A Review of the Measures Designed to Assess DSM-5 Personality Disorders

Adrian Furnham¹, Rebecca Milner¹, Reece Akhtar¹, Filip De Fruyt²

¹Research Department of Clinical, Educational and Health Psychology, University College London, London, UK
²Department of Developmental, Personality and Social Psychology, Ghent University, Ghent, Belgium

Email: a.furnham@ucl.ac.uk

Received 3 July 2014; revised 1 August 2014; accepted 24 August 2014

Abstract

The definition, classification and assessment of personality disorders (PDs) have attracted considerable debate for nearly 50 years. This paper attempts a comprehensive review of the instruments to assess all, or specific, individual disorders as described in DSM-5, including structured interviews and inventories. The review should be helpful for clinicians, researchers and also industrial and organizational psychologists, to screen and assess the personality pathology spectrum from subclinical manifestations to full blown personality pathology. A decision tree helpful to choose among the different measures is also provided.

Keywords

Personality Disorders, Personality Pathology, Measures, Questionnaires, Structured Interviews

1. Introduction

There has been much debate about personality disorders (PDs) over the years, particularly their definition, conceptualization, occurrence and assessment. Perhaps the greatest “shake up” in the way PDs were discussed has occurred in the move from DSM-IV (American Psychiatric Association, 2000) to DSM-5, proposing dimensional alternatives for the DSM-IV categorical diagnoses (Widiger, Livesley, & Clark, 2009). Notwithstanding this lively debate, DSM-5 preserved in its Section 2 the categorical PDs like distinguished in DSM-IV, whilst an alternative trait system is referred to Section 3 for further evaluation and research (American Psychiatric Association, 2013). Although, there is a great deal of activity developing and validating new instruments like the Personality Inventory for DSM-5 (PID-5; Krueger, Derringer, Markon, Watson, & Skodol, 2012) to assess and evaluate (Bagby, 2013) this new trait model, the diagnosis and assessment of categorical PDs is primarily advocated in the official nomenclature of the American Psychiatric Association.

How to cite this paper: Furnham, A., Milner, R., Akhtar, R., & De Fruyt, F. (2014). A Review of the Measures Designed to Assess DSM-5 Personality Disorders. Psychology, 5, 1646-1686. http://dx.doi.org/10.4236/psych.2014.514175
At the same time, the attention to PDs, both from an academic and societal perspective, expanded dramatically due to the impairing character of the diagnoses and the increasingly high financial costs involved in the treatment of patients with personality pathology (Gustavsson et al., 2012). Apart from attention from clinical psychologists to full-blown personality pathology, selection and human resources psychologists have become interested in subclinical manifestations of aberrant personality and the impact on individuals’ workplace functioning. This is because a substantial proportion of the general population and workforce has personality problems themselves or has to deal with (subclinically) disordered persons as colleagues or supervisors (Wille, De Fruyt, & De Clercq, 2013; De Fruyt, Wille, & Furnham, 2013b). Whereas clinical psychologists have been treating patients with one or more PDs and co-occurring pathology, industrial and organisational psychologists run career development programs to coach people on how to deal with the dark sides of their personality, a common need for all these professional groups is well-designed and psychometric sound assessment instruments. In addition, they also need criteria to choose among the different instruments currently available.

The present paper provides a broad review of current PD measures together with a decision tree to choose among them. Length constraints meant we could not consider proposed personality disorders like Depressive Personality Disorder. The aim is to be comprehensive and descriptive rather than (psychometrically or conceptually) critical which would involve a different paper. We have attempted to catalogue all measures, which has not been done before. The measures are in no way psychometrically equivalent though each paper has been peer reviewed.

Over the years a large number of measures have been devised for research and practice. The aim of this review is to alert psychologists and researchers to the range of instruments available to assess the categorically conceived PDs listed in DSM-5 and provide a set of criteria by which professionals may choose one over another. In the introduction of this paper we refer to DSM-5 PDs, though it should be clear that almost all measures were developed before the release of DSM-5, so we refer to these previous DSM-editions when describing these measures. The available PD measures differ on at least four major characteristics.

First, some instruments attempt to be comprehensive and measure all of the PDs currently (or previously) thought to exist, because the nature and number of PDs have shifted across the different DSM editions. Some “admit” disorders that others discount, but the usual number is around 10 - 15 disorders. On the other hand, some instruments set out simply to measure one very specific disorder. Second, there seem to be four most common methods to assess the PDs: structured diagnostic interviews, rating instruments for clinicians, self-report questionnaires and other-report questionnaires (Friedman, Olmanns, & Turkheimer, 2007). Thus, two use observer data (clinician, family) and two use self-report data approaches towards measurement. By far the most common however are questionnaires and structured interviews. Third, some measures are about subtypes of the PD in the sense that they are multidimensional measures that yield scores on different, but related facets of the disorder. For example, some measures and theorists may distinguish between grandiose and vulnerable, or communal and agentic Narcissistic PD (NPD; Gebauer, Sedikides, Verplanken, & Maio, 2012). Most measures, however, mimic DSM-5 categorical criteria and are not about the distinction among subtypes of a specific PD.

Fourth and finally, PD measures have been developed for essentially five target groups. The first group of users are clinicians attempting a reliable and valid diagnosis of a PD. The second is a related group, namely academic researchers who may be testing theories of the aetiology or prognosis of a PD eventually after treatment. Industrial and organisational psychologists form a third professional group interested in evaluating aberrant personality and subclinical forms of personality pathology in the context of personnel selection or career coaching and development. Finally, there are two other groups, namely “lay people” who may be interested in self-diagnosis, but also relatives of those with a specific PD requiring information about personality disorder symptoms and its prognosis.

There are, inevitably, a number of instruments on the web with unknown psychometric properties as well as various “popular books” that attempt to explain and describe the PDs for the lay public. The present review, however, primarily attempts a comprehensive overview for the first three groups interested in the professional assessment of personality pathology.

Before listing and discussing the different measures, we provide an overview of the DSM-5 PDs with their clinical labels and a short description in Table 1. This table further shows the labels and descriptions of PDs as they are used in a popular measure frequently used in occupational and career coaching and development settings (Hogan & Hogan, 1997; Furnham, Trickey, & Hyde, 2012). The remaining columns illustrate the labels used in books written by psychiatrists (Oldham & Morris, 1991), clinical psychologists (Miller, 2008) and I/O
Table 1. Different labels for traits associated with similar disorders.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Paranoid</td>
<td>Distrustful and suspicious of others; motives are interpreted as malevolent.</td>
<td>Sceptical; Cynical, distrustful and doubting others’ true intentions.</td>
<td>Vigilant; Vigilantes; Habitual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizoid</td>
<td>Emotional coldness and detachment from social relationships; indifferent to praise and criticism.</td>
<td>Reserved; Aloof, detached and uncommunicative; lacking interest in or awareness of the feelings of others.</td>
<td>Solitary; Oddballs; Aloof</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizotypal</td>
<td>Odd beliefs or magical thinking; behaviour or speech that is odd, eccentric or peculiar.</td>
<td>Imaginative; Acting and thinking in creative and sometimes odd or unusual ways.</td>
<td>Idiosyncratic; Creativity and vision; Eccentric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antisocial</td>
<td>Disregard for the truth; impulsivity and failure to plan ahead; failure to conform.</td>
<td>Mischiefous; Enjoying risk taking and testing the limits; needing excitement; manipulative, deceitful, cunning and exploitative.</td>
<td>Adventurous; Predators; Mischiefous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borderline</td>
<td>Inappropriate anger; unstable and intense relationships alternating between idealisation and devaluation.</td>
<td>Excitable; Moody and hard to please; intense but short-lived enthusiasm for people, projects or things.</td>
<td>Mercurial; Reactors; Volatility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histrionic</td>
<td>Excessive emotionality and attention seeking; self dramatising, theatrical and exaggerated emotional expression.</td>
<td>Colourful; Expressive, animated and dramatic; wanting to be noticed and needing to be the centre of attention.</td>
<td>Dramatic; Emoters; Melodramtic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Narcissistic</td>
<td>Arrogant and haughty behaviours or attitudes, grandiose sense of self-importance and entitlement.</td>
<td>Bold; Unusually self-confident; feelings of grandiosity and entitlement; over valuation of one’s capabilities.</td>
<td>Self-Confidence; Preeners; Arrogance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoidant</td>
<td>Social inhibition; feelings of inadequacy and hypersensitivity to criticism or rejection.</td>
<td>Cautious; Reluctant to take risks for fear of being rejected or negatively evaluation.</td>
<td>Sensitive; Shrinkers; Excessive Caution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent</td>
<td>Difficulty making everyday decisions without excessive advice and reassurance; difficulty expressing disagreement out of fear of loss of support or approval.</td>
<td>Dutiful; Eager to please and reliant on others for support and guidance; reluctant to take independent action or to go against popular opinion.</td>
<td>Devoted; Clingers; Eager to Please</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obsessive-Compulsive</td>
<td>Preoccupations with orderliness; rules, perfectionism and control; over-Conscientiousness and inflexible.</td>
<td>Diligent; Meticulous, precise and perfectionistic, inflexible about rules and procedures; critical of others;</td>
<td>Conscientious; Detailers; Perfectionistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive-Aggressive</td>
<td>Passive resistance to adequate social and occupational performance; irritated when asked to do something he/she does not want to.</td>
<td>Leisurly; Independent; ignoring people’s requests and becoming irritated or argumentative if they persist.</td>
<td>Leisurly; Spoilers; Passive Resistance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

psychologists (Dotlich & Cairo, 2003) to explain the PDs to lay people.

2. Available Measures

This paper covers the measures available, including those assessing all PDs, as well as each PD in turn. We also acknowledge the fact that there are instruments intended to measure the prevalence of specific symptoms of PDs, yet have excluded these from our analysis due to space constraints. Likewise, we have also excluded alternative dimensional conceptualisations of PDs and personality pathology (Clark, 2007), except when these methods are specifically targeted to assess the categorical DSM-5 PDs. We hence do not explicitly discuss and reiterate the discussion on alternative dimensional models of PDs (Widiger & Clark, 2000; Widiger & Costa, 2013), except
when these provide direct assessments of the categorical PDs. To our knowledge a review such as this has not been done before, though there are review papers that have reviewed some instruments at the same time (Clark & Harrison, 2001; McDermut & Zimmerman, 2008; Segal & Coolidge, 2007; Widiger & Boyd, 2009; Zimmerman, 2003).

Apart from usual bibliometric investigations, we emailed over 50 experts (mainly those on the editorial board of specialist PD journals) in the area showing them our list and asking if they knew of any measures that we were not aware of. This did yield half a dozen extra, and we are reasonably satisfied that we have been able to locate most important measures.

3. Measures of all the Personality Disorders (See Table 2)

3.1. Structured Interviews

The Structured Interview for DSM-III Personality Disorders (SIDP; Pfohl, Stangl, & Zimmerman, 1983) has largely fallen out of favour because of its focus on DSM-III PDs. Despite this, it has been shown to hold highly variable test-retest reliabilities ranging from .24 for obsessive-compulsive PD to .74 for histrionic PD, with an average level of .54 (First et al., 1995). Pfohl, Blum and Zimmerman (1997) adapted the SIDP at the advent of the DSM-IV, releasing The Structured Interview for DSM-IV Personality Disorders (SIDP-IV)—a fairly brief interview (lasting roughly 60 minutes) that features both a patient and an informant. This is beneficial as it helps gain a different perspective on the patient in question. There are two versions of the SIDP-IV: a diagnostic version and a “topical” version, though the only difference is the order of the questions. The benefit of including a topical version is that it includes natural questions that are designed to make interviewing defensive patients easier. Much like the International Personality Disorder Examination (IPDE; Loranger, 1999; see below), the SIDP-IV can also assess for Personality Disorder Not Specified (PDNOS) however the SIDP-IV will diagnose a PDNOS only when two or more disorders are one criterion short of the diagnostic threshold. Jane, Pagan, Turkheimer, Fiedler and Oltmanns (2006) found inter-rater reliability for each PD being greater than .70, a finding also supported by Damen, De Jong and Van Der Kroft (2004).

The Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV; Zanarini, Frankenburg, Sickel, & Yong, 1996). This semi-structured clinical interview assesses all DSM-IV PDs, and like most clinical interviews, specialised training is required before the interview can be administered. The interview has 108 items, with each disorder rated on a scale of 0 (disorder is absent) to 2 (disorder is present). If the totalled scores exceed a threshold the clinician can diagnose a disorder. The original paper cites internal consistency levels ranging from .64 to .93, with six of the disorders having levels greater than .70; acceptable levels of test-retest reliability with Kappa = .58 to 1 are reported over a 6-month period. These are also called dependability coefficients (Chmielewski & Watson, 2009). The DIPD-IV was used in the Collaborative Longitudinal Personality Disorders Study (CLPS).

The Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II; First, Gibbons, Spitzer, Williams, & Benjamin, 1997) is widely used and researched, unlike the DIPD-IV. The respondent typically first completes a questionnaire and interviewers then follow up responses. It is also the shortest interview (140 items), lasting minimally 30 minutes (the DIPD-IV lasts around 90 minutes). The SCID-II measures all DSM-IV PDs and the associated symptoms in the order they are presented in the DSM-IV. Some have criticised its brevity (Rogers, 2003). Investigations into the instrument’s reliability and validity have shown considerable support. Lobbestael, Leurgans and Amtz (2011) found mean kappa scores of .84. Moran et al. (2003) provided further support with mean kappa scores of .71, but others have reported lower Kappas (Hyler, Skoldol et al., 1990, 1992). Skodol et al. (1991) investigated the convergent validity of the SCID-II by comparing it to diagnoses made by the International Personality Disorder Examination (IPDE; Loranger, 1999). The authors found that the two instruments’ diagnoses for each PD correlated from .58 to .87, suggesting that both instruments measure the same PDs to a “reasonable” extent.

The Personality Disorder Interview (PDI-IV; Widiger, Mangine, Crobitt, Ellis, & Thomas, 1995) is another semi-structured interview that assesses each of the 94 personality disorder criteria displayed in the DSM-IV, making it a lengthy interview lasting around 90 - 120 minutes. Rogers (2001) supports the instrument’s extensive criteria, however criticises its sometimes sophisticated and complex language. This is a particularly valid concern when using the instrument with adolescents and cognitively impaired patients. Rogers (2001) also notes how, despite high levels of reliability, its little adoption within clinical environments has proven to be an
### Table 2. A review of measures that attempt to measure all the disorders.

