Clinical Validity of Prototype Personality Disorder Ratings in Adolescents

JARED A. DEFFIE,1 GREG HAGGERTY,2 SCOTT W. SMITH,3 LUIS BETANCOURT,2 ZAIN AHMED,2 AND KEITH DITKOWSKY2

1Department of Psychology, Emory University
2Von Tauber Institute for Global Psychiatry, Nassau University Medical Center
3Pathways to Housing, New York, New York

A growing body of research shows that personality pathology in adolescents is clinically distinctive and frequently stable into adulthood. A reliable and useful method for rating personality pathology in adolescent patients has the potential to enhance conceptualization, dissemination, and treatment effectiveness. The aim of this study is to examine the clinical validity of a prototype matching approach (derived from the Shedler Westen Assessment Procedure–Adolescent Version) for quantifying personality pathology in an adolescent inpatient sample. Sixty-six adolescent inpatients and their parents or legal guardians completed forms of the Child Behavior Checklist (CBCL) assessing emotional and behavioral problems. Clinical criterion variables including suicide history, substance use, and fights with peers were also assessed. Patients’ individual and group therapists on the inpatient unit completed personality prototype ratings. Prototype diagnoses demonstrated substantial reliability (median intraclass correlation coefficient = .75) across independent ratings from individual and group therapists. Personality prototype ratings correlated with the CBCL scales and clinical criterion variables in anticipated and meaningful ways. As seen in prior research with adult samples, prototype personality ratings show clinical validity across independent clinician raters previously unfamiliar with the approach, and they are meaningfully related to clinical symptoms, behavioral problems, and adaptive functioning.

Received August 3, 2013; Revised September 10, 2014.

Luis Betancourt is now with Novant Health Psychiatric Associates in Charlotte, NC; Zain Ahmed is now a research assistant based in Manhattan, NY; Keith Ditkowsky is now at Bronx Children’s Psychiatric Center/Albert Einstein College of Medicine in the Bronx, NY.

Address correspondence to Greg Haggerty, Von Tauber Institute for Global Psychiatry, Nassau University Medical Center, 2201 Hempstead Turnpike, Box 48, East Meadow, NY 11554; Email: ghag541@aol.com

Lewinsohn, Solomon, Seeley, & Zeiss, 2000; Marshall et al., 2001). Dimensional diagnostic approaches are currently being considered for the International Classification of Diseases (11th ed. [ICD–11]; Maj, 2011; Reed, 2010) and received widespread support (Kraemer, 2007; Reed, Correia, Esparza, Saxena, & Maj, 2011; Shedler et al., 2010; Spitzer, First, Shedler, Westen, & Skodol, 2008).

Westen, Shedler, and colleagues proposed a prototype rating approach designed to maximize rating accuracy while taking into consideration the cognitive characteristics of human assessors (Ortigo et al., 2010; Shedler & Westen, 2004; Westen et al., 2002; Westen & Shedler, 2000; Westen et al., 2006; Westen, Shedler, Bradley, & Defife, 2012). Cognitive research science on classification processes indicates that human thinking naturally relies on forms of cognitive prototype matching (Cantor & Genero, 1986; Horowitz, Post, de Sales French, Wallis, & Siegelman, 1981; Horowitz, Wright, Lowenstein, & Parad, 1981; Kim & Ahn, 2002; Rosch & Mervis, 1975). Rather than memorize symptom lists with arbitrary and variable cutoffs across disorders, diagnosticians can form mental representations of coherent syndromes, in which signs and symptoms can be linked by meaningful functional relations (Ahn, 1999). The prototype-matching method preserves a configurational or syndromal approach to personality classification (Blashfield, 1985; Horowitz, Post, et al., 1981; Pilkonis, 1988; Rounsaville et al., 2002) while allowing dimensional assessment on a scale from 1 (no match) through 5 (very good match). The method parallels diagnosis in many areas of medicine, where variables such as blood pressure are measured on a continuum, but physicians refer to certain ranges as “borderline” or “high.” Additionally, clinical utility findings suggest that many clinicians might prefer a prototype rating approach over current diagnostic methods. As illustrated in a recent survey
sponsored by the World Health Organization, psychiatrists prefer a diagnostic method that has many of the features associated with prototype diagnosis (Reed et al., 2011). With respect to PDs, multiple independent studies of clinicians rated prototype diagnosis substantially more useful, comprehensive, and clinically efficient than DSM–IV diagnosis or other dimensional alternatives (Rottman, Ahn, Sanislow, & Kim, 2009; Spitzer et al., 2008; Westen et al., 2006). More research has begun to focus on the reliability and validity of prototype diagnosis. Reliability of prototype personality ratings by independent observers of the same clinical material is high, with median interrater reliability across PDs of $r = .72$ (Westen, Defife, Bradley, & Hilsenroth, 2010). Across a number of studies, prototype ratings correlate highly with, and have similar correlates to, both categorical and dimensional diagnoses obtained by summing DSM–IV criteria or using self-report measures of specific syndromes (Defife et al., 2010; Dutra, Campbell, & Westen, 2004; Westen, Muderrisoglu, Fowler, Shedler, & Koren, 1997). All demographic information recorded on the CDF was gathered from a combination of the individual therapist, unit social worker, attending psychiatrist, and any collateral sources (i.e., parent or legal guardian, school psychologist, teachers, case worker) so that demographic information was based on all available reliable data.

