Can Personality Disorder Experts Recognize DSM-IV Personality Disorders From Five-Factor Model Descriptions of Patient Cases?

Benjamin M. Rottman, MS; Nancy S. Kim, PhD; Woo-kyoung Ahn, PhD; and Charles A. Sanislow, PhD

Background: Dimensional models of personality are under consideration for integration into the next Diagnostic and Statistical Manual of Mental Disorders (DSM-5), but the clinical utility of such models is unclear.

Objective: To test the ability of clinical researchers who specialize in personality disorders to diagnose personality disorders using dimensional assessments and to compare those researchers' ratings of clinical utility for a dimensional system versus for the DSM-IV.

Method: A sample of 73 researchers who had each published at least 3 (median = 15) articles on personality disorders participated between December 2008 and January 2009. The Five-Factor Model (FFM), one of the most-studied dimensional models to date, was compared to the DSM-IV. Participants provided diagnoses for case profiles in DSM-IV and FFM formats and then rated the DSM-IV and FFM on 6 aspects of clinical utility.

Results: Overall, participants had difficulty identifying correct diagnoses from FFM profiles (t51 = 12.36, P < .01), and the same held true for a subset reporting equal familiarity with the DSM-IV and FFM (t51 = 6.96, P < .01). Participants rated the FFM as less clinically useful than the DSM for making diagnoses, devising treatment plans, and communicating with professionals (all t69 > 2.19, all P < .05), but more useful for communicating with patients (t69 = 3.03, P < .01).

Conclusions: The results suggest that personality disorder expertise and familiarity with the FFM are insufficient to correctly diagnose personality disorders using FFM profiles. Because of ambiguity inherent in FFM profile descriptors, this insufficiency may prove unlikely to be attenuated with increased clinical familiarity with the FFM.

Five-Factor Model: A Trait-Based Dimensional System

The trait-based dimensional proposal for personality disorders that has received the most attention is the Five Factor Model (FFM) of personality (eg, Costa and McCrae's Revised NEO Personality Inventory; see Clark for a recent review), and it is therefore the example we chose to examine in the current research. Note, however, that our broader intent in comparing the DSM-IV and the FFM is to provide new information about the clinical utility of trait-based dimensional models in general, as will be discussed later.

As diagnostic/assessment systems for personality disorders, the current DSM-IV and FFM each have distinct benefits and disadvantages. The DSM-IV classifies maladaptive personality into 10 personality disorders, each defined by unique criteria. For example, to be diagnosed with antisocial personality disorder, one must have, pervasively and across contexts, at least 3 of the 7 symptoms shown in Figure 1A. This approach has an important advantage in terms of cognitive processing; using discrete categories is cognitively efficient. Instead of describing or remembering all the features and characteristics of each person, one can simply describe or remember a person as having antisocial personality disorder.

However, compared to the FFM and other dimensional systems, there are certain disadvantages to the DSM's categorical assessment, and there exist many useful reviews on this topic. These include diagnostic comorbidities that may be due to criterion overlap, arbitrary diagnostic thresholds of the number of criteria necessary to count as having a disorder, and clinical heterogeneity among people with the same diagnosis. These problems have led some to argue that DSM-IV disorder categories are neither discrete nor well defined. Some critics have argued that the DSM-IV personality disorders do not cover all important personality problems; yet, adding additional personality disorders could exacerbate the comorbidity problem. In sum, dissatisfaction with the current diagnostic system has generally been on the rise.

Two general approaches to dimensionalizing Axis II disorders have been proposed. One is to preserve personality disorder types (eg, borderline personality disorder) and assess how close a person is to a given type (type-based dimensional system). The other approach departs further from the DSM-IV by profiling a person along underlying traits, such as introversion (trait-based dimensional system). The current study examines the clinical utility of a trait-based dimensional system.
In contrast, the FFM does not presuppose any personality disorder categories and instead describes personality in a continuous manner along a set of 30 traits (or "facets") grouped into 5 overarching factors. Figure 1B shows the FFM profile of a prototypical patient with antisocial personality disorder. If a person has a high score on a given facet (eg, anxiousness), he/she is better described by the high adjectives (eg, fearful) than the low adjectives (eg, relaxed).

The FFM has a number of advantages regarding construct validity: it has been shown to be biologically based, universal, temporally stable, and related to life outcomes. Furthermore, because the FFM describes people continuously along 30 facets rather than with discrete disorders, it avoids many of the aforementioned disadvantages of the DSM-IV. For example, the issue of high comorbidity is irrelevant with the FFM because no categorical diagnoses are given; similarly, the problem of arbitrary diagnostic thresholds of personality disorders is also moot because the FFM does not implement cutoffs specifying the presence versus absence of a disorder. Other trait-based dimensional systems, although differing in the choice of specific traits, share these strengths with the FFM.