<table>
<thead>
<tr>
<th>Name of Instrument</th>
<th>Author</th>
<th>Number of Items</th>
<th>Dimensions</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured Interview for DSM-III Personality Disorders (SIDP)</td>
<td>Pfohl, Stangl, &amp; Zimmerman, 1983</td>
<td></td>
<td>All DSM-III Personality Disorders</td>
<td>Test-retest reliabilities range</td>
<td>.24 - .74, with an average of .54.</td>
</tr>
<tr>
<td>The Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV).</td>
<td>Zanarini, M. C., Frankenburg, F. R., Sickel, A. E., &amp; Yong, L. (1996).</td>
<td>108</td>
<td>All DSM-IV Personality Disorders</td>
<td>Internal Consistency ranges</td>
<td>between .64 to .93</td>
</tr>
<tr>
<td>The Structured Interview for DSM-IV Personality Disorders (SIDP-IV)</td>
<td>Pfohl, Blum, &amp; Zimmerman (1997)</td>
<td></td>
<td>All DSM-IV Personality Disorders</td>
<td>Test-Retest Reliability ranges</td>
<td>between .58 to 1.00</td>
</tr>
<tr>
<td>The Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II)</td>
<td>First, Gibbons, Spitzer, Williams, &amp; Benjamin, 1997</td>
<td>140</td>
<td>All DSM-IV Personality Disorders</td>
<td>Test-retest reliabilities have been found to range</td>
<td>between .71 - .84</td>
</tr>
<tr>
<td>Personality Disorder Interview (PDI-IV)</td>
<td>Widiger, Mangine, Corbit, Ellis, &amp; Thomas, 1995</td>
<td></td>
<td>Semi-Structured Interview Questions for assessment of each of the 94 individual personality disorder diagnostic criteria are presented.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>International Personality Disorder Examination (IPDE)</td>
<td>Loranger, 1998; Loranger, 1999; Loranger et al., 1987</td>
<td></td>
<td>The IPDE is a structured clinical interview that systematically surveys the phenomenology and life experiences relevant to the diagnosis of all DSM-IV Axis II (and ICD-10) personality disorders</td>
<td>Inter-rater reliabilities range between .81 - .92</td>
<td>(Lenzenweger, 1999).</td>
</tr>
<tr>
<td>The International Personality Disorder Examination Questionnaire (IPDEQ)</td>
<td>The World Health Organisation</td>
<td>99 Items</td>
<td>1) Work, 2) Self, 3) Interpersonal relationships 4) Affects, 5) Reality testing, 6) Impulse control</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The SCID-II/PQ was found to have moderate convergent validity with the NPD section of the SCID-II (Campbell & Miller, 2011). The SCID-II and the IPDE was shown to hold good convergent validity (.58 - .87; Skodol et al., 1991).

The diagnosis of any PD was highly reliable with phi > .92. However, diagnosis of non-specific PD was not reliable at all (phi close to 0) suggesting that this is a true residual category. Diagnoses of specific PDs were highly reliable with the exception of schizoid PD. Diagnosis of antisocial and Borderline PDs were perfectly reliable with phi equal to 1.00 (Fountoulakis et al., 2002).
<table>
<thead>
<tr>
<th>Test</th>
<th>Author/Reference</th>
<th>DSM-III-R Personality Disorders</th>
<th>Sensitivity/Specificity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iowa Personality Disorder Screen (IPDS)</td>
<td>Langbehn, Pföhl, &amp; Reynolds et al., 1999</td>
<td>11</td>
<td>All</td>
<td>The average internal consistency was .72 (Langbehn et al., 1999).</td>
</tr>
<tr>
<td>The personality assessment inventory (PAI)</td>
<td>Morey (1991)</td>
<td></td>
<td>344</td>
<td>Blind administration of the IPDS yielded excellent sensitivity (92%) and good specificity (79%), using a subset of five screening items (Langbehn et al., 1999).</td>
</tr>
<tr>
<td>Standardised Assessment of Personality: Abbreviated Scale (SAPAS)</td>
<td>Moran, Leese, Lee, Walters, Graham, Thornicroft, &amp; Mann (2003)</td>
<td>8</td>
<td>All</td>
<td>The personality assessment inventory (PAI)</td>
</tr>
<tr>
<td>The Hogan Development Survey</td>
<td>Hogan &amp; Hogan, 1997</td>
<td></td>
<td>11</td>
<td>The Hogan Development Survey</td>
</tr>
<tr>
<td>The Omnibus Personality Inventory (OMNI)</td>
<td>Loranger, 1994, 2002</td>
<td>375 items</td>
<td>All</td>
<td>The Omnibus Personality Inventory (OMNI)</td>
</tr>
</tbody>
</table>

**Clinical:**
1) Somatic Complaints
2) Anxiety
3) Anxiety-Related Disorders
4) Depression
5) Mania
6) Paranoia
7) Schizophrenia
8) Borderline Features
9) Antisocial Features
10) Alcohol Problems
11) Drug Problems

**Interpersonal Scales**
1) Dominance
2) Warmth

**Treatment Scales**
1) Aggression
2) Suicidal Ideation
3) Stress
4) Non-support
5) Treatment Rejection

**Validity Scales**
1) Infrequency
2) Negative Impression
3) Positive Impression
4) Inconsistency

The scales were found to significantly correlate with the prevalence of life-events in psychiatric patients, except for the mania and anxiety scales.

When compared to the SCID-II, the SAPAS was found to have a good balance of sensitivity (.73) and specificity (.9; Pluck, Sirdifield, Brooker & Moran, 2012).

The HDS can predict work success, as well as each factor being reduced to clusters A, B and C of the DSM-IV (Furnham et al., 2012).
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Authors</th>
<th>Items</th>
<th>Sub-Scales</th>
<th>Diagnosis</th>
<th>Test-retest Correlation Coefficients</th>
<th>Predictive Validity</th>
<th>Support</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schedule for Nonadaptive and Adaptive Personality (SNAP)</td>
<td>Clark, 1993</td>
<td>375</td>
<td>All DSM-IV Personality Disorders</td>
<td>Test-retest correlation coefficients over 7 to 14 months ranged from .59 to .84. (Melley, et al., 2002)</td>
<td>Internal consistency ranges between .43 - .70 (Hyler &amp; Lyons, 1988). Test-retest reliabilities are greater than .56 (Hurt et al., 1984)</td>
<td>Our results support the predictive validity of the diagnostic scales on the SNAP (Clark, 1993). SNAP scores can predict cluster A and C diagnoses.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Personality Diagnostic Questionnaire (PDQ)</td>
<td>Hyler et al., 1988</td>
<td>163</td>
<td>All DSM-III Personality Disorders</td>
<td>Test-retest reliabilities are greater than .76 (Uehara et al., 1997)</td>
<td>Compared to the SCID-II, Kappa levels were no lower than .50. The PDQ-R demonstrated high sensitivity and low specificity. (Abdin et al., 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Personality Diagnostic Questionnaire Revised (PDQ-R)</td>
<td>Hunt &amp; Andrews, 1992</td>
<td>152</td>
<td>All DSM-III Personality Disorders</td>
<td>Test-retest reliabilities averaging .67.</td>
<td>Excellent test-retest reliability (.90) has been established, as well as moderate internal consistency (.76). (Coolidge &amp; Merwin, 1992). Watson &amp; Sinha (1996)—a gender difference was found on the antisocial scale. Age differences were found for several PD scales in that younger respondents (17 - 24 years) scored higher than the older ones (25 - 57 years). The test has demonstrated reliability and validity and has been used in a variety of contexts, such as attributes of psychology majors in college (McCray, King, &amp; Bailly, 2005). A 50% concordance rate with clinicians’ diagnoses for 24 personality-disordered out-patients was found (Coolidge &amp; Merwin, 1992). The test has demonstrated reliability and validity and has been used in a variety of contexts, such as attributes of psychology majors in college (McCray, King, &amp; Bailly, 2005).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Personality Diagnostic Questionnaire 4th Edition (PDQ-4)</td>
<td>Hyler, 1994</td>
<td>99</td>
<td>13 Personality Disorder Scales (DSM-II) 3 Axis I Disorders</td>
<td>Test-retest reliabilities averaging .67.</td>
<td>PCA and CFA reveal that the internal structure of the test has dimensions similar to the full CATI. Therefore, the SCATI is an instrument of value for further research (Watson &amp; Sinha, 2007).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coolidge axis-II inventory (CATI)</td>
<td>Coolidge, 1984</td>
<td>225</td>
<td>All DSM-IV Personality Disorders</td>
<td>Despite the scales being reduced from an average of approximately 25 items to 5 items in length, the SCATI has good internal reliability (Watson &amp; Sinha, 2007). Multivariate analysis revealed that the internal structure of the SCATI is similar to the original. The SCATI is found to have good internal reliability.</td>
<td>The test has demonstrated reliability and validity and has been used in a variety of contexts, such as attributes of psychology majors in college (McCray, King, &amp; Bailly, 2005). A 50% concordance rate with clinicians’ diagnoses for 24 personality-disordered out-patients was found (Coolidge &amp; Merwin, 1992).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coolidge axis-two inventory (SCATI)</td>
<td>Coolidge, 2001</td>
<td>70</td>
<td>All DSM-IV Personality Disorders</td>
<td>PCA and CFA reveal that the internal structure of the test has dimensions similar to the full CATI. Therefore, the SCATI is an instrument of value for further research (Watson &amp; Sinha, 2007).</td>
<td>The test has demonstrated reliability and validity and has been used in a variety of contexts, such as attributes of psychology majors in college (McCray, King, &amp; Bailly, 2005). A 50% concordance rate with clinicians’ diagnoses for 24 personality-disordered out-patients was found (Coolidge &amp; Merwin, 1992).</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Schedule for Nonadaptive and Adaptive Personality (SNAP) | Clark, 1993 | 375 | All DSM-IV Personality Disorders | Test-retest correlation coefficients over 7 to 14 months ranged from .59 to .84. (Melley, et al., 2002) | Internal consistency ranges between .43 - .70 (Hyler & Lyons, 1988). Test-retest reliabilities are greater than .56 (Hurt et al., 1984) | Our results support the predictive validity of the diagnostic scales on the SNAP (Clark, 1993). SNAP scores can predict cluster A and C diagnoses. |
| The Personality Diagnostic Questionnaire (PDQ) | Hyler et al., 1988 | 163 | All DSM-III Personality Disorders | Test-retest reliabilities are greater than .76 (Uehara et al., 1997) | Compared to the SCID-II, Kappa levels were no lower than .50. The PDQ-R demonstrated high sensitivity and low specificity. (Abdin et al., 2011) |
| The Personality Diagnostic Questionnaire Revised (PDQ-R) | Hunt & Andrews, 1992 | 152 | All DSM-III Personality Disorders | Test-retest reliabilities averaging .67. | Excellent test-retest reliability (.90) has been established, as well as moderate internal consistency (.76). (Coolidge & Merwin, 1992). Watson & Sinha (1996)—a gender difference was found on the antisocial scale. Age differences were found for several PD scales in that younger respondents (17 - 24 years) scored higher than the older ones (25 - 57 years). The test has demonstrated reliability and validity and has been used in a variety of contexts, such as attributes of psychology majors in college (McCray, King, & Bailly, 2005). A 50% concordance rate with clinicians’ diagnoses for 24 personality-disordered out-patients was found (Coolidge & Merwin, 1992). The test has demonstrated reliability and validity and has been used in a variety of contexts, such as attributes of psychology majors in college (McCray, King, & Bailly, 2005). |
| The Personality Diagnostic Questionnaire 4th Edition (PDQ-4) | Hyler, 1994 | 99 | 13 Personality Disorder Scales (DSM-II) 3 Axis I Disorders | Test-retest reliabilities averaging .67. | PCA and CFA reveal that the internal structure of the test has dimensions similar to the full CATI. Therefore, the SCATI is an instrument of value for further research (Watson & Sinha, 2007). |
| Coolidge axis-II inventory (CATI) | Coolidge, 1984 | 225 | All DSM-IV Personality Disorders | Despite the scales being reduced from an average of approximately 25 items to 5 items in length, the SCATI has good internal reliability (Watson & Sinha, 2007). Multivariate analysis revealed that the internal structure of the SCATI is similar to the original. The SCATI is found to have good internal reliability. | The test has demonstrated reliability and validity and has been used in a variety of contexts, such as attributes of psychology majors in college (McCray, King, & Bailly, 2005). A 50% concordance rate with clinicians’ diagnoses for 24 personality-disordered out-patients was found (Coolidge & Merwin, 1992). | |
| Coolidge axis-two inventory (SCATI) | Coolidge, 2001 | 70 | All DSM-IV Personality Disorders | PCA and CFA reveal that the internal structure of the test has dimensions similar to the full CATI. Therefore, the SCATI is an instrument of value for further research (Watson & Sinha, 2007). | The test has demonstrated reliability and validity and has been used in a variety of contexts, such as attributes of psychology majors in college (McCray, King, & Bailly, 2005). A 50% concordance rate with clinicians’ diagnoses for 24 personality-disordered out-patients was found (Coolidge & Merwin, 1992). | |
Continued

<table>
<thead>
<tr>
<th>Millon Clinical Multiaxial Inventory-III (MCMI-III)</th>
<th>Millon, Millon, Davis &amp; Grossman (2006)</th>
<th>175</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 Personality Disorder Scales; 11 Moderate Personality Disorder Scales; 3 Severe personality Pathology Scales; 10 Clinical Syndrome Scales; 7 Moderate Syndrome Scales; 3 Severe Syndrome Scales Corrections Scales; 3 Modifying Indices; 2 Random Response Indicators; 42 Grossman Personality Facet Scales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examining the temporal stability of the personality disorder subscales from the MCMI: Psychiatric inpatients completed the MCMI at initial admission and at the next admission (within 2 years). The scales demonstrated adequate stability. High retest correlations for the MCMI clinical syndrome subscales—Overholser (1990).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Personality Beliefs Questionnaire (PBQ)</th>
<th>Beck &amp; Beck, 1991.</th>
<th>126</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Personality disorders:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Avoidant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Dependent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Obsessive-Compulsive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Histrionic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Passive-Aggressive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) Narcissistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7) Paranoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8) Schizoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9) Antisocial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10) Borderline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBQ administered to students, showing good internal consistency across scales; Cronbach’s alphas ranged from .77 to .93. Test-retest correlation coefficients over a month interval were high, ranging from .63 (passive-aggressive) to .82 (paranoid; Trull, et al., 1993).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The PBQ-Short Form (PBQ-SF)</th>
<th>Butler, Beck, &amp; Cohen, 2007</th>
<th>65</th>
</tr>
</thead>
<tbody>
<tr>
<td>All DSM-IV Personality Disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal consistency coefficients range between .81 - .92. Test-retest correlations range from .57 - .82 (Butler et al., 2007).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Minnesota multiphasic personality inventory (MMPI) for DSM-III</th>
<th>Morey, Waugh, &amp; Blashfield (1985)</th>
<th>165</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Histrionic</td>
<td>The complete version of these scales yielded internal consistency estimates superior to those obtained in examining the original clinical scales from the MMPI (from previous versions of the DSM). Hence, it appears that the derived scales are reliable as defined by internal consistency—Morey, Waugh, &amp; Blashfield, 1985.</td>
<td></td>
</tr>
<tr>
<td>2) Narcissistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Borderline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Antisocial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Depressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) Obsessive-Compulsive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7) Passive-Aggressive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8) Paranoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9) Schizotypy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10) Avoidant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11) Schizoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This study found that 5 of the 11 MMPI-PD scales correlated positively and significantly with the corresponding MCMI-PD scales, which gives limited support for the concurrent validity of the MMPI-PD scales when the MCMI-PD scales were used as the criterion measures. The Schizoid, Avoidant, Dependent, Histrionic, and Narcissistic scales achieved significant correlation. Overall, the study supports the limited validity and diagnostic utility of the MMPI-PD scales—Schuler, Snibbe, &amp; Buckwalter, 1994.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
obstacle when evaluating its validity. Widiger, Costa and Samuel (2006) argue that the PDI-IV’s strength lies within its manual and compared the PDI-IV’s manual to manuals of other semi-structured interviews. Most are lacking normative data, statistical evidence for reliability and validity, and practical guidance, issues covered in the PDI-IV’s manual.