**METHOD**

Newly admitted patients to Nassau University Medical Center’s adolescent inpatient unit and their legal guardians were approached by the staff for consent to participate in an assessment and outcome study. Patient and parent-report clinical measures were collected at admission. Patients on the unit were assigned to an individual and group psychotherapist in an ecologically valid manner based on clinician availability and caseload. Individual and group psychotherapy each consisted of meetings three times weekly. Individual and group psychotherapists independently completed clinical ratings on patient discharge.

**Participants**

Participants ($n = 66$) were adolescent inpatients seen at Nassau University Medical Center between August 2010 and June 2011 (43% of the total patients admitted to the unit). Thirty-two female and 34 male patients consented to participate the study with an average age of 15.8 years ($SD = 1.17$). The ethnicity of the sample was 61% White, 32% African American, 25% Hispanic, 2% Asian, and 5% other. A range of socioeconomic status was represented, with 49% identified as working class, 39% middle class, 9% upper middle class, and 4% poor. The socioeconomic status was reported by the individual using all available information known about the patient and his or her background. The average length of stay on the unit was 11 days ($SD = 5.14$). The primary diagnosis for these 66 patients was as follows: 62% mood disorders, 32% conduct disorder/oppositional defiant disorder, 3% posttraumatic stress disorder, 2% psychosis, and 1% impulse control disorder. We found no significant differences in age, gender, diagnosis, or scores on admission measure between those who consented to the study versus those who did not.

Participating clinicians included a licensed clinical psychologist with more than 5 years of experience and advanced clinical psychology doctoral students (interns and externs) who had completed advanced course work in assessment at a clinical psychology PhD program accredited by the American Psychological Association.

**Measures**

**Clinical Data Form.** The Clinical Data Form (CDF; Westen & Shedler, 1999) is a clinician report form that gathers information on demographic, diagnostic, etiological, and adaptive functioning variables. Data collected with the form concerning developmental history and life events shows strong agreement (cross-method validity) with data collected from patients ($r = .74–.96$; Defife, Drill, Nakash, & Westen, 2010). Adaptive functioning variables assessed via the CDF (e.g., Global Assessment of Functioning scores) have likewise shown high reliability and validity vis-à-vis ratings by independent observers (Defife et al., 2010; Dutra, Campbell, & Westen, 2004; Westen, Muderrisoglu, Fowler, Shedler, & Koren, 1997). All demographic information recorded on the CDF was gathered from a combination of the individual therapist, unit social worker, attending psychiatrist, and any collateral sources (i.e., parent or legal guardian, school psychologist, teachers, case worker) so that demographic information was based on all available reliable data.

**TABLE 1.—Individual and group therapist agreement on SWAP–A prototype personality ratings.**

<table>
<thead>
<tr>
<th>SWAP–A Prototype</th>
<th>ICC (1,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antisocial/psychopathic</td>
<td>.77</td>
</tr>
<tr>
<td>Emotionally regulated</td>
<td>.55</td>
</tr>
<tr>
<td>Avoidant/constricted</td>
<td>.77</td>
</tr>
<tr>
<td>Narcissistic</td>
<td>.84</td>
</tr>
<tr>
<td>Histrionic</td>
<td>.68</td>
</tr>
<tr>
<td>Inhibited/self-critical style</td>
<td>.73</td>
</tr>
</tbody>
</table>

*Note. SWAP–A = Shedler–Westen Assessment Procedure–Adolescent. n = 61.*
Inhibited/self-critical Histrionic Narcissistic Avoidant/constricted

prototypes.