However, considerable cognitive-processing challenges may be inherent to any trait-based dimensional system. Specifically, the facets or traits may be fundamentally ambiguous. Previous research in cognitive science suggests that descriptors are relative to the categories they describe (eg, large molecule versus large mountain; open hand versus open bottle; strong woman versus strong man). As a result, descriptors are inherently ambiguous without the context of an accompanying category. Translated to the domain of personality pathology, when an FFM facet is used without the context of a diagnostic category, it can be ambiguous in a similar way. For instance, a low score on the "gregariousness" FFM facet could correspond to paranoid fears (as in paranoid personality disorder), fear of not being liked by others (avoidant personality disorder), or indifference to others (schizoid personality disorder). A high score on the "anger" facet could correspond to either temper tantrums (histrionic personality disorder) or lack of control over anger (borderline personality disorder). While the features used in the DSM-IV diagnostic criteria are less likely to be ambiguous because the descriptors tend to refer to observable behaviors (eg, "perceives attacks on his or her character or reputation that are not apparent to others and is quick to react angrily or to counterattack") and are framed in the context of a diagnosis, FFM traits are unobservable (eg, in Figure 1B, "angry" or "bitter") and can be ambiguous if presented without any diagnostic context. This ambiguity in FFM patient descriptions could pose problems for clinical functions such as determining prognoses or developing treatment plans. Previous studies comparing the clinical utility of the FFM with that of the DSM-IV have not examined this issue of ambiguity inherent to FFM, with the exception of a recent study by our group, Rottman et al. Because the current study utilizes the task used in that work, we describe the task in detail here.

Back-Translation Task Used in Rottman et al

Both Rottman and colleagues’ study and the current study used a back-translation task to examine whether trait-based descriptions of patients may be clinically ambiguous. In the back-translation task, participants are presented with patient descriptions in the FFM format (Figure 1B), which were taken from previous studies in which experienced clinicians thought about a prototypical case of each of the 10 DSM-IV personality disorders or comorbid cases and rated each case on the 30 FFM facets. Then, participants are asked to "back-translate" these FFM descriptions by identifying any known DSM-IV disorders found in the descriptions.

The logic behind this back-translation task is the following. We begin with the assumption that practicing clinicians are familiar with the DSM-IV personality disorders. Presenting clinicians with the DSM-IV diagnostic criteria and having them make DSM-IV diagnoses tests the validity of this assumption. Having demonstrated that the DSM-IV personality disorders are familiar concepts to the clinicians, the next step is to determine whether they can recognize these known concepts from the FFM trait descriptions. If traits are indeed ambiguous, such that a score on one facet (eg, a low score on gregariousness) could correspond to multiple DSM-IV diagnoses, then clinicians should have difficulty identifying correct DSM-IV diagnoses from FFM descriptions alone (eg, Figure 1B). That is, identifying DSM-IV diagnoses from FFM profiles would be a 1-to-many mapping. If traits or a set of traits taken as a whole are not ambiguous, clinicians should be able to readily recognize the DSM-IV personality disorders from FFM descriptions alone.

The outcome of this task is not obvious and needs to be tested empirically for the following reasons. Previous studies have demonstrated that the FFM is comprehensive enough to reliably describe the DSM-IV personality disorders. For example, in Lynam and Widiger's study, experts in personality disorders were asked to consider prototypical cases of each of the DSM-IV personality disorders and to rate them on the 30 facets that constitute the FFM. Average interrater reliability was good, ranging from 0.48 to 0.66. Samuel and Widiger subsequently demonstrated that practicing clinicians could also describe the personality disorders in terms of the FFM with fairly high interrater agreement, ranging from 0.64 to 0.78. Samuel and Widiger also found extremely high agreement between the prototypes derived from practicing clinicians and those from experts in personality disorders. These studies suggest that clinicians can reliably translate existing concepts of personality disorders into FFM ratings. On the basis of these results, Lynam and Widiger stated that "the DSM-IV PDs can be understood from the dimensional perspective of the FFM." In a review article, Clark also stated that the DSM-IV personality disorders "can be characterized with the FFM conceptually ... and empirically." If the traits can capture the DSM-IV personality disorders in a reasonably unambiguous manner, clinicians should at least be able to recognize prototypical DSM-IV diagnoses from the
trait-based descriptions alone. The current study examines whether this is indeed the case.

One might criticize the back-translation task for relying on the DSM-IV diagnoses, the very concepts that the proponents of the trait-based dimensional systems propose eliminating due to the problems discussed earlier. This contention, however, is irrelevant to the aims of the current study; even if the DSM-5 does not use the same diagnostic categories as DSM-IV (or even if it eliminates diagnoses entirely), the back-translation task should nonetheless effectively assess whether there are ambiguities in traits. The back-translation task merely uses the categories already known to the clinicians as an established baseline and is agnostic as to the validity of these categories (see Discussion for more detail).

It may also be argued that there are ways to disambiguate or contextualize the traits with supplementary information, such as identifying dysfunctional behaviors associated with extreme trait scores (see the Discussion). Yet, the first order of business before endeavoring to implement such steps is to empirically examine whether or not traits are indeed ambiguous.

Rottman et al. presented the back-translation task to practicing clinical psychologists, psychiatrists, and clinical social workers. They found that on average, clinicians identified correct diagnoses for only 47% of prototypical cases and only 21% of comorbid cases when they were described by the FFM traits alone. This finding cannot be attributed to a lack of knowledge of the DSM-IV, because the same clinicians had relatively little difficulty identifying correct diagnoses presented in the DSM-IV format; on average, the clinicians identified correct diagnoses for 82% of prototypical cases and 60% of comorbid cases when they were written in the DSM-IV format. In other words, clinicians had difficulty disambiguating the meaning of FFM patient descriptions even for well-known, prototypical DSM-IV disorders. The clinicians also rated the FFM as less clinically useful than the DSM. In sum, it appears that the FFM requires supplementary contextual information for clinicians to effectively disambiguate the meanings of the FFM’s facets for any given patient.

Experts in Personality Disorders

Rottman and colleagues’ focus on practicing clinicians demonstrated some of the cognitive difficulties that would be faced by mental health professionals using the FFM to make personality disorder diagnoses. However, 2 important issues were not fully addressed in this previous work.