The International Personality Disorder Examination (IPDE; Loranger, 1999) is a structured interview that is able to assess PDs across both the DSM-IV and ICD-10. The IPDE scores individuals dimensionally (“negative”, “probable” and “definite”) (Rogers, 2001). It demonstrates excellent inter-rater reliabilities (.81 to .92; Lenzenweger, 1999). The IPDE’s strengths are that it can assess PDNOS, providing that the individual has scored on at least 10 different PD criteria, and has internationally validity as it was developed alongside the WHO. It is used in the Longitudinal Study of Personality Disorders (LSPD).

The International Personality Disorder Examination Questionnaire (IPDEQ, Loranger, 1999) is a screening tool to be used alongside the IPDE. Consisting of 99 items, it assesses PDs across six scales that represent everyday functioning (work, self, interpersonal relationships, affect, reality testing and impulse control). Loranger suggests that if an individual scores highly on at least three of the scales, then the IPDE should be subsequently used. Slade, Peters, Schneiden and Andrews (2006) found the IPDEQ to accurately predict anti-social PD, and Lewin, Slade, Andrews, Carr and Hornabrook (2005) found that the IPDEQ’s scores were not only psychologically sound, but also provided a good index for the likelihood of developing a PD in an epidemiological study, therefore suggesting its use outside of clinical environments. Fountoulakis et al. (2002) compared IPDEQ scores with the PD diagnoses, and found that the onset of specific PDs had strong phi coefficients > .91, suggesting good reliability.

The Iowa Personality Disorder Screen (IPDS; Langbehn et al., 1999) is an 11 item semi-structured interview which is essentially a screening instrument measuring DSM-III PDs. The interview only lasts around five minutes. The original authors found sensitivity validities being as high as 92% and specificity validities as high as 72%—a finding further supported by Trull and Amdur (2001) in a non-clinical population. Olsson, Sorebo, and Dahl (2011) also found that within psychiatric outpatients, the 11 items held an average internal consistency of .70, with positive prediction power averaging at .66 and correctly classifying PDs in comparison to the SCID-II on average at 64%. Similarly to the Personality Beliefs Questionnaire (PBQ-SF; Beck & Beck, 1991; see further in this paper), the IPDS stands up well against other instruments due to its extreme brevity and good statistical properties.

The Shedler-Westen Assessment Procedure 200 (SWAP-200). There are various versions of this measure including the SWAP-II and SWAP-II-A (Westen & Shedler, 2007), an adolescent version. Clinically experienced interviewers are required to sort the 200 personality descriptive items into 8 categories from most descriptive to not descriptive or irrelevant. A computer program then reports DSM-IV PD diagnoses, personality diagnoses for alternative, empirically derived personality syndromes, and dimensional trait scores. Westen and Shedler (2007) provide both reliability and validity evidence of both versions of the test.

3.2. Questionnaires (See Table 3)

The Personality Diagnostic Questionnaire-4 (PDQ-4, Hyler, 1994) consists of 99 items, that measure all 10 of the DSM-IV PDs. Okada and Oltmanns (2009) found acceptable test-retest validities over different three time periods, with an average of .67. Abdin et al. (2011) investigated the PDQ-4’s efficacy of being used as a screening instrument for mentally ill inpatients, by comparing its validity with the SCID-II. Generally there was moderate agreement between the two instruments with Kappa levels no lower than .50. They did find that the PDQ-4 held high sensitivity (the probability of the likelihood of being diagnosed with a PD across both instruments) and low specificity (the likelihood that both instruments detect an absence of PD). When looking at specific PDs, sensitivities ranged from poor (dependent PD; .30) to good (antisocial PD; .78). Abdin et al. (2011) who used a translation in Singapore concluded that the PDQ-4 is statistically robust enough to be used as a screening instrument.

There are earlier versions of the PDQ-4: the PDQ (Hyler, Rieder, Williams, Spitzer, Hendler, & Lyons, 1988) and the PDQ-R (Hyler & Rieder, 1987). The PDQ consists of 163 items, but it is in accordance with the DSM-III. Participants respond to items with true/false answers. The PDQ has been found to have poor levels of internal consistency (.43 to .70; Hyler & Lyons, 1988) and test-retest reliabilities larger than .56 (Hurt, Hyler, Frances, Clarkin, & Brent, 1984). The PDQ-R was created as a response to the changes found within the
Table 3. A review of measures that attempt to measure specific disorders.

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Name of Test</th>
<th>Author</th>
<th>Items</th>
<th>Scales</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borderline</td>
<td>Zanarini Rating Scale For Borderline Personality Disorder (ZAN-BPD)</td>
<td>Zanarini (2003).</td>
<td>10</td>
<td>1) Affective</td>
<td>Internal consistency of the ZAN-BPD was found to be high (Cronbach’s α = .85)—Zanarini, 2003.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) Cognitive</td>
<td></td>
<td>The convergent validity of the ZAN-BPD with the scales of the SCL-90, both of which assess symptoms in the past week, is high (p &lt; .001). This is particularly so for the overall borderline psychopathology scores and the scores pertaining to affective symptomatology—Zanarini, 2003.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) Impulsive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4) Interpersonal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Revised Diagnostic Interview for Borderlines (DIB-R)</td>
<td>Zanarini, Gunderson, Frankenburg, &amp; Chauncey (1989)</td>
<td>10</td>
<td>1) Affect</td>
<td>Excellent kappas were found in each of the three inter-rater reliability substudies for the vast majority of borderline symptoms assessed by the DIB-R. More specifically, one-third of the BPD symptoms assessed had a kappa in the excellent range and the remaining two-thirds had a kappa in the fair-good range (.57-.73). More specifically, all five dimensional measures of borderline psychopathology had intraclass correlation coefficients in the excellent range for all four reliability substudies. Taken together, the results of this study suggest that both the borderline diagnosis and the symptoms of BPD can be diagnosed reliably when using the DIB-R—Zanarini, Frankenburg, &amp; Vujanovic, 2002.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2) Cognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) Impulse action patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4) Interpersonal relationships</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD)</td>
<td>Zanarini, Vujanovic, Parachini, Boulanger, Frances, Frankenburg, &amp; Hennen (2003)</td>
<td>10</td>
<td>1) Dependent</td>
<td>Test-retest reliability of the MSI-BPD was found to be good. More specifically, a highly significant correlation was found between the number of items endorsed at Time 1 and Time 2 (Spearman’s rho = .72, p &lt; .0001) (based on the total ranks from the two administrations of the scale—Zanarini et al., 2003.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) Avoidant</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4) Histrionic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Themes of dependency, helplessness, distrust, fears of rejection/abandonment/losing emotional control, and extreme attention-seeking behaviour.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Borderline Personality Disorder Beliefs Scale</td>
<td>Butler, Brown, Beck, &amp; Grisham, (2002)</td>
<td>14</td>
<td></td>
<td>The internal reliability of the 14 items for the 84 BPD patients was moderately high (alpha = .89)—Butler, Brown, Beck, &amp; Grisham, 2002.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The full BPQ is a highly reliable test. Subscale reliabilities are generally satisfactory for scales of this length, though with a few exceptions, particularly Quasi-Psychotic States. Several of the subscales show high positive skew, noting that the standard error of skew for the three samples was respectively .18, .16, and .20. No transformations were performed, and the skewness of the data should be taken into account when considering the parametric statistics reported in subsequent analyses—Poreh, Rawlings, Claridge, Freeman, Faulkner, & Shelton (2006).

Discriminant validity was examined in Sample 2. Pearson correlations were calculated between the BPQ sum score and the total scores on both the MMPI-2 STY subscale (r = .48) and the SPQ (r = .45). While both coefficients were highly significant, their moderate size suggests satisfactory discriminant validity for the BPQ. Convergent validity was examined by correlating the BPQ with the MMPI-2 BPD in Sample 2, yielding a high correlation of .85. The BPQ was correlated with the STB in Samples 3 and 4, yielding coefficients of .72 and .78 respectively. These results suggest high convergent validity for the BPQ—Poreh, Rawlings, Claridge, Freeman, Faulkner, & Shelton (2006).

The BEST exhibits face validity by assessing thoughts and behaviours typical of BPD. Evidence for content validity has been indirectly established because the items are derived from the DSM-IV. At the screening visit, the BEST correlated strongly with the ZAN-BPD score, SCL-90-R total score, the SAS total score, the CGI severity score, and both the GAS and BDI scores—Pfohl, Blum, John, McCormick, Allen, & Black (2009).
The FFBI subscales were shown to have good internal consistency—Mullins-Sweat, Edmundson, Sauer-Zavala, Lynam, Miller, & Widiger (2012).

The FFBI subscales were shown to have good convergent, discriminant, and incremental validity—Mullins-Sweat, Edmundson, Sauer-Zavala, Lynam, Miller, & Widiger (2012).

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Anxious Uncertainty</td>
<td></td>
</tr>
<tr>
<td>2) Dysregulated Anger</td>
<td></td>
</tr>
<tr>
<td>3) Despondence</td>
<td></td>
</tr>
<tr>
<td>4) Self-Disturbance</td>
<td></td>
</tr>
<tr>
<td>5) Behaviour Dysregulation</td>
<td></td>
</tr>
<tr>
<td>6) Affective Dysregulation</td>
<td></td>
</tr>
<tr>
<td>7) Fragility</td>
<td></td>
</tr>
<tr>
<td>8) Dissociative Tendencies</td>
<td></td>
</tr>
<tr>
<td>9) Distrustfulness</td>
<td></td>
</tr>
<tr>
<td>10) Manipulativeness</td>
<td></td>
</tr>
<tr>
<td>11) Oppositional</td>
<td></td>
</tr>
<tr>
<td>12) Rashness</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paranoid Personality Disorder Features Questionnaire</th>
<th>Mullins-Sweat, Edmundson, Sauer-Zavala, Lynam, Miller, &amp; Widiger (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Mistrust/ suspiciousness</td>
<td></td>
</tr>
<tr>
<td>2) Antagonism</td>
<td></td>
</tr>
<tr>
<td>3) Introversion</td>
<td></td>
</tr>
<tr>
<td>4) Hypersensitivity</td>
<td></td>
</tr>
<tr>
<td>5) Hypervigilance</td>
<td></td>
</tr>
<tr>
<td>6) Rigidity</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Avoidant UNABLE TO FIND</th>
<th>Mullins-Sweat, Edmundson, Sauer-Zavala, Lynam, Miller, &amp; Widiger (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12) Internal consistency estimates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Internal consistency estimates were high, with coefficient alpha equal to .88. Corrected item-to-total correlations ranged from .40 to .74, and a mean interitem correlation of .41 suggests homogeneity for the IM-SZ. Interrater agreement was estimated with an average, one-way random effects ICC, based on 123 cases in which interviewer and observer ratings were available, and was acceptable, $r = .69$, indicating that the traits measured are relatively robust across raters. Individual item ICCs were somewhat lower; the mean item ICC was .53—Kosson, Blackburn, Byrnes, Park, Logan, & Donnelly (2008).

Results provide preliminary indications that the measure of SZPD correlates with other measures in specific ways that suggest preliminary construct validity. Scores on the IM-SZ appear independent of age, education, and SES yet correlated negatively with IQ. Similarly, IM-SZ scores were generally independent of criminal activity yet correlated in predicted ways with ratings on several components of psychopathy—Kosson, Blackburn, Byrnes, Park, Logan, & Donnelly, 2008.
### Narcissistic Personality Inventory-16 (NPI-16)


1. Narcissism-consistent

α = .72; mean inter-item correlation = .13—Ames, Rose, & Anderson (2006).

Divergent validity: Narcissism was not significantly correlated with dispositionalism.

Convergent validity: Men scored higher in narcissism than women.

Predictive validity: Narcissism was significantly and positively correlated with participants’ estimates of their own relative performance in a judgment task—Ames, Rose, & Anderson (2006).

### Narcissistic Personality Inventory-40 (NPI-40)

Raskin & Terry (1988)

1. Authority
2. Exhibitionism
3. Superiority
4. Entitlement
5. Exploitativeness
6. Self-sufficiency
7. Vanity.

In Raskin & Terry (1988), alpha composite reliability scores of .83, .74, .80, and .90 were calculated.

Construct validity: In order to investigate the validity of the seven NPI component scales, we correlated them with the IPAR criterion variables. In the observer domain, five of the seven component scales showed significant positive correlations with the Trait Ranking for Narcissism. These correlations were Authority (r = .44, p < .001), Exhibitionism (r = .42, p < .001), Self-Sufficiency (r = .34, p < .01), Exploitativeness (r = .28, p < .05), and Superiority (r = .24, p < .05) —Raskin & Terry (1988).

