persisting symptoms after mild head injury: A review of postconcussive syndrome. *Journal of Clinical and Experimental Neuropsychology, 8,* 323–346.

Coolidge, F. L. (1984). *Coolidge Axis II Inventory.* U.S. Copyright TXU 182–026, Washington, DC.

Coolidge, F. L., Bracken, D. D., Taylor, L. R., Smith, K., & Peters, B. (1985). Personality disorders in Alzheimer's disease: A hierarchical cluster analysis approach. *Journal of Clinical and Experimental Neuropsychology, 7,* 616.

Coolidge, F. L., & Merwin, M. M. (1989, April). *Self-assessment of the cognitive and emotional sequelae of brain damage.* Paper presented at the meeting of the Western Psychological Association, Reno.

Coolidge, F. L., Merwin, M. M., & Philbrick, P. B. (1989, November). *Validation of a self-report, brain dysfunction scale of the Coolidge Axis II Inventory.* Paper presented at the meeting of the National Academy of Neuropsychologists, Washington, DC.

Costanzo, R. M., & Becker, D. P. (1986). Smell and taste disorders in head injury and neurosurgery patients. In H. L. Meiselman & R. S. Rivlin (Eds.), *Clinical measurement of taste and smell* (pp. 565–578). New York: Macmillan.

Cronbach, L. J. (1951). Coefficient alpha and the internal structures of tests. *Psychometrika, 16,* 297–334.

Ewing-Cobbs, L., Levin, H. S., Eisenberg, H. M., & Fletcher, J. M. (1987). Language functions following closed-head injury in children and adolescents. *Journal of Clinical and Experimental Neuropsychology, 9,* 575–592.

Gasparrini, W. G., Satz, P., Heilman, K., & Coolidge, F. L. (1978). Hemispheric asymmetries of affective processing as determined by the Minnesota Multiphasic Personality Inventory. *Journal of Neurology, Neurosurgery, and Psychiatry, 41,* 470–473.

Grana, A. S., Coolidge, F. L., & Merwin, M. M. (1989). Personality profiles of the morbidly obese. *Journal of Clinical Psychology, 45,* 762–765.

Heilman, K. M., Saffran, E., & Geschwind, N. (1971). Closed head trauma and aphasia. *Journal of Neurology, Neurosurgery, and Psychiatry, 34,* 265–269.

Hosman, P. A. (1989). *Convergent validity of the Spielberger Anxiety Scale, the MMPI Anxiety scale, and the Coolidge Axis II Inventory.* Unpublished honor's thesis, University of Colorado, Colorado Springs.

Hyler, S. E., Rieder, R. O., Williams, J. B. W., Spitzer, R. L., Hendler, J., & Lyons, M. (1988). The Personality Diagnostic Questionnaire: Development and preliminary results. *Journal of Personality Disorders, 2,* 229–237.

Lezak, M. D. (1983). *Neuropsychological assessment.* New York: Oxford University Press.

Lezak, M. D. (1987). Relationships between personality disorders, social disturbances, and physical disability following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 2,* 57–69.

Lucero, D. L. (1989). *The construction and validation of a depression scale for the CATI.* Unpublished honor's thesis, University of Colorado, Colorado Springs.

Merwin, M. M., & Coolidge, F. L. (1987, March). *Coolidge Axis Two Inventory Scale reliabilities.* Paper presented to the meeting of the Colorado Psychological Association, Denver.

Merwin, M. M., & Coolidge, F. L. (1989, November). *Factor analysis of a self-report inventory for the assessment of brain dysfunction.* Paper presented at the meeting of the Colorado Psychological Association, Denver.

McGlynn, S. M., & Schacter, D. L. (1989). Unawareness of deficits in neuropsychological syndromes. *Journal of Clinical and Experimental Neuropsychology, 11,* 143–205.

Mellsop, G., Varghese, F., Joshua, S., & Hicks, A. (1982). The reliability of Axis II of DSM-III. *American Journal of Psychiatry, 139,* 1360–1361.

Merikangas, K. R., & Weissman, M. M. (1986). Epidemiology of DSM-III Axis II personality disorders. In A. J. Frances & R. E. Hales (Eds.), *American Psychiatry Association: Annual review. Vol. 5: Psychiatry update series* (pp. 258–278). Washington, DC: American Psychiatric Press.

Millon, T. (1977). *Millon Clinical Multiaxial Inventory.* Minneapolis: National Computer Systems.

Millon, T. (1985). The MCMI provides a good assessment of DSM-III disorders: The MCMI-II will prove even better. *Journal of Personality Assessment, 49,* 379–391.

Millon, T. (1987). *Millon Clinical Multiaxial Inventory-II manual.* Minneapolis: National Computer Systems.

Mitton, N. M. (1991). *A scale of dissimulation on the CATI: Reliability and validity.* Unpublished master's thesis, University of Colorado, Colorado Springs.

Morey, L. C., Waugh, M. H., & Blashfield, R. K. (1985). MMPI scales for DSM-III personality disorders: Their derivation and correlates. *Journal of Personality Assessment, 49,* 245–251.

Nelson, L. D., Satz, P., Mitsushina, M., Van Gorp, W., Cicchetti, D., Lewis, R., & Van Lancker, D. (1989). Development and validation of the neuropsychology behavior and affect profile. *Journal of Consulting and Clinical Psychology, 1,* 266–272.

Piersma, H. L. (1986). The factor structure of the Millon Clinical Multiaxial Inventory (MCMI) for psychiatric inpatients. *Journal of Personality Assessment, 50,* 587–594.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spitzer, R. L., Forman, J. B. W., & Nee, J. (1979). DSM-III field trials: I. Initial interrater diagnostic reliability. *American Journal of Psychiatry, 136,* 815–817.

Streiner, D. L., & Miller, H. R. (1988). Validity of MMPI scales for DSM-III personality disorders: What are they measuring? *Journal of Personality Disorders, 2,* 238–242.

Widiger, T. A., Williams, J. B. W., Spitzer, R. L., & Francis, A. (1985). The MCMI as a measure of DSM-III. *Journal of Personality Assessment, 49,* 366–378.

Frederick L. Coolidge
Department of Psychology
P.O. Box 7150
University of Colorado
Colorado Springs, CO 80933–7150

Received February 22, 1991
Revised May 5, 1991