First, many practicing clinicians, such as those tested by Rottman et al., quite likely specialize in disorders other than personality disorders (e.g., Axis I disorders) and, as such, may have been unable to use the FFM to its full potential in that study, which focused solely on personality disorders. In contrast, research in cognitive science would suggest that clinical-research personality disorder experts who have specialized in building knowledge and theories about the causal workings of personality disorders relevant to FFM facets could better identify important correlations between scores on FFM facets for personality disorders. Furthermore, identifying important correlations between FFM facets could help personality disorder specialists integrate the information across the 30 facets and form a more coherent concept of a patient, benefiting diagnosis and other clinical functions. For instance, although a low score on the “gregariousness” facet may be ambiguous on its own, a combination of low “gregariousness” and low “trust” scores may indicate that a patient has paranoid personality disorder, whereas a combination of low “gregariousness” and high “self-consciousness” scores may indicate that a patient has avoidant personality disorder. A similar finding has been demonstrated in chess experts, who are able to quickly perceive combinations of chess pieces and positions as meaningful “chunks” bound by relations such as attack and defense. In sum, having specialized knowledge in personality disorders may help reduce the effects of ambiguity in the FFM, in which case personality disorder experts should be able to overcome these challenges of working with the FFM. If this is true, then conceivably the problems with ambiguity documented by Rottman et al. are not necessarily inherent to the FFM, but rather could be attributed to the background of the clinicians in that previous study, a factor that might readily be overcome with specialized training. To test this possibility, in the current study, personality disorder researchers were studied to tap into a population highly likely to have maximal knowledge about personality disorders.

The second critical issue addressed in the current study is that Rottman and colleagues’ clinicians self-reported being relatively unfamiliar with the FFM and also being considerably less familiar with the DSM than with the DSM-IV. It remains possible that these clinicians had a harder time working with the FFM simply because the system was new to them. If so, it may be that any potential cognitive difficulties with a trait-based assessment system would be attenuated once the system becomes more familiar. The personality disorder researchers tested in the current study, in contrast, should be familiar with both the FFM and the DSM-IV. We were also able to identify a subset of researchers reporting equal familiarity with the FFM and DSM-IV.

If expert knowledge contributes to perceived clinical utility of the FFM, then the current study provides a more comprehensive test of the utility of the FFM compared to Rottman and colleagues’ study of practicing clinicians. However, if a group of personality disorder researchers, including those with notable FFM expertise, have difficulty disambiguating FFM descriptions, this outcome would suggest that the cognitive difficulties previously attributed to the FFM in Rottman and colleagues’ study are not likely to be overcome with experience or increased knowledge of personality disorders or through more extensive exposure to and experience with the FFM.

METHOD

Participants

In line with previous research, we identified people with specialized knowledge of personality pathology by conducting
Failure to conform to social norms with respect to lawful behaviors as indicated by repeatedly performing acts that are grounds for arrest.

Deceitfulness, as indicated by repeated lying, use of aliases, or conning others for personal profit or pleasure.

Impulsivity or failure to plan ahead.

Irritability and aggressiveness, as indicated by repeated physical fights or assaults.

Reckless disregard for safety of self or others.

Consistent irresponsibility, as indicated by repeated failure to sustain consistent work behavior or honor financial obligations.

Lack of remorse, as indicated by being indifferent to or rationalizing having hurt, mistreated, or stolen from another.

B. Five-Factor Model Description

<table>
<thead>
<tr>
<th>Facets</th>
<th>Low Adjectives</th>
<th>High Adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism Facets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>2.00</td>
<td>relaxed, unconcerned, cool</td>
</tr>
<tr>
<td>Angry Hostility</td>
<td>3.93</td>
<td>even-tempered</td>
</tr>
<tr>
<td>Depressiveness</td>
<td>2.70</td>
<td>optimistic</td>
</tr>
<tr>
<td>Self-consciousness</td>
<td>1.63</td>
<td>self-assured, glit, shameless</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>4.22</td>
<td>controlled, restrained</td>
</tr>
<tr>
<td>Vulnerability</td>
<td>2.07</td>
<td>clear-thinking, fearless, unflappable</td>
</tr>
<tr>
<td>Extraversion Facets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warmth</td>
<td>2.00</td>
<td>cold, aloof, indifferent</td>
</tr>
<tr>
<td>Gregariousness</td>
<td>3.48</td>
<td>withdrawn, isolated</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>4.07</td>
<td>unassuming, quiet, resigned</td>
</tr>
<tr>
<td>Activity</td>
<td>4.00</td>
<td>passive, lethargic</td>
</tr>
<tr>
<td>Excitement-Seeking</td>
<td>4.30</td>
<td>cautious, monotonous, dull</td>
</tr>
<tr>
<td>Positive Emotions</td>
<td>3.52</td>
<td>placid, anhedonic</td>
</tr>
<tr>
<td>Openness Facets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td>3.48</td>
<td>practical, concrete</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>2.78</td>
<td>uninvolved, no aesthetic interests</td>
</tr>
<tr>
<td>Feelings</td>
<td>2.41</td>
<td>constricted, unaware, alexythmic</td>
</tr>
<tr>
<td>Actions</td>
<td>4.07</td>
<td>routine, predictable, habitual, stubborn</td>
</tr>
<tr>
<td>Ideas</td>
<td>3.26</td>
<td>pragmatic, rigid</td>
</tr>
<tr>
<td>Values</td>
<td>3.48</td>
<td>traditional, inflexible, dogmatic</td>
</tr>
<tr>
<td>Agreeableness Facets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>1.70</td>
<td>skeptical, cynical, suspicious, paranoid</td>
</tr>
<tr>
<td>Straightforwardness</td>
<td>1.41</td>
<td>cunning, manipulative, deceptive</td>
</tr>
<tr>
<td>Altruism</td>
<td>1.41</td>
<td>stingy, selfish, greedy, exploitative</td>
</tr>
<tr>
<td>Compliance</td>
<td>1.81</td>
<td>oppositional, combative, aggressive</td>
</tr>
<tr>
<td>Modesty</td>
<td>1.70</td>
<td>confident, boastful, arrogant</td>
</tr>
<tr>
<td>Tendermindedness</td>
<td>1.52</td>
<td>tough, callous, ruthless</td>
</tr>
<tr>
<td>Conscientious Facets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>2.52</td>
<td>lax, negligent</td>
</tr>
<tr>
<td>Order</td>
<td>2.74</td>
<td>haphazard, disorganized, sloppy</td>
</tr>
<tr>
<td>Outfulness</td>
<td>1.52</td>
<td>casual, undependable, unethical</td>
</tr>
<tr>
<td>Achievement Striving</td>
<td>2.33</td>
<td>aimless, desultory</td>
</tr>
<tr>
<td>Self-Discipline</td>
<td>1.85</td>
<td>hedonistic, negligent</td>
</tr>
<tr>
<td>Deliberation</td>
<td>1.96</td>
<td>hasty, careless, rash</td>
</tr>
</tbody>
</table>