### Narcissistic Personality Inventory

Raskin & Hall, 1979

54 (each item being a pair of statements)

1. Leadership/Authority (LA)
2. Superiory/Arrogance (SA)
3. Self-Absorption/Self-Admiration (SS)
4. Exploitativeness/Entitlement (EE)

The reliability and validity of the NPI has been established in clinical (e.g., Priftera & Ryan, 1984) and non-clinical samples (Emmons, 1984; Raskin & Hall, 1981).

Validity evidence includes the successful differentiation between narcissistic and normal adolescents (t = 3.93; p < .001)—Mullins & Kopelman (1988).

Solomon (1982)—confirms the validity of the MMPI and are consistent with O. Kernberg’s (1976) emphasis on the impoverishment of object relations in the dynamics of narcissistic personalities and with H. Kohut’s (1971, 1976) view of nightmares as a reflection of self-feeling, self-state, or level of narcissistic development.
Continued

<table>
<thead>
<tr>
<th>Scale</th>
<th>Author(s)</th>
<th>Year</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murray’s Narcissism Scale</td>
<td>Murray</td>
<td>1938</td>
<td>itori (1985) found the alpha reliability of this scale to be .76.</td>
</tr>
<tr>
<td>The Hypersensitive Narcissism Scale (HSNS)</td>
<td>Hendin, Cheek</td>
<td>1997</td>
<td>Cronbach’s alpha for the HSNS was .82—Arble, 2008. The reported reliability of the HSNS was Covert Narcissism good, with alpha coefficients ranging from .72 to .76. Furthermore, the HSNS, the Narcissism-Hypersensitivity Scale, and the NPD demonstrated significant positive correlations with each other (r = .50) and negative correlations (r = −.18) with the NPI (Raskin &amp; Hall, 1979).</td>
</tr>
<tr>
<td>Antisocial Personality Questionnaire</td>
<td>Blackburn &amp; Fawcett</td>
<td>1999</td>
<td>1) Self-Control, 2) Self-Esteem, 3) Avoidance, 4) Paranoid Suspicion, 5) Resentment, 6) Aggression, 7) Deviance, 8) Extraversion</td>
</tr>
<tr>
<td>Psychopathy Checklist (PCL)</td>
<td>Hare</td>
<td>1980</td>
<td>Coefficient alpha ranges from .77 to .87 for the total sample, from .79 to .88 for patients, and from .75 to .84 for normals, suggesting satisfactory internal consistencies for all scales—Blackburn &amp; Fawcett, 1999.</td>
</tr>
</tbody>
</table>

The HSNS was found to have a strong negative correlation with the RSES, a measure of self-esteem. Conversely, the HSNS was found to have a strong positive correlation with the measure of social inhibition FNE, a measure of shame (PFQ2), and a measure of masochism (SDPS). The HSNS also produced a moderate positive correlation with the Social Incompetence scale and a slight positive correlation with the Egocentricity scale. Thus, elevations on HSNS seemed indicative of a wide range of object related disturbances—Arble, 2008.

Comparing correlational patterns for the PNI and NPI, we conclude preliminary evidence exists to support the construct validity of the PNI as a measure of pathological narcissism. Results for the PNI provide additional evidence for its construct validity. Several PNI scales assess grandiose characteristics reflecting, in part, a range of vindictive, domineering, intrusive, and overly-nurturant interpersonal problems. Several PNI scales assess vulnerable characteristics reflecting, in part, cold, avoidant, and exploitable interpersonal problems. We conclude that the associations between the PNI and IIP-C contribute evidence in support of the PNI’s construct validity—Pincus et al., 2009.

The APQ scales clearly differentiate the mentally disordered offender and normal groups, but also discriminate within the offender sample. AMANOVA reveals a highly significant overall difference between the MI, PD and normal groups (Rao’s R (16, 712) = 15.45, p < .0001), and univariate ANOVAs indicate significant effects for all scales (Blackburn & Fawcett, 1999).
Continued

Hare Psychopathy Checklist-Revised (PCL-R)  
Hare (1991)  

Factors:  

To assess inter-rater reliability, intra-class correlation coefficients (ICCs; Shrout & Fleiss, 1979) were calculated for subsamples of subjects rated by more than one rater (see sample descriptions for №). They ranged from .78 to .94 ($M = .86$) for a single rating and from .87 to .97 ($M = .93$) for the average of two ratings (computations based on the Spearman-Brown Prophecy Formula). Internal consistency was assessed by Cronbach’s coefficient alpha, which ranged from .85 to .89 ($M = .88$), and by the mean inter-item correlation, which ranged from .23 to .30 ($M = .27$). These results indicate that the PCL-R, like the PCL, can be considered a homogeneous, uni-dimensional scale (Harpur et al., 1989)—Hare, Harpur, Hakstian, Forth, Hart, & Newman, 1990.

The associations found between the convergent validity measures and the PCL–R when it was used dimensionally support the validity of the instrument. As predicted, on-Psychoticism, SRPS, SMAST, MPQ-Constraint, and SO scale scores were significantly related to PCL–R scores. Overall, the pattern of results for both the Caucasian and African American women provided support for the convergent validity of the PCL–R in this sample—Vitale, Smith, Brinkley, & Newman, 2002.

Levenson’s self-report psychopathy scale (SRPS)  

Cronbach’s alphas for the total sample were .85 for the total SRPS, .83 for SRPS factor 1 and .69 for SRPS factor 2—Brinkley, Schmitt, Smith, & Newman, 2000.

The SRPS and the PCL–R were significantly correlated and both showed similar patterns of correlations to measures of substance abuse and criminal versatility. SRPS was predictive of performance of a passive avoidance task—Brinkley, Schmitt, Smith, & Newman, 2000.

Psychopathic Personality Inventory Revised (PPI-R)  

Internal consistency is adequate for the PPI-R Total score and the PPI-R Content scale scores, with coefficient alpha ranging from .78 - .92 for the community/college sample. For the offender sample, internal consistency estimates for the Total and Content scale scores ranged from .72 - .84—Lilienfeld & Widows, 2005.

Evidence of construct validity was obtained via significant correlations between the PPI-R and self-report measures of pathological and nonpathological personality functioning, DSM-IV Antisocial Personality Disorder, interpersonal problems, sensation-seeking, substance use, and offense variables—Lilienfeld & Widows, 2005.

The Elemental Psychopathy Assessment  
Lynam, Gaughan, Miller, Mullins-Sweat, & Widiger (2011)  

Antagonism:  
1) Distrust  
2) Manipulation  
3) Self-centeredness  
4) Opposition  
5) Arrogance  
6) Callousness  
Conscientiousness:  
1) Disobliged  
2) Impulsivity  
3) Rashness  
Extraversion:  
1) Coldness  
2) Dominance  
3) Thrill-Seeking  
Neuroticism:  
1) Unconcern  
2) Anger  
3) Self-Contemptment  
4) Self-Assurance  
5) Urgency  
6) Invulnerability  

Final scales were fairly reliable. Seventeen of the 18 scales had alpha coefficients above .7; 14 of these were above .8. The average alpha for the 18 scales was .81. The scales were also fairly homogeneous. Average interitem correlations ranged from .16 for Arrogance to .48 for Dominance, with an average of .34. Average item-total correlations ranged from .31 for Arrogance to .62 for Thrill-Seeking, with an average of .51.

Concurrent validity was shown by correlations with other self-report psychopathy measures; SRP, LSRP, PPI-R. The measure contains two validity scales: Infrequency scale, and a Too Good to Be True scale. It has an alpha level of .95.
In the present sample, the KR-20 internal consistency reliability coefficient for the full SPQ-B was .83 (for biological relatives) and .82 in non-psychiatric comparison participants—Compton, Chien, & Bollini, 2007.

High convergent validity (.59 to .81), discriminative validity (.63) and criterion validity (.68)—Raine (1991).

Correlations between SPQ-B total and subscale scores and the derived SCID-II SPD criteria scores were examined next in a subsample of participants for whom SCID-II data were available (from the first study; n = 58). The SPQ-B total score was significantly correlated with the SCID-II SPD criteria total score ($p = .49$, pb.001).—Compton, Chien, & Bollini, 2007.
Short forms of Wisconsin Schizotypy Scale


60 1) The magical ideation scale.
2) Perceptual aberration scale.
3) Revised social anhedonia scale.
4) Physical anhedonia scale.

The short forms had higher alpha values than would be predicted by the Spearman-Brown formula, which suggests that they retained the relatively more effective items from full scales—Winterstein, Silvia, Kwapił, Kaufman, Reiter-Palmon & Wigert, 2011.

All scales have highly adequate internal consistency as assessed by coefficient alpha.

The positive and negative symptom dimensions correlated modestly (r = .13)—Winterstein, Silvia, Kwapił, Kaufman, Reiter-Palmon & Wigert, 2011.

Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE)

Mason, Claridge, & Jackson (1995)

1) Unusual experiences
2) Cognitive disorganisation
3) Introvertive anhedonia
4) Impulsive non-conformity

Results indicate that the OCI-R is a psychometrically sound and valid measure of OCD and its various symptom presentations. The instrument also evidenced good convergent validity, and performed well in discriminating OCD from other anxiety disorders The OCI-R is recommended as an empirically validated instrument that can be used in a range of clinical and research settings for research on OCD—Abramowitz & Deacon, 2006.

Test-retest reliability:

Spearman's correlations—Patients diagnosed with OCD ranged from .74 to .91. Nonanxious controls ranged from .57 to .87—Reid, Rosen, Arnold, Larson, Mason, Murphy, Storch, 2011.

Discriminative validity: Comparisons of means for the total distress score revealed that the OCD group reported greater distress on the OCI than did the other 3 groups. Comparisons of means revealed that the OCD group reported more frequent OCD symptoms than did the remaining 3 groups on all but the Hoarding subscale.

Convergent validity: The validity of the OCI total distress and frequency scales was assessed by correlating (Pearsons correlations) the OCI scores of the OCD patients with scores obtained on other measures of OCD symptoms: Y-BOCS, CAC, MOCI. For the OCD sample, the coefficients were quite high, ranging from .41 to .93 with MOCI—Foa, Kozak, Salkovskis, Coles, Amir, 1998.

Obsessive-Compulsive Inventory (OCI) for the controls, distress (OCD, r = .87; controls, r = .89) and frequency (OCD, r = .84; controls, r = .90) total scores.

Test-retest reliability for the subscales exceeded .80, with the exception of Ordering distress (r = .77) and Ordering frequency scores (r = .79) in the OCD sample and the Doubting distress (r = .77) and Hoarding distress (r = .68) scores in the control sample—Foa, Kozak, Salkovskis, Coles, Amir, 1998.
Obsessive Beliefs Questionnaire (OBQ)  

1. Overestimation of threat.  
2. Tolerance of uncertainty.  
3. Importance of thoughts.  
4. Control of thoughts.  
5. Responsibility.  
6. Perfectionism.

Cronbach's alpha for those with OCD and control samples were between .80 and .96—Obsessive Compulsive Cognitions Working Group, 2001.

The multivariate effect was significant, Wilks' Lambda = .039, F(30, 1806) = 75.28, \( p < .05 \). Univariate analyses were then conducted for each subscale. All of these one-way ANOVAs were also significant. Follow up Student-Newman-Keuls tests showed that the OCD group scored significantly higher than the other four groups in all cases with one exception: OCDs did not score higher than anxious controls on Perfectionism. Student controls scored higher than all community controls on Overestimation of Threat, and higher than Greek community controls on Responsibility. These results support the known groups validity of the scales by distinguishing people who have OCD from those who do not—Sica, Coradeschi, Sanavio, Dorz, Manchisi & Novara (2004).

The OCD group scored significantly higher than the other groups on the VOCI total score, Contamination, Checking, Just Right, and Indecisiveness subscales.

The OCD group scored significantly higher than the non-clinical controls but not the A/D control group on the Obsessions and Hoarding subscales—Thordarson, Radomsky, Rachman, Shafran, Sawchuk & Hakstian, 2004.

Convergent and discriminant validity: OCD sample—Total score highly correlated with other multidimensional self-report measures of OCD and moderately correlated with measures of anxiety, depression and worry. Convergent and discriminant validity were excellent for the Contamination and Checking subscales, but weaker for Obsessions. Discriminant validity of Hoarding was excellent, while its convergent validity was not examined—Gönner, Ecker, Leonhart, & Limbacher, 2010.

Scores of the VOCI and its subscales demonstrated acceptable to excellent internal consistency (\( .81 \leq \alpha \leq .95 \))—Gönner, Ecker, Leonhart, & Limbacher 2010b. The VOCI and its subscales appear to have excellent test-retest reliability in the OCD sample, with all coefficients .9 or above, and with the one-sided .95 confidence interval >.81 in all cases, despite, on average, a long test-retest interval. For the students, on the other hand, test-retest reliability is poor, ranging from .5 to .6. In the student sample, these low correlations may be due to range restriction. Most of the students had mean VOCI item scores between 0 and 1, with the exception of four participants, who appeared to be bivariate outliers. Excluding the four outliers improved the test-retest correlations for some scales (e.g., VOCI total, \( r = .62, p < .001 \)), but not others (e.g., VOCI Checking, \( r = .44 \)).—Thordarson, Radomsky, Rachman, Shafran, Sawchuk & Hakstian, 2004.
Continued

<table>
<thead>
<tr>
<th>Test</th>
<th>Authors</th>
<th>Sample Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vancouver Obsessional-Compulsive Inventory Revised (VOCI-R)</td>
<td>Gönner, Ecker, Leonhart, &amp; Limbacher (2010)</td>
<td>30</td>
<td>Scores of the 30-item total scale and all symptom subscales showed good to excellent internal consistency (.82 ≤ α ≤ .95) — Gönner, Ecker, Leonhart &amp; Limbacher, 2010.</td>
</tr>
<tr>
<td>Maudsley Obsessional-Compulsive Inventory (MOCI)</td>
<td>Hodgson &amp; Rachman (1977)</td>
<td>30</td>
<td>Test-retest reliability, 50 participants, completing the questionnaire one month apart (r = .80) — Hodgson &amp; Rachman, (1977).</td>
</tr>
<tr>
<td>Padua-Inventory-Palatine Revision (PI-PR)</td>
<td>Gönner, Ecker &amp; Leonhart (2010).</td>
<td>24</td>
<td>The total scale and all subscales had acceptable to excellent internal consistency (.78 ≤ α ≤ .93) Acceptable corrected item-total correlations (r &gt; .40) were found for the PI-PR (with one exception: Ruminationsubscale, Item 121) — Gönner, Ecker &amp; Leonhart, 2010</td>
</tr>
<tr>
<td>Padua Inventory (PI)</td>
<td>Sanavio (1988)</td>
<td>60</td>
<td>The α coefficient was .90 in male and .94 in female participants—Coles (2003). Test-retest correlations were 0.78 for males, and .83 for females who filled out the inventory twice at a 30-day interval—Sanavio (1988).</td>
</tr>
<tr>
<td>Leyton Obsessional Inventory</td>
<td>Cooper (1970)</td>
<td>69</td>
<td>The Leyton Obsessional Inventory (LOI) is a card-sort task that has shown good reliability and validity estimates (Cooper, 1970; Taylor, 1995).</td>
</tr>
</tbody>
</table>

Convergent validity: Scores of the total scale and the five symptom subscales demonstrated excellent convergent validity. All VOCI-R subscales showed moderate to high correlations with the OCI-R total scale (r, range = .47 - .57) and with the PI-PR total scale (r, range = .43 - .55).