Shedler–Westen Assessment Procedure–Adolescent–Prototypes. Shedler and colleagues empirically derived adolescent PD prototypes known as the Shedler–Westen Assessment Procedure–Adolescent–Prototypes (SWAP–A Prototypes; Westen, et al., 2003) from a large national sample of clinicians (n = 296) treating child and adolescent patients in their practice. The clinicians used a 200-item Q-sort instrument to describe an adolescent patient in their treatment for "enduring maladaptive patterns of thought, feeling, motivation, or behavior—that is, personality." They applied Q-factor analysis to identify empirically distinct diagnostic clusters and generated five PD types (i.e., Antisocial/Psychopathic, Emotionally Dysregulated, Avoidant/Constricted, Narcissistic, and Histrionic) and a higher functioning personality style (i.e., Inhibited/Self-Critical). For each cluster, a paragraph-long prototype was constructed using the most empirically descriptive items (e.g., see Figure 1). In this study, individual and group therapy clinicians rated the prototypes on a 5-point scale ranging from 1 (little or no match) between the patient and the prototype (description does not apply) to 5 (a very good match; patient exemplifies the description; a prototypical case).

RESULTS

We examined the level of agreement of prototype ratings between the adolescent patient’s individual and group therapists. Because two sets of prototype ratings can be perfectly correlated without showing actual diagnostic agreement (e.g., prototype rating scores of 2, 4, 2, 4 are perfectly correlated with another rater’s scores of 1, 5, 1, 5, but show no actual diagnostic agreement), we calculated one-way random average measures intraclass correlation coefficients [ICC(1,2)] for the ratings on each prototype scale. ICCs of <.40 are considered poor, .40 to .59 are fair, .60 to .73 are good, and .74 is considered excellent (Fleiss, 1981). Even though this study uses a cross-method and not strictly interrater reliability measurement design (i.e., rating scores were not based on two observations of the same clinical material, but on independent clinical observations of behavior in two different treatment contexts: individual vs. group therapy), the ICCs (see in Table 1) were in the fair to excellent range for interrater reliability (Fleiss, 1986), with a median ICC at ICC = .75. There was also an exceptional degree of rating consistency across raters, with 84% to 97% of prototype scale ratings falling within a 1-point margin across the two raters. Table 2 provides the means and standard deviations of prototype personality ratings (averaged across individual and group therapist) and a prevalence percentage of adolescents meeting or exceeding the diagnostic “caseness” rating (i.e., a prototype rating score of 4 or greater). As might be expected on an adolescent acute inpatient unit, emotionally dysregulated PD was the most represented prototype rating, followed by antisocial/psychopathic PD, whereas the higher functioning inhibited/self-critical style was least prevalent.

We report correlations between personality prototype ratings and YSR and parent-report form summary scale variables in Table 3. The clinician prototype ratings show significant associations in distinct patterns across both self- and parent-reported behavioral problems. Avoidant/constricted and the higher functioning variant, inhibited/self-critical style, were both inversely related to externalizing behavior problems assessed from either the patient’s or parent’s perspective. Narcissistic PD prototype ratings were significantly correlated with parent-, but not self-reported externalizing behavior problems. Antisocial PD was inversely related to internalizing behaviors via self-report and strongly positively associated with externalizing behavior problems on parent reports. Conversely, parents of emotionally dysregulated patients reported significantly fewer externalizing problems, but the patients themselves reported a significantly greater degree of internalizing problems.

We also investigated the relationship between the prototype ratings and the subscales of the CBCL and YSR (Tables 4 and
5, respectively). Ratings on antisocial/psychopathic PD were positively correlated with parent-rated Rule Breaking and Aggressive Behavior scales, whereas self-reports did not show this correlation. Emotionally dysregulated PD ratings were positively related to the YSR Anxious/Depressed scale. Avoidant/constricted PD ratings were inversely related to parent-rated CBCL Rule Breaking and Aggressive Behavior scales and the self-reported YSR Aggressive Behavior scale. Narcissistic PD ratings were positively related to parent-rated CBCL Rule Breaking and Aggressive Behavior scales, but not on self-reported YSR. Inhibited/self-critical prototype ratings were positively correlated with the self-reported Anxious/Depressed YSR scale and negatively correlated on both the self-reported YSR and parent-rated CBCL Rule Breaking and Aggressive Behavior scales.