*Descriptions from DSM-IV-TR\(^1\) (1A) and Samuel and Widiger\(^26\) (1B); reprinted with permission from the American Psychiatric Association.\(^1\)\(^{25}\)

Figure 1. DSM-IV and FFM Descriptions of a Prototypical Case of Antisocial Personality Disorder\(^a\)

A. DSM-IV Description

a search in the PsycINFO database for authors who had published at least 3 articles with the keyword “personality disorder” in peer-reviewed journals and who had published at least 1 article from January 2006 through mid-November 2008 (the time during which our search was conducted). We then excluded those for whom we could not find contact information and those who were highly likely to already be familiar with Rottman and colleagues’ study. Recruitment e-mails were sent to the remaining 476 researchers in December 2008. At the beginning of the study, we requested that participants verify that they consider personality disorders to be among their primary research interests and that they have been conducting research on personality disorders for at least 4 years. This verification allowed us to exclude those who collaborated on personality disorder articles only because of expertise in other fields (eg, statisticians). Seventy-three participants completed the experiment. The experiment took 29 minutes on average, and participants received either a $60 gift certificate to an online retailer or a $60 check.

**Materials and Design**

Twelve different cases were described in both the FFM and DSM styles depicted in Figure 1. Ten described prototypical patients, each having only 1 of the 10 DSM-IV...
personality disorders. The remaining 2 were comorbid cases with 2 personality disorders each; these were included because comorbid cases have been argued to be more representative of real-world patients.33

The FFM facet scores were taken from previous studies in which practicing clinicians thought about prototypical personality disorder cases26 and about comorbid case vignettes22 and rated each on the 30 FFM facets. The FFM-style descriptions presented to participants contained both the mean rating for each facet obtained from these studies and a plot of the facet scores, anchored by high (eg, “fearful, apprehensive” for anxiousness) and low (eg, “relaxed, unconcerned, cool” for anxiousness) adjectives (the same descriptions used by Rottman et al,25 eg, Figure 1B). For the DSM-style descriptions, each prototypical case comprised all the DSM-IV-TR diagnostic criteria for that personality disorder (eg, Figure 1A). The comorbid DSM-style descriptions were taken from a pretest by Rottman et al,25 in which clinicians identified all the DSM-IV-TR personality disorder symptoms they found to be present in the comorbid vignettes.22

The 12 cases were divided into 2 groups, each containing 5 prototypical cases and 1 comorbid case. For diversity, each group included at least 1 disorder from each of the 3 clusters of personality disorders in the DSM-IV, and the diagnoses of the comorbid case did not match the diagnoses of any of the prototypical cases in the group. To the extent possible, we also matched the 2 groups of prototypical cases for difficulty of diagnosis, as previously determined.25

Each participant saw 1 group of 6 cases presented in the FFM style and the other group in the DSM style. Thus, descriptive style (DSM vs FFM) was a within-subject variable. The pairing of cases with descriptive style, presentation order of the 2 groups, and order of the styles were counterbalanced across participants. The order of the 6 cases within each group was randomized.

Procedure

The study was performed online using Qualtrics software (Qualtrics Labs, Inc, Provo, Utah). Participants were told that they would be presented with descriptions of adult patients and were asked to imagine that these patients were referred to them along with a patient description from a previous consultation. Participants were told that the patients “do not have schizophrenia or any other psychotic disorder, and their symptoms do not occur due to the direct effect of any general medical condition.” This instruction was included to prevent participants from avoiding giving personality disorder diagnoses for reasons not of experimental interest (eg, in the DSM-IV, a schizoid personality disorder diagnosis is not allowed if it occurs exclusively during the course of schizophrenia). Finally, participants were instructed not to consult the DSM-IV or other references during the experiment.

Next, participants were presented with the first group of 6 cases in either the DSM or FFM style. After each individual case, participants were asked to “provide any DSM-IV diagnoses you believe this patient to have.” Participants also rated their confidence in each diagnosis on a 7-point scale (1 = “not confident at all,” 4 = “somewhat confident,” and 7 = “very confident”).