Discriminant validity: the VOCI-R total scale (r, range = .34-.39) and the five symptom scales with measures of depression, anxiety and worries (r, range = .07-.31) indicate only a low to moderate construct overlap— Gönner, Ecker, Leonhart, & Limbacher 2010

For 30 obsessionals, we had Leyton Inventory Scores (Cooper, 1970) and found that our “total obsessionality” score correlated significantly with the Leyton Symptom Score (.6)—Hodgson & Rachman, (1977).

In the OCD sample the PI-PR and its subscales demonstrated good to excellent convergent validity with corresponding OCI-R scales. On the whole, the PI-PR and its subscales demonstrated very satisfying divergent validity with measures of depression, anxiety, and worries in the OCD sample, except for PI-PR Ruminating and PI-PR Harming Impulses and Obsessions.

In the anxiety disorder sample, the PI-PR and its subscales demonstrated good to excellent convergent validity with corresponding OCD scales— Gönner, Ecker & Leonhart, 2010

When assessing the convergent and divergent validity of the PI substantial correlations were found between related scales of the PI and the MOCI. The substantial correlation between the total PI scores and the two subscales of the SCL-90-R, Sensitivity and Hostility, are in line with the findings reported in Millar (1983) —van Oppen (1992).

“Within the limits imposed by these selected criterion groups, this inventory is valid, since it differentiates between the obsessional patients and the normals with very little overlap” —Cooper, 1970.
Continued

The total Yale-Brown Scale score was significantly correlated with two of three independent measures of obsessive-compulsive disorder and weakly correlated with measures of depression and of anxiety in patients with obsessive-compulsive disorder with minimal secondary depressive symptoms. Results from a previously reported placebo-controlled trial of fluvoxamine in 42 patients with obsessive-compulsive disorder showed that the Yale-Brown Scale was sensitive to drug-induced changes, and that reductions in Yale-Brown Scale scores specifically reflected improvement in obsessive-compulsive disorder symptoms. Together, these studies indicate that the 10-item Yale-Brown Scale is a reliable and valid instrument for assessing obsessive-compulsive disorder symptom severity, and that it is suitable as an outcome measure in drug trials of obsessive-compulsive disorder—Goodman, Price, Rasmussen, Mazure, Fleischmann, Hill, Heninger & Charney (1989).

Spearman rank correlations were used to compute convergent validity in a sub-group of obsessive compulsive patients ($r = .96$). The Obsessive Compulsive Thoughts Checklist correlated positively with the Compulsive Activity Checklist ($\rho = .62; p < .0001$). The total score was also positively correlated with the Yale-Brown Obsessive Compulsive Scale (YBOCS total scale: $\rho = .42; p < .0001$; YBOCS obsession scale: $\rho = .40; p < .0001$; YBOCS compulsion scale: $\rho = .37; p = .0002$). The factor structure was studied both on the sample which included agoraphobic patients and controls ($n = 141$) and on the obsessive compulsive patients only ($n = 122$). In both analyses, three identical factors were found:

- factor 1: checking/perfectionism-orderliness
- factor 2: responsibility/dread of harming others
- factor 3: washing/contamination.

Results support the validity and the internal consistency of the Obsessive Compulsive Thoughts Checklist—Bouvard, Cottraux, Mollard, Arthur, Lachance, Guerin, Sauteraud & Yao (1997).
Continued

**Five Factor Obsessive-Compulsive Inventory (FFOCI)**

1. Perfectionism
2. Fastidiously
3. Punctiliousness
4. Workaholism
5. Doggedness
6. Ruminative Deliberation
7. Detached Coldness
8. Risk Aversion
9. Constricted
10. Inflexibility
11. Dogmatism
12. Excessive Worry

The 12 scales demonstrated convergent correlations with established measures of OCPD and the FFM—Samuel, Riddell, Lynam, Miller, & Widger, 2012.

**Dependent Personality Questionnaire (DPQ)**

Dependent Personality Inventory (DPI) — Huber (2005)

1. Difficulty making everyday decisions.
2. Needs others to assume responsibility for most major areas of life.
3. Has difficulty expressing disagreement with others because of fear of loss or support from others.
4. Has difficulty initiating projects or doing things on own because of lack of confidence in judgment or abilities rather than a lack of motivation or energy.
5. Goes to excessive lengths to obtain nurturance and support from others to the point of volunteering to do things that are unpleasant.
6. Feels uncomfortable or helpless alone because of exaggerated fears of being unable to care for self.
7. Urgently seeks another relationship as a source of care and support when a close relationship ends.
8. Is unrealistically preoccupied with fears of being left to take care of self.

The internal consistency of the revised Dependent Personality Inventory (DPI-R) to .90, making the test highly reliable.— Huber (2005).

**Dependent Personality Questionnaire (DPQ)**

Dependent Personality Questionnaire — Tyrer, Morgan, & Cicchetti (2004).

The DPQ was also a good predictor of the diagnosis of dependent personality disorder, with sensitivity, specificity, predicted positive, and predicted negative accuracies of 87%.— Tyrer, Morgan, & Cicchetti (2004).

**Minnesota Multiphasic Personality Inventory-2 Social Introversion Subscales (MMPI-2 Si1, 2, 3)**

1) Difficulty making everyday decisions.
2) Needs others to assume responsibility for most major areas of life.
3) Has difficulty expressing disagreement with others because of fear of loss or support from others.
4) Has difficulty initiating projects or doing things on own because of lack of confidence in judgment or abilities rather than a lack of motivation or energy.
5) Goes to excessive lengths to obtain nurturance and support from others to the point of volunteering to do things that are unpleasant.
6) Feels uncomfortable or helpless alone because of exaggerated fears of being unable to care for self.
7) Urgently seeks another relationship as a source of care and support when a close relationship ends.
8) Is unrealistically preoccupied with fears of being left to take care of self.

Respective reliabilities for Si1, Si2, and Si3 of .82, .77, and .77 for college men and .82, .75, and .77 for college women—Ben-Porath et al. (1989).

Construct validity is satisfactory for the DPI as our results support the findings of previous studies that suggest two distinct constructs form dependent personality disorder—Gude, Hoffart, Hedley & Ro, 2004.
DSM-III-R. The PDQ-R consists of 152 items that are also answered by a forced choice true/false paradigm. Uehara, Sakado and Sato (1997) found test-retest reliabilities ranging from .76 to 1.0. Hyler, Skodol, Oldham, Kellman and Doidge (1992) found satisfactory levels of sensitivity and moderate levels of specificity for most of the PDs. They did note however that the PDQ-R did yield many false-positives, concluding that this cannot be used diagnostically as a replacement of a structured interview.

The Coolidge Axis-II Inventory (CATI; Coolidge, 1984) is a 225 item self-report scale, measuring DSM-III-R PDs (13 PDs) alongside three axis-I disorders (anxiety, depression and brain dysfunction). Coolidge and Merwin (1992) found an average test-retest reliability of .90 and a Cronbach Alpha level of .71 suggesting that this instrument is highly reliable in its efficacy to predict PDs. Watson and Sinha (1996) did find some gender bias regarding the antisocial PD scale and age biases for younger respondents (17 - 24 years old) in comparison to older respondents (25 - 57 years old). Ramaniah and Sharpe (1998) demonstrated through varimax rotation that each of the CATI’s scales could be mapped onto the five factor model of personality, providing support for its dimensional nature. Silberman, Roth, Segal and Burns (1997) compared the convergent validities between the CATI and the Millon Clinical Multiaxial Inventory II (MCMI-II; Millon, 1983) in a sample of elderly inpatients. Generally they found that validities between the two widely ranged, however 7 out of the 13 PDs had coefficients higher than .54.

The Short Coolidge Axis-II Inventory (SCATI; Coolidge, 2001) is a 70 item abbreviation of the CATI. Watson and Sinha (2007) found internal reliabilities to range from .46 to .72 (averaging at .61), which is impressive in comparison to the CATI’s (.70 to .86, averaging at .78; Watson and Sinha, 1996). Much like the CATI, evidence for gender biases in the SCATI were also found, with Cohen effect sizes ranging from .25 (schizoid PD) to 1.13 (sadistic). The authors were also able to map the SCATI’s factors onto Eysenck and Eysenck’s (1975) PEN model of personality. Lastly, the authors used confirmatory factor analysis (CFA) to demonstrate that the SCATI and the CATI are convergent.

The Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon, Millon, Davis & Grossman, 2009) is a 175 item questionnaire that consists of 14 PD scales (11 moderate PD scales and 3 severe personality pathology scales), 10 clinical syndrome scales, five correction scales (designed to identify random response styles and modifying indices) and 42 Grossman personality facet scales. Participants respond via true/false answers. There are two earlier versions of the MCMI-III, the MCMI (Millon, 1983) and the MCMI-II (Millon, 1987) that hold acceptable psychometric properties, but for the purpose of this review the MCMI-III will be of primary focus due to its accordance to the DSM-IV. One standout feature of MCMI-III is its theoretical anchoring.

The MCMI-III is built upon four domains of evolutionary theory: existence, adaptation, reproduction and abstraction. Another differentiating factor of the MCMI-III is that diagnoses are made based on respondents scoring higher than the base rate score of 84. Retzlaff (1996) investigated the MCMI-III’s diagnostic validity and found that the instrument’s positive predictive power varied widely. For example, validity coefficients ranged from .00 (sadistic personality) to .27 and .32 (narcissistic and histrionic respectively). This suggested that the instrument’s predictive validity could be improved. Millon, Davis and Millon (1997) however report very different positive predictive validities, ranging from .33 (delusional) to .93 (drug abuse), with an average of .64 across all PD and axis I disorders. Millon et al. (1997) also report Kappa levels ranging from .23 (anxiety) to .84 (paranoid), suggesting highly discrepant levels of test-retest reliability.

The Personality Beliefs Questionnaire (PBQ; Beck & Beck, 1991) is a self-report questionnaire that measures beliefs associated with 9 DSM-III-R PDs across 126 items. Like the MCMI-III, the PBQ is based on a foundation of cognitive theory whereby disorders are maintained by maladaptive thinking styles. Trull et al. (1993) found that the PBQ had alphas ranging from .77 to .93, with test-retest reliabilities from .63 (passive aggressive) to .82 (paranoid)—a finding further supported by Connan et al. (2009). The PBQ has also good discriminant validity, with psychiatric patients scoring significantly higher on the scale associated with their mental illness in contrast to any other scale on the questionnaire (Beck et al., 2001). This finding was also replicated by Jones, Burrell-Hodgson and Tate (2007) who compared the PBQ to MCMI-III diagnoses on 164 psychiatric patients. In addition to the PBQ, Butler, Beck and Cohen (2007) created a short form of the PBQ (PBQ-SF) that features only 65 items and measures all Axis II PDs except borderline PD. Similarly to the PBQ, the PBQ-SF has been found to hold desirable psychometric properties: Cronbach alphas ranged from .81 to .92, and test-retest correlations ranged from .57 to .82. Both the PBQ and the PBQ-SF seem to have the most impressive levels of reliability and validity in comparison to the other self-report instruments reported in this paper, with the added bonus of having a detailed theoretical grounding. The PBQ-SF is of use to both clinicians (when it is not feasible to
administer the PBQ) and academics (the PBQ-SF can be easily inserted into a battery of self-report instruments without taking too much time or space). A recent review of the scale was very positive (Bhar, Beck, & Butler, 2011).

**Minnesota Multiphasic Personality Inventory for the DSM-III** (MMPI; Morey, Waugh, & Blashfield, 1985). The MMPI features a scale for each of the DSM-III PDs measured via 556 items. The original authors found good levels of internal consistency ranging from .67 (compulsive) to .85 (avoidant). When Schuler, Snibbe and Buckwalter (1994) investigated the MMPI’s concurrent validity by correlating its scales with diagnoses made via the MCMI, only five out of the eleven MMPI scales were positively correlated: schizoid, avoidant, dependent, histrionic and narcissistic. Although this suggests some evidence of concurrent validity, it is limited at best. Much like the CATI, clinicians and academics wanting to use this instrument should be aware that it is constructed around the DSM-III.

The **Omnius Personality Inventory** (OMNI; Loranger, 1994, 2002) is a 375 item self-report instrument that assesses all ten DSM-IV PDs, and also the traditional five-factor model and 25 normal personality traits (e.g. warmth, trustfulness, and modesty). A practical strength of the OMNI is that it is used in conjunction with computer software that can automatically generate a comprehensive evaluation report. It can therefore be deployed in a variety of clinical, occupational and academic settings easily and quickly. Despite this, the instrument is not widely used. It has shown to demonstrate acceptable levels of internal consistency and reliability (Lenzenweger, Loranger, Korfine, & Neff, 1997).

The **Schedule for Nonadaptive and Adaptive Personality** (SNAP; Clark, 1993) is a 375-item questionnaire that consists of 15 scales; 12 of which are focused on maladaptive traits and the remaining scales assess negative and positive temperament and disinhibition. The original SNAP assessed the DSM-III-R PDs and the SNAP-2 assesses both the 10 primary PDs in DSM-IV and also the PDs in the appendix. The SNAP is designed to measure dimensional correlates to DSM-IV PDs, and has also 11 diagnostic scales for the DSM-IV-TR PDs. Clark (1993) found that the SNAP scores correlated on average .54 to DSM-IV diagnoses—a coefficient surprisingly high for a self-report instrument. Melley, Oltmanns and Turkheimer (2002) investigated the test-retest reliability and predictive validity of the SNAP; there were satisfactory levels of temporal stability (.58 to .81), however the authors found mixed support for its predictive validity. The SNAP scores did modestly predict cluster A and C PD onset. SNAP-2’s predictive validity has been shown to be higher than both the NEO PI R and the DSM-IV diagnoses themselves (Morey, Hopwood et al., 2007, 2012).