Finally, Table 6 reports correlations between prototype personality ratings and clinical reports of pathological behavior problems recorded via the CDF. Patients with higher antisocial/psychopathic PD ratings were significantly more likely to have school discipline problems, physical fights with peers, and substance abuse histories. Emotionally dysregulated PD ratings were significantly correlated with a prior suicide attempt, and narcissistic PD patients were more likely to have school discipline problems. Avoidant/constricted PD ratings were inversely related to substance abuse problems, and patients with a more inhibited/self-critical style were less likely than other patients on the unit to have maladaptive problems with school discipline, physical fights, or substance abuse.

**DISCUSSION**

These results provide evidence for the clinical validity of prototype PD ratings in clinical practice with adolescents. Even with a novel rating approach previously unfamiliar to the clinicians in this sample, prototype personality ratings showed substantial agreement across independent clinical raters. Whereas field trials and interrater interview studies comparing diagnoses made by different structured interviews and questionnaires for the same PDs administered days or a few weeks apart have shown low correlations and even lower kappa coefficients indicating convergent diagnoses (Clark, Livesley, & Morey, 1997; Pilkonis et al., 1995; Skodol, Oldham, Rosnick, Kellman, & Hyler, 1991), we found high ICCs (median ICC = .75) between two assessments made by independent assessors from naturalistic inpatient practice. Although the study design focused on clinical utility and validity (i.e., ratings were made based on material from independent clinical data sources), these results extend prior work on the interrater reliability of prototype personality ratings made from the same videotaped clinical hours (Westen et al., 2010). Furthermore, prototype ratings showed significant associations with clinical problems rated across informants, including self-, parent, and clinical report data, adding to increasing evidence of the validity, reliability, and clinical utility of prototype rating in research on PDs (Rottman et al., 2009; Skodol et al., 2011; Westen et al., 2006; Westen et al., 2012).

We found some noteworthy results because we used self-report and parent-reported data. The lack of awareness of one’s behavioral impact on others characteristic of narcissistic PD was possibly reflected in the finding that the narcissistic PD prototype rating was significantly correlated with parent- but not self-reported externalizing behavior problems. Antisocial/psychopathic and emotionally dysregulated PD ratings showed interesting divergent correlations between self- and parent-report CBCL forms. Perhaps the complicated self- other interpersonal disturbances prevalent in both these forms

**Table 4.—Correlations between the mean SWAP–A prototype scores and the Child Behavior Checklist scales.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Antisocial/psychopathic</td>
<td>-.34*</td>
<td>-.10</td>
<td>-.28</td>
<td>.14</td>
<td>.12</td>
<td>.31</td>
<td>.55**</td>
<td>.43**</td>
</tr>
<tr>
<td>Emotionally dysregulated</td>
<td>.10</td>
<td>.11</td>
<td>.09</td>
<td>-.28</td>
<td>-.32</td>
<td>-.48**</td>
<td>-.54**</td>
<td>-.53**</td>
</tr>
<tr>
<td>Avoidant/constricted</td>
<td>-.03</td>
<td>.14</td>
<td>-.07</td>
<td>.00</td>
<td>-.31</td>
<td>-.27</td>
<td>-.44**</td>
<td>-.46**</td>
</tr>
<tr>
<td>Narcissistic</td>
<td>.00</td>
<td>-.25</td>
<td>-.01</td>
<td>-.05</td>
<td>.13</td>
<td>.25</td>
<td>.47**</td>
<td>.33**</td>
</tr>
<tr>
<td>Histrionic</td>
<td>.04</td>
<td>-.27</td>
<td>.00</td>
<td>-.26</td>
<td>.05</td>
<td>-.17</td>
<td>-.03</td>
<td>-.17</td>
</tr>
<tr>
<td>Inhibited/self-critical</td>
<td>.20</td>
<td>.03</td>
<td>.13</td>
<td>-.04</td>
<td>-.15</td>
<td>-.38*</td>
<td>-.50**</td>
<td>-.48**</td>
</tr>
</tbody>
</table>

*Note.* All values in bold are statistically significant to the $p < .05$ level. SWAP–A = Shedler–Westen Assessment Procedure–Adolescent. $n = 35$. **p < .05. ***p < .01.