After the first group of cases was presented, participants rated the utility of the descriptive system that they just saw by answering the following 6 questions on a 5-point scale (1 = “not at all,” 2 = “slightly,” 3 = “moderately,” 4 = “very,” 5 = “extremely”):25

1. How informative are these descriptions in making diagnoses for these people?
2. How informative are these descriptions in devising treatment plans for these people?
3. How useful do you feel the system used to describe these people would be for communicating information about them with other mental health professionals?
4. How useful do you feel the system used to describe these people would be for communicating information about them to themselves?
5. How useful is the system used to describe these people for comprehensively describing all the important personality problems they have?
6. How useful was the system used to describe these people for describing their global personalities?

Participants then performed the same series of tasks for the second group of cases. Finally, participants provided demographic information and rated their own familiarity with the DSM-IV and FFM systems, respectively, on a 7-point scale (1 = “not familiar at all,” 4 = “moderately familiar,” 7 = “extremely familiar”). Participants gave informed consent, and this study was approved by the Yale Institutional Review Board (New Haven, Connecticut).

RESULTS

Demographics

Seventy-three researchers (51 PhDs, 16 MDs, 2 MD/PhDs, 3 MAs, and 1 MSW) participated between December 2008 and January 2009. Forty-eight of these researchers also saw patients. This subset of clinician-researchers had been in practice for a mean of 15 (SD = 9) years and worked specifically with patients with personality disorders a mean of 13 (SD = 11) hours weekly. On average, they received their highest degree in 1994, about 14 years before this study was conducted. Participants had published a median of 15 articles on personality disorders (mean = 24; range, 3–160) and had been conducting research on personality disorders for a mean of 15 (SD = 8) years. Overall, participants reported being more familiar with the DSM than the FFM (t27 = 7.70, P < .01). However, the current participants were more familiar with the DSM (mean = 6.40, SD = 0.95) and, more importantly, with the FFM (mean = 4.97, SD = 1.66) than were the clinicians in the study by Rottman et al25 (mean = 5.68, SD = 1.26, for the DSM; mean = 2.17, SD = 1.65, for the FFM, t17.4.2 = 4.89, P < .01, equal variances not assumed for the DSM; t25.2 = 12.24, P < .01, for the FFM). Furthermore, a 2 (DSM vs FFM) × 2 (clinicians vs researchers) analysis of variance
(ANOVA) revealed a significant interaction ($F_{1,252} = 71.91, P < .01$); although the clinicians in the study by Rottman et al$^{25}$ were much more familiar with the DSM-IV than the FFM, this difference was markedly smaller for the researchers in the current study. Whereas the clinicians in the study by Rottman et al$^{25}$ rated themselves as significantly below the midpoint of “moderately familiar” with the FFM ($t_{183} = 14.94, P < .01$), the researchers in the current study rated themselves significantly above the midpoint ($t_{23} = 5.10, P < .01$). Both main effects were also significant. The current researcher-participants were more familiar with the DSM than FFM ($F_{1,252} = 127.03, P < .01$), and generally gave higher familiarity ratings than Rottman and colleagues’ clinicians$^{26}$ ($F_{1,252} = 400.02, P < .01$).

For each of the analyses below, we will also refer to a subgroup of participants who rated themselves as equally familiar with the FFM and DSM (mean = 6.42, SD = 0.93, for both systems), again significantly above the midpoint of “moderately familiar” ($t_{23} = 12.74, P < .01$). This subgroup consists of 24 researchers: 18 PhDs, 2 MDs, 2 MD/PhDs, and 2 MAs. On average, they received their highest degree in 1991, 17 years before this study was conducted, and had published a median of 20 articles on personality disorders.

### Diagnostic Accuracy

The prototypical cases were analyzed by averaging across the 5 prototypical cases seen by each individual in each system. Thus, a score of 1 means that a participant gave correct diagnoses for all 5 cases and a score of 0 means that the participant gave no correct diagnoses for any case. Participants almost always gave the correct diagnosis in the DSM condition (mean = 0.99, SD = 0.06) and were much more accurate in the DSM condition than in the FFM condition (mean = 0.62, SD = 0.25; $t_{23} = 12.36, P < .01$) (Figure 2). See also Table 1 for the results broken down by personality disorder. This pattern of means held true across all 10 disorders.

#### Table 1. Mean Number of Correct and Incorrect Diagnoses per Case by Personality Disorder

<table>
<thead>
<tr>
<th>Personality Disorder</th>
<th>Correct</th>
<th>Incorrect</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DSM</td>
<td>FFM</td>
<td>DSM</td>
<td>FFM</td>
</tr>
<tr>
<td>Cluster A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paranoid</td>
<td>0.98</td>
<td>0.61</td>
<td>0.25</td>
<td>0.76</td>
</tr>
<tr>
<td>Schizoid</td>
<td>0.97</td>
<td>0.58</td>
<td>0.12</td>
<td>0.93</td>
</tr>
<tr>
<td>Schizotypal</td>
<td>0.98</td>
<td>0.67</td>
<td>0.50</td>
<td>0.76</td>
</tr>
<tr>
<td>Cluster B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antisocial</td>
<td>1.00</td>
<td>0.76</td>
<td>0.13</td>
<td>0.91</td>
</tr>
<tr>
<td>Borderline</td>
<td>1.00</td>
<td>0.55</td>
<td>0.03</td>
<td>0.98</td>
</tr>
<tr>
<td>Histrionic</td>
<td>1.00</td>
<td>0.20</td>
<td>0.06</td>
<td>0.65</td>
</tr>
<tr>
<td>Narcissistic</td>
<td>0.98</td>
<td>0.63</td>
<td>0.15</td>
<td>0.88</td>
</tr>
<tr>
<td>Cluster C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoidant</td>
<td>1.00</td>
<td>0.75</td>
<td>0.21</td>
<td>0.85</td>
</tr>
<tr>
<td>Dependent</td>
<td>0.98</td>
<td>0.64</td>
<td>0.13</td>
<td>0.79</td>
</tr>
<tr>
<td>Obsessive-Compulsive</td>
<td>1.00</td>
<td>0.88</td>
<td>0.00</td>
<td>0.48</td>
</tr>
</tbody>
</table>

*For correct diagnoses, responses were scored as either 0 (the correct diagnosis was not provided) or 1 (the correct diagnosis was provided). There was no upper limit on the number of incorrect diagnoses participants could provide per case.