The **Wisconsin Personality Disorder Inventory-IV** (WISPI; Klein et al., 1993; Klein & Benjamin, 1996). The WISPI-IV is an updated version of the WISPI-III and WISPI-III-R (Klein et al., 1993) self-report inventories using 204 items to assess DSM-IV criteria for PDs relying on an analysis of DSM items according to Benjamin’s Structural Analysis of Social Behavior model (SASB, Benjamin, 1996). Its validity with the SCID-II has been examined in adult psychiatric inpatients, showing poor convergence at the level of categorical diagnoses, but better convergent and discriminant validity for 5 out of 11 WISPI-IV dimensional scales (Smith et al., 2011).

The **Personality Assessment Inventory** (PAI; Morey, 1991), is a 344 item questionnaire presented with a four point Likert scale, used in both clinical and community settings. It measures 22 non-overlapping scales that are arranged into four clusters: clinical (this comprises DSM-IV PDs alongside addictive disorders), interpersonal (this seeks to measure interpersonal strategies), treatment (this cluster provides insight into the various efficacies of certain clinical treatments in relation to the individual’s personality, or other risk factors not held in the domain of clinical disorders) and validity (attempts to identify “faking good”, defensive or exaggerated responses to the questionnaire). Essentially it aims to assess Borderline and Antisocial Personality Disorder.

Boone (1998) found acceptable levels of internal consistency for the clusters (averaging .82) and the subscales (averaging .66). Some research has questioned its validity. Slavin-Mulford et al. (2012) researched the inventory’s convergent and discriminant validity by correlating PAI scales with the prevalence of life-events in psychiatric patients. The majority of the scales did hold meaningful correlations with at least one life-event, except for the mania and anxiety scales. The aforementioned research suggests that despite the self-report methodology, the PAI has some utility in a variety of environments. It is important to note that this instrument is not designed to be used as a tool for diagnosing disorders like the DIPD-IV; instead it provides an insight into the individual’s personality and temperament in a variety of contexts.

The **Standardised Assessment of Personality: Abbreviated Scale** (SAPAS; Moran et al., 2003) is an eight item screening interview that aims to provide a dimensional score as to whether the individual has a PD in general, rather than screening for specific disorders. Scores can range from 0 to 8; any scores higher than 3 indicate the
high possibility of a PD. The benefit of such a short interview is that it can be used in clinical environments when pressed for time. Hesse and Moran (2010) compared SAPAS scores with a variety of comprehensive personality inventories and found that SAPAS scores did regress on most DSM-IV PDs when controlling for demographic variables, suggesting convergent validity. However SAPAS scores were less likely to be associated with cluster B disorders (Antisocial, Histrionic and Borderline).

The Hogan Development Survey (HDS; Hogan & Hogan, 1997) is a self-report scale that renames the DSM-IV PDs into lay terms and is also contextualised for the work environment. Just like the Hogan Personality Inventory (HPI; Hogan & Hogan, 1992), the HDS is not a clinical instrument; instead it is mainly used for coaching, leadership development, and personnel selection. Furnham, Trickey and Hyde (2012) found various facets of the HDS to predict work success. Furnham et al. also found that the 11 scales can be clustered into three formations that are similar to clusters A, B, and C suggested by the DSM-IV. Over a dozen published studies have attested to its reliability, validity and dimensional structure (De Fruyt, Wille, & Furnham, 2013b).

The Tendances Dysfonctionelles-12 (TD-12; Rolland & Pichot, 2007) inventory examines the DSM-IV PDs, supplemented with the passive-aggressive and the depressive PDs. TD-12 is developed for the assessment of personality tendencies that may potentially harm and affect personal, social and professional functioning and is, like the HDS, mostly used in personnel selection and professional development contexts. The significance of the description of these personality tendencies for understanding behavior and performance at work is further described by Furnham and Taylor (2004) and Miller (2008).

### 3.3. Multidimensional Measures Targeting Categorical PDs

A series of dimensional approaches towards the description of personality pathology have been suggested as alternatives to the categorical DSM-IV PD diagnoses, including general dimensional models such as the FFM and more specific personality pathology dimensional representations like the SNAP, the DAPP-BQ or the DSM-5 trait model described in DSM-5 Section 3. Although these approaches advocate a dimensional instead of a categorical description of personality (pathology), in-between proposals to bridge categorical and dimensional diagnostics were proposed for most of these models. How such dimensional models translate into categorical PDs is described here.

**FFM-based measures of Axis-II PDs.** Widiger and Costa (2013) recently updated and summarized the available evidence for using a general trait model like the FFM for the description of PDs relying on the assumption that the distinction between general traits and personality pathology reflects more a quantitative than a qualitative difference (Simms & Clark, 2006). Samuel and Widiger (2008), for example, recently meta-analyzed the associations between FFM facets and DSM-IV PDs, demonstrating that most PDs could be described in terms of a particular set of FFM facets. Miller and colleagues (2005) corroborated on such findings and proposed an easy-to-use system to describe DSM-IV PDs in terms of aggregates of a specific set of FFM facets per PD. Scoring in a more extreme range on such a FFM PD count (for example 1.5 SD beyond the mean) is considered indicative of a specific PD, requiring further attention. Bastiaansen, Rossi and De Fruyt (2012) examined the concurrent validity of different FFM PD counts in an attempt to optimize this proposed scoring system. The utility of these FFM PD counts has been further supported in the meantime for both clinical and professional developmental diagnostic purposes. Miller and colleagues (2010) demonstrated the utility of this system for clinical decision making, whereas Wille et al. (2013) and De Fruyt et al. (2009, 2013b) investigated the applicability of the counts to screen for aberrant traits observable in the working population to identify dark side personality tendencies that may hinder performance or functioning at work.

**DSM-5 trait model.** A trait-set for describing personality pathology structured under the five broad dimensions of Negative Affectivity, Detachment, Disinhibition, Antagonism and Psychoticism is described in section 3 of DSM-5 (APA, 2013) for further review and evaluation. Although the labels for this five-factor structure are different from the defining dimensions of the FFM, there is strong support that at least four dimensions are conceptually and empirically related to the FFM dimensions, with some disagreement on the association between the FFM Openness to experience and the DSM-5 Psychoticism dimension (De Fruyt et al., 2013a). Krueger et al. (2012) developed the Personality Inventory for DSM-5, assessing 25 traits that can be combined to assess either six categorical PDs (borderline, avoidant, schizotypal, antisocial, obsessive-compulsive and narcissistic) or lead to a diagnosis of personality disorder trait-specified, when patients demonstrate elevated trait levels.
4. Assessments of Single Personality Disorders

The review of instruments assessing single PDs follows the order in which PDs are listed in DSM-5

4.1. Paranoid

There is much debate surrounding the paranoid personality disorder due to its shared symptomology with disorders such as schizotypal PD, narcissistic PD, and schizophrenia. The paranoid PD is marked by an entrenched mistrust towards others. Only one instrument was found that specifically diagnoses the paranoid PD.

The Paranoid Personality Disorder Features Questionnaire (Useda, 2002) is a 23-item inventory that measures six scales: mistrust, antagonism, introversion, hypersensitivity, hyper vigilance and rigidity. The authors intend for the six dimensions to closely map the current literature and DSM-IV criteria. There is a shortage of papers showing the efficacy of the instrument. However the original author did find that the instrument showed satisfactory test-retest reliabilities after a six-week period.

4.2. Schizoid

There appears to be only one measure of the schizoid PD. This may be due again to the controversy between distinguishing between schizoid and schizotypal PDs and schizophrenia. Nevertheless, the DSM-5 defines schizoid PD to be characterised by a lack of interest in social relationships and a stunted range of emotions. This contrasts with schizotypal PD that is characterised by unusual thinking styles and paranoia.

The Interpersonal Measure of Schizoid Personality Disorder (IM-SZ) (Kosson, Blackburn, Byrnes, Park, Logan, & Donnelly (2008)) consists of 12 items that measure various aspects of interpersonal interaction (e.g. rapport, absence of spontaneity in speech, poor interpersonal hygiene, and lack of verbal responsiveness). In two cross-validation studies (total N = 731), acceptable levels of internal reliability were achieved (.88), with inter-rater agreement (the inventory was completed after a semi-structured interview that was focused on the individual’s quality of held interpersonal relationships) to yield a Kappa level of .69. The authors also found the measure to hold good construct validity, but do call for further validations.

4.3. Schizotypal

As already mentioned there is a difference between schizoid and schizotypal PD (STPD), with the latter characterised by abnormal thought patterns, paranoia, and referential thinking. Our review identified seven specific measures of STPD, with each measure offering varying levels of, or focusing on specific, dimensions. Of all the instruments reviewed, the variety of STPD instruments is a strength as it allows the PD to be investigated from varying approaches and perspectives.

The Referential Thinking Scale (REF; Lenzenweger, Bennett, & Lilienfeld, 1997) is a unidimensional questionnaire, featuring 34 items that measures simple and guilty ideas of reference. A referential idea is a thought that is perceived to originate from within the individual. A simple referential idea is when the individual believes that other people are observing something about themselves that they would rather remain private. A guilty referential idea is when the individual feels that they are being accused of some wrongdoing. The authors also note that referential thinking is not exclusive to schizotypal PD, therefore the instrument is designed to measure referential thought in the clinical domain, and be wholly independent of normative referential thought processes such as self-monitoring and self-consciousness. The original paper cites Cronbach’s alpha levels to be .83 and a test-retest reliability of .86. Furthermore the original authors found convergent validity for the instrument by demonstrating high REF scores to be associated with increased levels of schizophrenia-related psychological deviation, magical ideation and perceptual aberration. The measure also held weak correlations with unrelated measures of normative self-awareness, suggesting that the REF is measuring a psychologically independent aspect of referential thought processes.

The Schizotypal Personality Questionnaire (SPQ; Raine, 1991) is a 74 item self-report questionnaire that measures all nine schizotypal traits that are laid out in the DSM-III-R. These are: Ideas of reference, excessive social anxiety, odd beliefs, unusual perceptual experiences, odd behaviour, no close friends, odd speech, constricted affect and suspiciousness. Some have questioned the heterogeneous structure of schizotypal PD (Chmielewski & Watson, 2008). The measure is designed to act as a screening tool for the diagnosis of the disorder, alongside furthering research into the field by gathering data on the individual subscales of the PD. The
measure produces an overall SPQ score as well as a score for each subscale. Participants respond via a forced-choice “yes/no” paradigm. Cronbach’s alpha level for the total measure was .91, with each subscale varying from .63 (constricted affect) to .81 (odd beliefs). Raine also found a test-retest correlation of .82. The instrument was validated by comparing its scores with the SCID-II omnibus measure of PD. Raine found that of the top 10% of scorers on the SPQ, fifty-five percent were also diagnosed with schizotypal PD by the SCID-II (suggesting high criterion validity as the bottom 10% of SPQ scorers received no SCID-II diagnosis). Convergent validity was also supported as each of the nine subscales (including the total score) held significant positive correlations with SCID-II scores (total SPQ score $r = .68$, $p < .001$).

Raine and Benishay (1995) developed the Schizotypal Personality Questionnaire-Brief (SPQ-B). It was designed to be a shorter version of the SPQ so that it could be used as a screening measure. Featuring only 22 items, it measures three dimensions: interpersonal deficits, cognitive-perceptual deficits and disorganisation (which combine to produce a total measure). The SPQ-B only takes two minutes to complete (lending its usefulness to academics and clinicians prior to a confirmatory interview). The original paper cites a Cronbach’s alpha level of .76, with Compton, Chien and Bollini (2007) finding a level of .86 within a psychiatric sample. Compton and colleagues also found the SPQ-B to be positively correlated with the SCID-II STPD subscale ($r = .49$, $p < .001$). Much like the SPQ, while it is a very robust measure psychometrically and across different samples (Axelrod, Grilo, Sanislow, & McGlashan, 2001), it still remains based on the DSM-III-R, though it remained largely unchanged in DSM-IV.

The Structured Interview for Schizotypy (SIS) (Kendler, Lieberman, & Walsh, 1989) is an interview-based research instrument for assessing symptoms of Schizotypal PD. The interview differed from other interviews at the time of its development in that it includes contextualised assessments of the pathological nature of specific symptoms (e.g. referential ideas) and symptom probes that aim to make faking good appear non-deviant. Based on the DSM-III-R, it features five kinds of items: closed-options (i.e. Likert scales), field-coded questions (i.e. an open ended question is asked and the interviewer codes the answer with a number that lays on a continuum), global symptom scores (measured via a one to seven Likert scale), specific symptom scores (the interviewer rates the responder’s severity of behaviours) and global scores on broad categories of behaviours. There are 19 sections to the SIS, measured across 16 dimensions. The interview takes around one hour to complete, and interviewers must be especially trained. The authors found inter-rater reliability to be high for the subscales ($r = .87 \pm .12$) except magical thinking ($r = .79$ to $.67$), as assessed across two clinical samples (total $N = 58$). Discriminant validity was demonstrated, as across three pilot studies, SIS scores discriminated between participants who had schizophrenic relatives and those who did not.

Wisconsin Schizotypy Scales (Winsterstein et al., 2011). The magical ideation, perceptual aberration, revised social anhedonia and physical anhedonia scales each contain 15 items and measure a single dimension. They are designed solely for academic use. The authors found each scale to have high Cronbach’s alpha levels (.84 to .88). The scales correlated with the HEXACO-60 (Ashton & Lee, 2009), measures of curiosity (Kashdan et al., 2009), sensation seeking (Hoyle, Stephenson, Palmgreen, Lorch, & Donohew, 2002) and hypomania (Eckblad & Chapman, 1986). These measures were thought to represent the positive and negative dimensions of schizotypy. In agreement with previous studies (Kwapil, Barrantes-Vidal, & Silvia, 2008), modest correlations were found between the four scales and the aforementioned measures.

The Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE; Mason, Claridge, & Jackson, 1995). The O-LIFE views schizotypy to lie on a continuum between normality and abnormality. It was designed to be primarily used within a normal population. It comprises four dimensions: unusual experiences, cognitive disorganisation, introvertive anhedonia and impulsive nonconformity, with each ranging between 24 - 30 items. Cronbach’s alpha level for the scales ranged between .77 and .89. The O-LIFE’s validity has been supported in a variety of laboratory studies ranging from neurological functioning, performance on reasoning tasks, face-processing and childhood abuse (Avons, Nunn, Chan, & Armstrong, 2003; Sellen, Oaksford, & Gray, 2005; Mason & Claridge, 1999; Startup, 1999).