**Table 5.—Correlations between the mean SWAP–A prototype scores and the Youth Self-Report scales.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Antisocial/psychopathic</td>
<td>-.37**</td>
<td>-.32*</td>
<td>-.25</td>
<td>-.13</td>
<td>-.16</td>
<td>-.11</td>
<td>-.25</td>
<td>.06</td>
</tr>
<tr>
<td>Emotionally dysregulated</td>
<td>.37**</td>
<td>.20</td>
<td>.14</td>
<td>.07</td>
<td>.09</td>
<td>.06</td>
<td>-.09*</td>
<td>-.04</td>
</tr>
<tr>
<td>Avoidant/constricted</td>
<td>.04</td>
<td>.09</td>
<td>-.12</td>
<td>.09</td>
<td>-.12</td>
<td>.08</td>
<td>.27</td>
<td>-.35**</td>
</tr>
<tr>
<td>Narcissistic</td>
<td>-.18</td>
<td>-.11</td>
<td>.11</td>
<td>-.24</td>
<td>.02</td>
<td>.06</td>
<td>.22</td>
<td>-.02</td>
</tr>
<tr>
<td>Histrionic</td>
<td>.02</td>
<td>-.03</td>
<td>.04</td>
<td>-.17</td>
<td>-.01</td>
<td>-.22</td>
<td>-.10</td>
<td>-.09</td>
</tr>
<tr>
<td>Inhibited/self-critical</td>
<td>.29*</td>
<td>.24</td>
<td>.06</td>
<td>.02</td>
<td>.07</td>
<td>-.03</td>
<td>-.43**</td>
<td>-.31*</td>
</tr>
</tbody>
</table>

*Note.* All values in bold are statistically significant to the $p < .05$ level. SWAP–A = Shedler–Westen Assessment Procedure–Adolescent. $n = 52$. **p < .05. ***p < .01.
of personality pathology are highlighted in these findings, with antisocial PD inversely related to internalizing behaviors via self-report and strongly positively associated with externalizing behavior problems on parent reports.

Twenty-seven percent of patients in our sample scored a 4 or 5 on at least one PD prototype. The prevalence of PDs in adolescent clinical samples is understudied (Shiner & Tackett, 2014). Only two studies (Feenstra, Busschbach, Verheul, & Hutsebaut, 2011; Grilo et al., 1998) to date have investigated the prevalence rates of PDs with these populations and found that between 41% and 64% of patients in a clinical setting have diagnosed PDs. One explanation for the differences in prevalence rates between our sample and the other two studies might be that our sample did not contain as many patients demonstrating personality pathology. Our study contains a smaller sample than the two other mentioned studies and, as such, we could have conducted the study during a period in which fewer patients admitted were experiencing a diagnosable PD. Our sample might not be indicative of the overall personality pathology experienced on the inpatient unit. Future research will need to be done with larger samples taken over a longer period of time to assess this. Another possibility is that prototypes yield appropriately conservative PD diagnoses. Past research on prevalence used DSM criteria for PD diagnosis. The prototypes used here and the DSM criteria for PD diagnosis were constructed differently. The method in which the prototypes were constructed was based on data on naturally occurring clusters. The categories used in the DSM are based on clinical convention derived from clinicians’ experiences with patients.

Clinicians are reluctant to implement the existing PD diagnostic system with its laundry list of symptoms, cumbersome algorithms, overlapping criteria, and descriptive vagaries. Using a prototype system, clinical assessors could briefly and efficiently generate a classification rating that indicates for each syndrome both the extent to which the patient resembles the prototype and whether the patient matches the prototype strongly enough to receive a categorical classification.

Our study is a first step in many ways. This is the first time, to our knowledge, that the prototype diagnoses have been evaluated by using them in the way in which they would be potentially used in everyday clinical work (i.e., based on information from clinical contact and not structured interviews). Given the small sample, this study is by no means a definitive statement on the utility and applicability of dimensional personality prototype ratings. However, our findings do highlight the cross-observer and concurrent validity of personality prototype ratings, adding to a growing body of work supporting alternative approaches to personality classification for research and applied practice. Future research on this prototype approach will require larger samples. Additionally, future studies would do well to compare prototype ratings results from structured interviews with ratings done by clinicians who have been treating the patients. Because our sample was based on an average length of stay of 11 days, studies would need to evaluate this prototype matching approach in outpatient samples in which clinicians will know the patient for longer periods. That being said, our results are encouraging because even with somewhat limited exposure, clinicians showed good reliability with the prototype ratings.

ACKNOWLEDGMENT
An earlier version of this article was presented at the annual meeting of the Society for Personality Assessment on March 17, 2012.

FUNDING
Preparation of this article was supported by grants from the National Institute of Mental Health R01-MH78100 and the American Psychoanalytic Association’s Fund for Research. The authors have no financial conflicts of interest to disclose.

REFERENCES