Abbreviations: DSM = Diagnostic and Statistical Manual of Mental Disorders, FFM = Five Factor Model.

The comorbid cases were analyzed by examining the proportion of correct diagnoses within each condition (ie, the DSM or the DSM-style comorbid case). Since there are 2 correct diagnoses for a given comorbid case, a score of 1 means that a participant correctly identified both diagnoses, a score of 0.5 means that a participant identified 1 of the 2 correct diagnoses, and a score of 0 means that a participant identified neither of the correct diagnoses. For comorbid cases, participants were again more likely to give the correct diagnoses in the DSM (mean = 0.77, SD = 0.26) than in the FFM condition (mean = 0.48, SD = 0.33; $Z = 5.03, n = 73, P < .01$) (Figure 2). Nonparametric Wilcoxon signed rank tests were used to analyze the accuracy of comorbid cases because there are few levels of the outcome variables.

Figure 2 also presents results from Rottman and colleagues’ investigation$^{25}$ of clinicians not necessarily specializing in personality disorders for comparison with the current results. Due to differences in design, inferential statistics are not possible. As can be seen, across the prototypical and comorbid cases, the personality disorder researchers in the current study provided more accurate diagnoses than Rottman and colleagues’ practising clinicians,$^{25}$ but importantly, they did so in both the DSM and FFM conditions. Other methods of counting correct or incorrect diagnoses (eg, not counting features or traits, or not counting “obsessive-compulsive” as Axis II obsessive-compulsive personality disorder but rather as Axis I obsessive-compulsive disorder) would not change the main results.

Incorrect diagnoses (Figure 3), defined as any DSM-IV diagnosis mismatching the correct diagnosis and any non-DSM-IV diagnosis, were examined. Participants could provide any number of incorrect diagnoses per case. Participants gave significantly more incorrect diagnoses per prototypical case in the DSM (mean = 0.79, SD = 0.48) than DSM condition (mean = 0.16, SD = 0.35; $t_{17} = 9.82, P < .01$). Again, this pattern of means held true across all 10 disorders (see Table 1). For the comorbid cases, they also gave more incorrect diagnoses.
per case in the FFM (mean = 0.81, SD = 0.84) than DSM condition (mean = 0.30, SD = 0.64; Z = 3.95, n = 73, P < .01). Again, these results differed little from those of practicing clinicians in Rottman and colleagues’ study except for being generally more accurate across both conditions.

Next, the frequencies of correct and incorrect diagnoses within the subgroup of participants who rated themselves as equally familiar with the DSM-IV and FFM were examined. The results closely replicate those for the entire sample, suggesting that familiarity did not influence diagnostic accuracy. The participants in this subgroup more frequently gave correct diagnoses for prototypical cases in the DSM (mean = 1.00, SD = 0.00; these participants always gave the correct diagnosis in the DSM) than FFM (mean = 0.62, SD = 0.27; t3 = 6.96, P < .01) and gave more incorrect diagnoses for the FFM (mean = 0.75, SD = 0.54) than the DSM (mean = 0.13, SD = 0.25; t3 = 5.15, P < .01). The same results hold for comorbid cases: participants more frequently gave correct diagnoses in the DSM (mean = 0.75, SD = 0.26) than FFM condition (mean = 0.48, SD = 0.31; Z = 2.97, n = 24, P < .01), and gave more incorrect diagnoses in the FFM (mean = 0.75, SD = 0.94) than DSM condition (mean = 0.13, SD = 0.34; Z = 2.58, n = 24, P = .01).

Correlational analyses were also conducted between familiarity ratings and frequency of correct/incorrect diagnoses for the entire set of participants. The most important reason to look at these correlations is to determine whether familiarity with the FFM increases accuracy in identifying diagnoses from FFM patient profiles. If so, such a correlation would suggest that familiarity with the FFM facilitates being able to form a coherent image of a patient from an FFM patient profile. However, this possibility was not supported. Familiarity with the FFM did not correlate significantly with providing correct diagnoses in the FFM condition (r = 0.08 and r = 0.11 for prototypical and comorbid cases, respectively, NS). Familiarity with the FFM also did not help participants avoid providing incorrect diagnoses in the FFM condition (r < 0.01 and r = −0.01 for prototypical and comorbid cases respectively, NS).

**Confidence in Diagnoses**

A 2 (correct vs incorrect diagnosis) × 2 (DSM vs FFM) repeated-measures ANOVA revealed that participants were more confident making diagnoses in the DSM than in the FFM condition (F1,25 = 45.15, P < .01, ηp2 = .64) and more confident for correct than incorrect diagnoses overall (F1,25 = 75.25, P < .01, ηp2 = .75). This within-subjects analysis can be conducted for only the 26 participants who gave at least 1 correct and 1 incorrect diagnosis (and consequently their corresponding confidence ratings) in both the DSM and FFM. To increase the number of subjects who could be included in this analysis, prototypical and comorbid cases were both included. For example, to obtain the average confidence rating for correct diagnoses in the FFM condition, the mean was computed over whichever of the 6 FFM cases (5 prototypical and 1 comorbid) participants provided correct diagnoses.