The Five-Factor Measure of Schizotypal Personality Traits (FFM STPT; Edmundson, Lynam, Miller, Gore, & Widiger, 2011) groups nine scales, constructed as maladaptive variants of FFM general traits, including Social Anxiousness (the STPT variant of FFM N1: Anxiety), Social Discomfort (the STPT variant of FFM N4: Self-consciousness), Social Anhedonia (low E1: Warmth), Social Isolation and Withdrawal (low E2: Gregariousness), Physical Anhedonia (Low FFM E6: Positive emotions), Aberrant Perceptions (O1: Fantasy), Odd and Eccentric (O4: Actions), Aberrant Ideas (O5: Ideas), and Interpersonal Suspiciousness (low FFM A1: Trust).
Together, these scales are called the Five-Factor Schizotypal Inventory (FFSI). The FFSI-scales showed good psychometric properties, including support for its convergent, divergent and incremental validity beyond already existing measures.

### 4.4. Antisocial

Antisocial PD (APD) continues to be of a popular interest to researchers and clinicians of various fields. Eight measures of APD will be detailed below, however there is some inconsistency in terminology within the domain. For instance psychopathy can be referenced as a specific subtype of APD or as a synonym (Pickersgill, 2010).

The *Antisocial Personality Questionnaire* (APQ; Blackburn & Fawcet, 1999) is a 125-item self-report inventory that is designed to measure APD holistically in criminal offender populations. The measure features eight scales: *self-control, self-esteem, avoidance, paranoid suspicion, resentment, aggression, deviance and extraversion*, with Cronbach’s alpha levels ranging from .77 to .87. Validated in both a clinical and normal population, all scales were found to hold concurrent validity with the MCMI alongside predicting the age of an inmate’s first criminal offence and the length of detention.

The *Psychopathy Checklist* (PCL; Hare, 1980) has become the standard measure of psychopathy in clinical environments since it was created. When categorised as a subtype of APD, Hare defines psychopathy to be characterised by callous and unemotional traits. The PCL assesses two factors: the first focusing on the individual’s grandiose, callous and manipulative personality, and the second focusing on the individual’s deviant and impulsive life history. It has 22 items (each of which represent a facet of either factor) that form the basis of a semi-structured interview. To administer this checklist, interviewers must first take a short course to become accredited. Each item is scored on a range from 0 to 3, where higher scores represent severity. With scores ranging from 0 to 44, scores above 30 indicate a diagnosis of clinical psychopathy. Hare found that the PCL held a Cronbach’s alpha level of .90 and interrater reliability levels of .89, demonstrating good psychometric properties for a clinical interview. Hart and Hare (1989) demonstrated discriminant validity as the PCL scores were related to substance abuse.

The *Psychopathy Checklist-Revised* (PCL-R; Hare et al., 1990) is based on the PCL, however only features 20 items. It remains as a clinical interview that requires specific training to be administered, with scores over 30 indicating the presence of psychopathy. Furthermore, it still assesses the same two factors. The PCL-R removed two items that were found to hold low correlations with the total PCL score, as well as slightly modified the scoring criteria. The PCL-R has been well validated, and has superseded its precursor in popularity. Being validated in five clinical samples (N = 925), the average inter-rater reliability and Cronbach’s alpha level was .86 and .88 respectively. Salekin, Rogers, and Sewell (2006) found that PCL and PCL-R scores were significant predictors of future violence and aggression. Vitale, Smith, Brinkley and Newman (2002) found convergent validity with Eysenck’s Personality Questionnaire, in particular the psychoticism scale (Eysenck & Eysenck, 1975). The PCL-R is often favoured over the PCL due to its psychometric refinements. Furthermore, it is worth highlighting that the PCL/PCL-R measures a slightly different construct to the APQ: the APQ measures APD as defined by the DSM-IV, whereas the PCL-R measures psychopathy as a subtype of APD.

The *Levenson’s Self-Report Psychopathy Scale* (SRPS; Levenson, Kiehl, & Fitzpatrick, 1995) is a 26-item self-report questionnaire that measures *primary psychopathy* (a synonym for factor 1 of psychopathy as measured by the PCL-R) and *secondary psychopathy* (a synonym for factor 2). Unlike the PCL-R, the SRPS was designed to be used in a nonclinical population. Brinkley, Schmitt, Smith and Newman (2000) found the measure to have Cronbach alphas of .83 for primary psychopathy, .69 for secondary, and .85 for the total measure. The authors also found the SRPS to positively correlate with the PCL-R and demonstrate similar correlations to measures of substance abuse and criminal versatility which may be evidence of its poor discriminant validity. Akhtar, Ahmetoglu, and Chamorro-Premuzic (2013) demonstrated its predictive validity by using the measure to predict entrepreneurial outcomes, therefore providing support for the SRPS’ utility in nonclinical samples.

The *Psychopathic Personality Inventory Revised* (PPI-R; Lilienfield & Widows, 2005) is another self-report measure of personality traits associated with psychopathy. Featuring 157 items, the PPI-R contrasts both the PCL-R and SRPS as it measures three dimensions of psychopathy: *fearless dominance* (traits that reflect social potency, shallow affect and stress immunity), *impulsive antisociality* (traits that reflect impulsive nonconformity, externalisation of blame and Machiavellian behaviours), and *cold heartedness* (lack of empathy). The original authors found Cronbach’s alpha levels to range from .79 to .92. Uzieblo, Verschuere, Van den Bussche, and
Crombez, (2010) further found that the measure held convergent and discriminant validity with the SRPS; all three factors of the PPI-R were positively correlated with primary psychopathy, fearless dominance and impulsive anti-sociality correlated with secondary psychopathy. The authors created the PPI-R to challenge the monopoly the PCL-R had within the field and to address a variety of theoretical challenges within the field. The measure was designed to be used within both non- and clinical samples.

The Elemental Psychopathy Assessment (EPA; Lynam et al., 2011) is a 178 self-report measure that features 18 psychopathy subscales. The subscales are grouped into four unidimensional factors that are based on the FFM: antagonism, conscientiousness, extraversion and neuroticism. The authors chose this approach as they argued that psychopathy could be understood if it is related to the basic units of personality. For each of the subscales, Cronbach’s alpha levels ranged between .63 and .88. In a clinical sample, the original authors showed that the EPA correlated with three measures of psychopathy, including the PPI-R and SRPS (mean r = .81). EPA scores were also correlated with externalised behaviours such as alcohol abuse and antisocial behaviour in a prison sample. The benefit of using the EPA over other measures of psychopathy and APD is that its framework of the FFM allows other assessments that are based upon the FFM to be congruent with each other.

The Inventory of Callous-Unemotional Traits (ICU; Frick, 2004). The ICU differentiates itself from psychopathy, as it focuses on an individual’s empathy and caring behaviours, particularly in adolescents. Therefore the measure tries to capture behaviour across 24 items to help identify those most at risk. There are four scales; careless, callous, unemotional and uncaring. Participants respond via a 4-point Likert scale. Kimonis et al. (2008a, 2008b) found the total 24 items to hold a Cronbach’s alpha of .81, with the subscales ranging between .53 and .81. They also found the measure to hold concurrent validity as it was found to positively correlate with proactive and reactive aggression, delinquency and sexual offences, while negatively correlating with empathy and positive affect.

The Business-Scan 360 (B-Scan 360; Babiak & Hare, 2012) has been developed to investigate psychopathic features in business settings. This 360 degree tool is designed for managers, subordinates, and peers to assess corporate psychopathy in others. Initially, the B-Scan consisted of 113 items, to be rated on a 5-point Likert Scale, ranging from 1 (strongly Disagree) to 5 (Strongly Agree). Mahieu and colleagues (2012) collected data in two large independent samples of business personnel who rated their supervisors on the original B-Scan items and on several external variables. They identified a preliminary 20-item B-Scan scale that is consistent with the four PCL-based factors of psychopathy, being Interpersonal (superficial, grandiose, deceitful), Affective (lacks remorse, lacks empathy, doesn’t accept responsibility for actions), Lifestyle (impulsive, lacks realistic goals, irresponsible), and Antisocial (poor behavior controls, adolescent antisocial behavior, adult antisocial behavior). Because the B-Scan factors are meant to have utility in an organizational environment, these four factors were relabelled as Manipulative/Unethical, Callous/Insensitive, Unreliable/Unfocused, and Intimidating/Aggressive. In both samples, internal consistencies of the scales and total score were acceptable to excellent, ranging between .70 (Intimidating/Aggressive) to .99 (Callous/Insensitive). Initial data on the external validity seem promising as the psychopathic features measured with the B-Scan seem to be related to deviant behaviors at the workplace, including organizational retaliatory behavior, bullying, and interpersonal deviance (Mahieu et al., 2012).

4.5. Borderline

To our knowledge there are seven specific measures of borderline PD (BPD), nearly all adopt a multidimensional approach to measuring BPD. However there is some variation between the number and labels of the dimensions measured.

The Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD; Zanarini et al., 2003), features nine items where respondents answer via a 5-point Likert scale. It was found to have reasonably good levels of internal consistency and convergent validity with other BPD measures. The ZAN-BPD defines BPD as consisting of four dimensions: affective, cognitive, impulsive and interpersonal. The scale is designed to be used within a clinical setting to quickly measure changes in symptomatology over time.

The Diagnostic Interview for Borderlines (DIB-R; Zanarini, Gunderson, Frankenburg, & Chauncey, 1989) is a semi-structured interview that lasts between 50 - 90 minutes. Using 132 items, the DIB-R measures the same four dimensions as the ZAN-BPD but in considerably more depth. The DIB-R is one of the most widely used instruments for BPD diagnoses due to it being: in the public domain, possessing established psychometric qualities such
as good interrater reliabilities (Kappa levels range between .57 and .73; Zanarini, Frankenbug & Vujanovic, 2002) and longitudinal predictive validity of remission rates (Zanarini, Frankenbug, Hennen, & Silk, 2003).

The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD; Zanarini et al; 2003) is a 10-item self-report scale that is designed to complement clinical interview instruments as it identifies the potential of a possible diagnosis. Its brevity, test-retest reliability (.72), good sensitivity (correctly identified 81% of diagnoses) and specificity (correctly identified 85% of non-diagnoses) make it a useful screener.

The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD; Zanarini et al; 2003) is a 10-item self-report scale that is designed to complement clinical interview instruments as it identifies the potential of a possible diagnosis. Its brevity, test-retest reliability (.72), good sensitivity (correctly identified 81% of diagnoses) and specificity (correctly identified 85% of non-diagnoses) make it a useful screener.

The McLean Screening Instrument for Borderline Personality Disorder Beliefs Scale (BPDBS; Butler, Brown, Beck, & Grisham; 2002) differs from the aforementioned BPD instruments, as the BPDBS is influenced by cognitive behavioural therapy (Beck & Freeman, 1990) as it is constructed out of 14 items found on the Personality Belief Scale (PBS; Beck & Beck, 1991). Butler and colleagues (2002) identified that these 14 items could discriminate between BPD patients and patients with other PDs. Although not widely used, the BPDBS’s partnership with cognitive therapy offers a different approach for clinicians and researchers alike.

The McLean Screening Instrument for Borderline Personality Disorder Beliefs Scale (BPDBS; Butler, Brown, Beck, & Grisham; 2002) differs from the aforementioned BPD instruments, as the BPDBS is influenced by cognitive behavioural therapy (Beck & Freeman, 1990) as it is constructed out of 14 items found on the Personality Belief Scale (PBS; Beck & Beck, 1991). Butler and colleagues (2002) identified that these 14 items could discriminate between BPD patients and patients with other PDs. Although not widely used, the BPDBS’s partnership with cognitive therapy offers a different approach for clinicians and researchers alike.

The Borderline Personality Questionnaire (BPQ; Poreh et al., 2006) is an 80 item self-report scale. Unlike the aforementioned instruments that focus around four dimensions, the BPQ has nine. Although there is some similarity (affective instability, impulsivity and relationships), it introduces dimensions such as intense anger, suicide/self-mutilation and quasi-psychotic states. When compared with the MMPI and the SPQ, it showed significant coefficients of .48 and .45 respectively, suggesting acceptable discriminant validity. Similarly, convergent validity with the MMPI yielded a coefficient of .85.

The Borderline Evaluation of Severity Over Time (BEST; Pfohl, Blum, St. John, McCormick, Allen, & Black; 2009) is a 15 item self-report instrument that differs from other measures of BPD as it is focused on how symptomatology changes over time, rather than diagnosing the disorder. The BEST features three subscales each focusing on either: problematic thoughts and emotions that are characteristic of BPD (e.g. suicidal flirtations), problematic behaviours, or the use of positive behaviours. Each item is rated on a 5-point Likert scale. The tool measures respondents’ behaviour over the previous week, explaining its utility for therapists and clinicians. The original paper cites test-retest reliabilities ($r = .61, p < .001$), alongside face validity (due to its focus on the thoughts and feelings of the individual), and concurrent validity (it was shown to correlate significantly with the ZAN-BPD).

The Borderline Evaluation of Severity Over Time (BEST; Pfohl, Blum, St. John, McCormick, Allen, & Black; 2009) is a 15 item self-report instrument that differs from other measures of BPD as it is focused on how symptomatology changes over time, rather than diagnosing the disorder. The BEST features three subscales each focusing on either: problematic thoughts and emotions that are characteristic of BPD (e.g. suicidal flirtations), problematic behaviours, or the use of positive behaviours. Each item is rated on a 5-point Likert scale. The tool measures respondents’ behaviour over the previous week, explaining its utility for therapists and clinicians. The original paper cites test-retest reliabilities ($r = .61, p < .001$), alongside face validity (due to its focus on the thoughts and feelings of the individual), and concurrent validity (it was shown to correlate significantly with the ZAN-BPD).

The Five Factor Inventory of Histrionic Traits (FFI; Mullins-Sweatt, Edmundson, Sauer-Zavala, Lynam, Miller, & Widiger; 2012). This measure is unique in that it builds its structure around the five-factor model of personality, promoting a new multi-dimensional approach of BPD. The authors argue that by viewing BPD as comprising a constellation of maladaptive traits rather than a homogenous category, the full range of the disorder can be assessed. It assesses eight scales, each with 10 items. Items vary from five factor-based traits such as rashness, distrustfulness, and manipulativeness, to more traditional behavioural and affective dysregulation. The original paper cites internal consistency levels no less than .77 and good convergent validities with the PDQ, PAI and SNAP. It was also shown to hold incremental validity at predicting BPD over the NEO PI-R and the BPD scale of the PAI.

4.6. Histrionic

The Five Factor Measure of Histrionic Traits (FFM-HIS; Tomiatti, Gore, Lynam, Miller, & Widiger, 2012). Parallel to maladaptive variants of FFM schizotypal and borderline traits, 13 Five Factor Histrionic trait scales were constructed as maladaptive variants of FFM general traits, including Neediness for Attention and Rapidly Shifting Emotions, Intimacy Seeking, Attention Seeking, Social Butterfly, Flirtatiousness, Melodramatic-Emotionality, Romantic Fantasies, Touched Feely Suggestibility, Vanity, Disorderliness, and Impressionistic Thinking. Tomiatti and colleagues (2012) reported good psychometric qualities, adequate convergent validity for 11 out of the 13 FFHI scales and incremental validity over their respective NEO-PI-R facet scales for 12 out of 13 scales in accounting for PDQ-4 histrionic variance.

4.7. Narcissistic

Narcissistic PD (NPD) has generated a lot of interest in the past 40 years, with various instruments dedicated to identify and diagnose this PD. There are many different approaches to NPD, debating whether it is multidimensional and its place within the clinical sphere.

Murray’s Narcissism Scale (1938) is a 20-item self-report scale that measures narcissism in terms of overt,
Narcissistic vulnerability comprises of three facets: contingent self-esteem, job satisfaction demonstrating evidence of predictive validity. Hendin and Cheek (1997) found the measure to have a Cronbach’s alpha of .76 and found 10 of the items from Murray’s scale to significantly correlate with MMPI’s narcissism scale, demonstrating concurrent validity. They constructed a Hypersensitive Narcissism Scale (HSNS) that measures the covert aspect of narcissism. This was shown to have a Cronbach’s alpha of .82 and correlated with the five-factor model of personality, MMPI and the exploitativeness/entitlement dimension of the NPI-40. Priftira and Ryan (1984) found the NPI to be strongly correlated with the MCMI narcissism scale within a clinical sample.

Raskin and Terry (1988) further developed the NPI, producing the NPI-40, which has become the most popular measure to assess NPD. The NPI-40 features just 40 items with a Guttman alpha of internal consistency statistic of .83. Most interesting about this version of the NPI is that seven dimensions were also identified: authority, exhibitionism, superiority, vanity, exploitativeness, entitlement, and self-sufficiency—all of which were found to have internal consistency levels no lower than .50. The total scale and its dimensions were found to correlate with various trait rankings on self-confidence, physical attractiveness, pleasure seeking and assertiveness as measured by other instruments included in the Institute of Personality Assessment and Research (IPAR) battery.

Arms, Rose and Anderson (2006) reduced the NPI-40 to a 16-item, uni-dimensional measure. The NPI-16 incorporates the seven factors found within the NPI-40, however it produces a single score that simply represents how narcissistic the individual is. The NPI-16 was found to have a Cronbach’s alpha level of .72, which is satisfactory as the authors found the NPI-40 to have a level of .84. Both the NPI-40 and NPI-16 correlated with self-ratings of attractiveness, competence, and big five measures, suggesting evidence of predictive validity.

Margolis-Thomas Measure of Narcissism (M-T; Margolis & Thomas, 1980). The M-T, an unpublished master’s thesis, differs from the NPI as it is able to make clinical diagnoses. The scale features 60 items, and measures the six dimensions of NPD defined by the DSM-III. Each item is a paired statement, with one being narcissistic and the other not, and respondents choose the statement they believe is most true to themselves. The original authors cite an internal consistency coefficient of .84, with the measure successfully differentiating between adolescents with and without NPD, suggesting concurrent validity. Mullins and Kopelman (1988) created a short version that consisted of only 24 paired samples, as they wanted a smaller, more efficient battery. This scale had an internal consistency level of .69. This short version was inversely related to life, self-, family and job satisfaction demonstrating evidence of predictive validity.

The Pathological Narcissism Inventory (PNI; Pincus, Ansell, Pimentel, Cain, Wright, & Levy, 2009). The PNI differs from various other narcissism measures as it aims to distinguish between pathological/clinical and normal narcissism. Comprising of 52 items, it assesses two overarching dimensions: narcissistic grandiosity and narcissistic vulnerability. Narcissistic grandiosity features four pathological facets that are not dissimilar to those found in the NPI-40: entitlement, rage, exploitativeness, grandiose fantasy, and self-sacrificing self-enhancement. Narcissistic vulnerability comprises of three facets: contingent self-esteem, hiding the self, and devaluing. Therefore the PNI is similar to Murray’s original conception of narcissism that consists of overt and covert behaviours. The PNI was found to have a Cronbach’s Alpha of .95, with all the facets scoring between .75 and .93. Concurrent validity was found as PNI scores held significant correlations with NPI and HSNS scores (this is true for both total and facet scores). PNI scores were also found to predict empathy, aggression and low moral values, as well as psychiatric variables such as number of suicide attempts, no-shows to therapy sessions, and whether psychiatric medication is being taken.

Five Factor Narcissism Inventory (FFNI; Glover et al., 2012). Like its sister inventories, the FFNI groups 15
narcissism trait scales that are constructed as maladaptive and narcissistic variants of FFM facets, including Reactive Anger, Shame, Indifference, Need for Admiration, Exhibitionism, Thrill-Seeking (also represented in the EPA; Lynam et al., 2011), Authoritativeness, Grandiose Fantasies, Cynicism/Distrust (also represented in the EPA; Lynam et al., 2011), Manipulativeness, Exploitativeness, Entitlement, Arrogance, Lack of Empathy, and Acclaim-Seeking. Good to excellent internal consistencies were reported, including adequate convergent validity for 14 out of 15 FFNI scales. It was concluded that the 15 different scales provided a comprehensive and multifaceted description of narcissistic pathology.

There are also other less well known measures of NPD (Ashby, Lee, & Duke, 1979; Richman & Flahery, 1990).

4.8. Avoidant

No specific measure of avoidant personality could be found, except for the Five-Factor Avoidant Assessment (FFAvA; Lynam, Loehr, Miller, & Widiger, 2012), proposing 10 maladaptive avoidant variants of FFM traits, including scales assessing Evaluation Apprehension, Despair, Mortified, Overcome, Social Dread, Shrinking, Risk Averse, Joyless, Rigidity, and Timorous. Initial validation results are promising, showing homogenous and reliable scales, that are strongly related to their respective NEO-PI-R facets, and with all FFAvA scales, except Timorous, showing expected relationships with AVD measures enclosed in comprehensive PD inventories.

4.9. Dependent

The DSM-5 characterises DPD as displaying maladaptive clinging behaviour towards others for care, advice and support. Our review identified four instruments that specifically measured DPD.

The Minnesota Multiphasic Personality Inventory-2 Social Introversion Subscales (MMPI-2 Si1, 2, 3; Ben-Porath, Hostetler, Butcher, & Graham, 1989). The instrument consists of 38 items and is based on the subscales of MMPI-2. Using item-level factor-analysis, Ben-Porath and colleagues produced three exclusive subscales that are designed to collectively measure DPD, in particular the avoidance of being alone. The three scales, Shyness/Self Consciousness, Social Avoidance, and Self/Other Alienation, were found to hold acceptable internal consistency coefficients (ranging between .75 and .82). The authors found that two subscales predicted 80% of the variance in a study on social introversion. Further ad hoc validity for the role of shyness comes from Lorant, Henderson, and Zimbardo (2000) who found that in a clinical sample (N = 107), 60 participants that were found to be shy, also had a personality disorder, with DPD being the most common. Unfortunately there have not been many studies to further validate the three subscales. It could even be criticised in that the scales are not directly measuring DPD—rather a facet, or another comorbid construct altogether.

The Dependent Personality Questionnaire (DPQ; Tyrer, Morgan, & Cicchetti, 2004) is an eight item self-report questionnaire. The instrument is intended to be used as a screening tool to identify patients that potentially have DPD. Participants rate themselves using a 4-point Likert scale that ranges from 0 to 3. Although the original authors do not cite any statistics on the instrument’s reliability, they demonstrate that the DPQ holds good diagnostic validity. The DPQ held an overall diagnostic accuracy of 87.5% in psychiatric patients diagnosed with DPD according to the ICD-10 version of the Personality Assessment Schedule (Tyrer, 2000). The DPQ’s diagnostic sensitivity, specificity, predicted positive and negative accuracies were all 87.5%. Compared to matched controls, patients diagnosed with the disorder had a mean score of 13 (controls comparably scored a mean of 7).

The Dependent Personality Inventory (DPI; Huber, 2005). The DPI is a 55-item questionnaire that measures seven independent factors representing various symptoms of DPD as defined by the DSM-IV, including: difficulty making decisions, assuming responsibility, difficulty expressing disagreement, difficulty initiating projects, seeking support from others and feeling helpless and alone. The original paper found the DPI to have a high internal consistency with a Cronbach’s alpha of .90.

The Five-Factor Measure of Dependent Traits (FFM DPT; Gore et al., 2012). Gore and colleagues created 12 scales to assess FFM dependent traits, including Separation Insecurity, Pessimism, Shamefulness, Helplessness, Intimacy Needs, Unassertiveness, Gullibility, Selflessness, Subservience, Self-effacing, Ineptitude, and Negligence. Internal consistencies of the scales were good, and scales correlated with their NEO-PI-R equivalents, also demonstrating discriminant validity towards different NEO-traits, and incremental validity beyond their corresponding NEO PI-R facets to explain variance in the SNAP DPD scale.
4.10. Obsessive-Compulsive

Obsessive-compulsive PD (OCPD) is rather similar to the Axis 1 obsessive-compulsive disorder (OCD), and therefore they are often confused. Indeed it has been suggested that the evidence for the construct validity of these two scales was mixed (Phillips et al., 2010). However, where OCD can be defined as a mental disorder characterized by ego-dystonic, intrusive and time-consuming obsessions and compulsions, OCPD rather reflects an ego-syntonic personality style including stable traits such as perfectionism and rigidity. Assessing these disorders is difficult due to their comorbidity with (each other and) other disorders as well as their heterogeneity.

The Five Factor Obsessive-Compulsive Inventory (FFOCI; Samuel, Riddell, Lynam, Miller, & Widiger, 2012) is a self-report measure that maps OCPD onto the FFM. The 12 dimensions of the instrument each represent a maladaptive, lower-order version of a big five trait. The scale’s Cronbach’s alpha levels ranged between .77 and .87. The scales were shown to hold convergent validity with the MCMI-III (r = .58, p < .01), PDQ-4 (r = .50, p < .01) and the SNAP-2 (r = .66, p < .01). The instrument was also found to hold incremental validity over the MCMI-III.

5. Discussion

This paper has been a PD assessment “housekeeping and auditing” exercise, targeting comprehensiveness rather than reflecting usage frequency or the publication track-record of instruments. Such criteria probably more reflect whether a measure has been included in one of the four major longitudinal studies on PDs resulting in an increased number of publications, instead of reflecting their quality status. We made the choice to go back as far as DSM-III covering old and new measures. We also did not do a “quality control” in terms of journal ratings as these vary over time and have been criticised.

This review has made a number of things clear: (a) there are multiple options to assess PDs, both comprehensively and specifically, (b) objectives of users may largely vary, going from general screening towards differentiated (clinical) diagnostics, (c) methods vary in terms of in-depth assessment and necessary time, with structured interviews generally preferred for clinical assessment over self-report inventories, (d) there exist a number of “fore wash” assessment methods, that may be followed by more in-depth assessment when appropriate, (e) also the contexts of the assessment may vary, ranging from clinical to more personal developmental and occupational related questions, and finally (f) methods may differ in terms of whether they assess primarily pathological trait variance or also tap into general traits.

To facilitate choosing among measures, a decision tree was developed distinguishing three major entries, i.e. research, clinical assessment and assessment with a developmental purpose (Figure 1). Researchers can choose between categorically-based versus more dimensionally-based PD measures usually relying on individuals’ self-reports. A similar dichotomy is available for clinical diagnostics and decision making, with the same dimensional measures available like for research purposes. Categorically-based clinical assessment may first involve a pre-screening followed by more specific PD assessment focussing on single PDs. Our review has made clear that there are various options for the majority of the PDs. A broader clinical assessment procedure may also include a comprehensive PD assessment, usually done via structured interviews, assessing both nature and severity of personality pathology in terms of DSM-5 categorical constructs. Finally, an increasing number of psychologists are interested to use personality pathology assessment instruments and methods to identify an individual’s personal and professional needs, in order to help develop and mould the sharp sides of their personalities. These are often related to subclinical forms of personality pathology, also called aberrant personality tendencies (Wille et al., 2013; De Fruyt et al., 2013ab). Given the idea of a spectrum or continuum between dimensions of general and maladaptive personality traits (De Bolle et al., 2012) and that the US Legislation (Americans with Disabilities Act, ADA, 1990) prohibits that clinical measures are used in personnel selection and development assessments, FFM general trait based measures of personality dysfunction seem to be most useful. In addition, also a number of DSM-IV based personality measures have been developed to understand personality functioning at work, contextualizing the item content and/or instructions with a work-frame. Also for these purposes, different options are available and increasingly used.

The purpose of the present review was not only to list the available PD measures, but also to illustrate the
increased attention and use of such assessment devices for a broader range of purposes including clinical and non-clinical assessment and research. This expanded scope is a direct consequence of advances in the conceptualization of PDs from distinct categorical entities to the consideration of personality pathology dimensions that are more quantitatively than qualitatively different from normal trait variation. Despite the many suggestions to replace the categorical PDs by a dimensional system including more specific traits, DSM-5 continues the categorical conceptualisation of PDs making our review of assessment methods still timely. We hope that professionals and researchers will find our review helpful in choosing an appropriate assessment method for their purposes, but will consider at the same time also the trait based system of personality pathology described in DSM-5 Section 3.

References


A. Furnham et al.


Coolidge, F. L. (2001). Personal Communication. Questionnaire and Scoring Key for the Short Coolidge Axis-II Inventory. (Provided by the Author)


Jones, S. H., Burrell-Hodgson, G., & Tate, G. (2007). Relationships between the Personality Beliefs Questionnaire and


Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.