In addition, there was an interaction (F1,25 = 53.15, P < .01, ηp2 = .68) indicating that, although participants were much more confident in correct than incorrect diagnoses in the DSM condition, there was a much smaller difference in confidence between correct and incorrect diagnoses in the FFM condition (Figure 4). These findings suggested that participants were more aware of the accuracy of their diagnoses in the DSM than FFM condition. Familiarity with the FFM was not significantly correlated with confidence for correct or incorrect diagnoses.

**Clinical Utility Ratings**

The mean clinical utility ratings broken down by the DSM and the FFM condition are presented in Figure 5. Paired t tests revealed that participants found the DSM-IV to be more useful than the FFM on 3 measures: prognosis, treatment plans, and communicating with professionals (all t69 > 2.19, all P < .05). Participants rated the FFM as more useful than the DSM for communicating with patients (t69 = 3.03, P < .01), possibly because the DSM-IV disorder names are considered
to be stigmatizing and because the FFM facets are common terms rather than technical disorder names. There was no difference between the DSM and FFM for comprehensively describing all important personality problems and global personality description (P values > .10). All of these patterns of results also hold when including data from only the condition presented first.

The clinical utility ratings for the subset of participants who were equally familiar with the DSM-IV and FFM were examined (Figure 6). One participant in the subset of 24 did not include utility ratings for the DSM and is excluded from analyses. Paired t tests showed that the FFM was not rated as more useful than the DSM for 4 of the 6 aspects of clinical utility. Participants in the subgroup did rate the FFM as more useful than the DSM for communicating with patients and describing global personality, which makes sense because the FFM is based on common adjectives describing personality and was meant to describe all types of personality, not just pathological personality. In summary, however, the clinical utility ratings do not suggest that participants found the FFM to be clearly more useful than the DSM, a finding that would be necessary to support a switch to the FFM on grounds of increased clinical utility.

DISCUSSION

Recent work has shown that the FFM poses cognitive challenges and has relatively low clinical utility for practicing clinicians, if presented without context to disambiguate FFM traits. However, these previous findings were obtained in a broad sample of practicing clinicians who rated themselves as much more familiar with the DSM-IV than the FFM and who may not necessarily specialize in personality disorders per se (as opposed to other, Axis I disorders). In the current study, we examined whether people with specialized experience and knowledge about personality disorders—personality disorder researchers—especially those who are equally familiar with the FFM and the DSM, are able to overcome the cognitive challenges of the FFM. The current results, in conjunction with those reported by Rottman et al., suggest that FFM traits alone may be too ambiguous as a diagnostic tool for practicing clinicians.

In the current study, experts in personality disorders, as established by their record of published research and self-identified primary interests, had difficulty identifying even highly familiar, prototypical DSM-IV diagnoses from DSM profiles. Correlational analyses ruled out the possibility that participants’ degree of familiarity with the FFM was likely to be responsible for the observed problems in identifying correct diagnoses from FFM profiles. A subgroup of participants reporting equal familiarity with the DSM-IV and FFM also had difficulty in identifying prototypical DSM-IV diagnoses from DSM profiles. This finding is consequential because it suggests that even equal familiarity with the FFM and DSM-IV is insufficient to form a coherent image of a patient from an FFM profile alone.

One could argue that, to the extent that the DSM-IV personality disorders lack validity, it is not particularly important to be able to use them to conceive of a case. Yet, completely abandoning them would pose considerable disruption from a practical standpoint (eg, disruption of ongoing research, difficulty for clinicians in implementing past research findings into clinical practice). Perhaps most importantly,
clinicians have been working with a categorical personality disorder system since 1980; they cannot simply turn off their prior knowledge and experience, nor could it conceivably be desirable for them to do so. Furthermore, as mentioned earlier, certain DSM-IV personality disorders, particularly borderline personality disorder, have been acknowledged to be useful constructs even by proponents of the FFM.29 Yet, only 55% of researchers in the current study were able to identify the prototypical FFM trait pattern as borderline. Such difficulty in recognizing useful constructs in FFM case profiles suggests a problem with the FFM’s clinical utility.

The researcher-participants in the current study also judged the clinical utility of the FFM to be low in a number of aspects, further suggesting that they found the FFM descriptors to be ambiguous. Specifically, participants judged an abstract FFM patient description (eg, a neurotic, anxious, and introverted person) to be less useful in making treatment plans and predictions about the course and outcome of the patient than a DSM description. Participants also thought that the disorder category names of the DSM-IV greatly facilitated communication between mental health professionals who know the terminology, although they thought that the commonplace adjectives used by the FFM were more useful for communicating with patients who, presumably, are less likely to know diagnostic terminology.

Investigating how people use the FFM may reveal additional issues to be considered in formulating potential diagnostic systems incorporating the FFM or other trait-based dimensional systems. In the current study, participants were able to identify some DSM-IV disorders much better than others when examining FFM profiles (see Table 1). For example, histrionic personality disorder was correctly identified only 20% of the time on the basis of an FFM profile. In contrast, obsessive-compulsive personality disorder was correctly identified 88% of the time from an FFM profile. Obsessive-compulsive personality disorder may be particularly easy to identify since the conscientiousness facets all receive very high scores (between 4 and 5), whereas none of the conscientiousness facets receive above a 4 in any of the other personality disorder prototypes.26 (In fact, a cluster analysis of the 10 FFM prototypes revealed that obsessive-compulsive personality disorder is the most distinctive of the 10 personality disorders.) Because the FFM conscientiousness facets are diagnostic of obsessive-compulsive personality disorder, a high score on the conscientiousness facets is not ambiguous—they primarily occur only for obsessive-compulsive personality disorder. Thus, highly distinctive and diagnostic facets allowed our participants to more easily recognize the disorder. Participants’ poor performance on histrionic personality disorder may be due to the fact that its facet scores are quite moderate, and thus it may be hard to determine which facets are clinically relevant.

Future Research

To what extent will other trait-based dimensional systems face the same cognitive challenges relating to ambiguity demonstrated in the current study? Many other systems have scales that can be mapped closely onto the FFM facets.34 For example, like the FFM, the Schedule for Nonadaptive and Adaptive Personality (SNAP-2)27-35 and the Dimensional Assessment of Personality Pathology (DAPP)8,36 have “impulsivity” and “mistrust”/“suspiciousness” scales. Given the degree to which such systems overlap with the FFM, we speculate that they may contain a similar degree of ambiguity. Future research, however, will be necessary to definitively assess whether or not this is indeed the case.

Another crucial future research direction is to identify ways in which the trait descriptors can be successfully disambiguated. One possible remedy is to combine trait-based dimensional systems with type-based ones (eg, the prototypes of DSM-IV personality disorders tested by Spitzer et al23). For example, a clinician might first determine how similar a patient is to different personality prototypes (eg, borderline, antisocial, etc) along a dimensional scale and then use a trait-based assessment to further describe the patient. The idea is that the initial prototype assessment would instantiate the more specific meanings of the traits for this patient. For example, rather than thinking about a patient as “withdrawn,” a clinician could think of the patient as “withdrawn due to paranoid fears” (as in the paranoid type) or “withdrawn due to indifference to others” (as in the schizoid type). We suggest that in general, instantiated descriptors are likely to be more clinically meaningful and useful for clinicians than uninstantiated (and thereby ambiguous) ones.

Finally, future work might test whether optional steps proposed to supplement the FFM will successfully reduce ambiguity. One such proposal is to assess dysfunctional behaviors associated with abnormal trait scores (eg, dysfunctional behaviors such as “overspending” or “excessive gambling” or “excessive use of drugs” are associated with high impulsiveness scores; behaviors such as “readily perceives malevolent intentions within benign, innocent remarks or behaviors” or “is often involved in acrimonious arguments with friends” are associated with low trust scores29; see Clark10 for a review). Unfortunately, no empirical studies have yet examined whether clinicians find supplemental dysfunctional behaviors to be useful and whether they actually can use these supplements. For instance, although Samuel and Widiger26 and Lynam and Widiger27 have shown that researchers and clinicians can reliably assess FFM traits for prototypical cases of the DSM-IV personality disorders, no studies have yet examined whether they can also reliably identify which dysfunctional behaviors are associated with these prototypical cases.

Once the translatability of these dysfunctional behaviors from the familiar DSM-IV constructs is empirically established, additional research can further examine whether the descriptions based on these dysfunctional behaviors are unambiguous enough to be translated back to the DSM-IV constructs, as in the current study. It may be that dysfunctional behaviors would help to clarify the context of an extreme trait and improve clinicians’ ability to back-translate to the DSM. On the other hand, the existing catalog of dysfunctional behaviors29,37 may not clarify the context of

J Clin Psychiatry e9


PSYCHIATRIST.COM
the traits, because it may not have been developed with the ambiguities of traits in mind. For example, the existing catalog lists only 2 dysfunctional behaviors associated with low ratings on the gregariousness trait—"is socially isolated" and "has no apparent support network due to his or her own social withdrawal"—but clinicians may not be able to determine, based on these specified dysfunctions, whether this low gregariousness is paranoid fear, fear of not being liked by others, or indifference to others. To give yet another example, high excitement seeking is associated with the following dysfunctional behaviors: "engages in a variety of reckless and even highly dangerous activities; behavior is rash, foolhardy, and careless" and "easily bored; excessive thrill seeker." It is not clear whether clinicians will be able to use these dysfunctional behaviors to differentiate between narcissistic, antisocial, borderline, and histrionic personality types. Since all existing studies on the clinical utility of dimensional systems have focused on the trait description itself without consideration of the associated dysfunctions, if inclusion of such dysfunctions is to be considered for the DSM-5, future research must empirically determine whether the dysfunction assessment does indeed disambiguate traits.

**Author affiliation:** Department of Psychology, Yale University, New Haven, Connecticut (Mr Rottman and Dr Ahn); Department of Psychology, Northeastern University, Boston, Massachusetts (Dr Kim); and Department of Psychology, Wesleyan University, Middletown, Connecticut (Dr Sanislow).

**Potential conflicts of interest:** The authors report no competing interests.

**Funding/support:** This research was supported by a National Science Foundation Graduate Research Fellowship (Mr Rottman) and National Institute of Mental Health Grants MH08-1407 (Dr Kim), MH77377 (Dr Ahn), and MH73708 (Dr Sanislow).

**Previous presentation:** Presented in part in the 31st Annual Conference of the Cognitive Science Society; June 2009; Amsterdam, The Netherlands.

**Acknowledgment:** The authors thank Rachel Litwin, BA, for help identifying and recruiting the personality disorder researchers and coding data; she reports no conflict of interest.

**REFERENCES**


