Published in final edited form as:

J Psychopathol Behav Assess. 2009 December ; 31(4): 358–367. doi:10.1007/s10862-009-9130-8.

Comparison of Three Self-Report Measures of Personality Pathology

Mayumi Okada and Thomas F. Oltmanns

Washington University in St. Louis

Abstract

Levels of convergence among three measures of personality pathology, the Personality Diagnostic Questionnaire-4+ (PDQ-4+), the Structured Clinical Interview for DSM-IV Axis II Personality Disorders Questionnaire (SCID-IIQ) and the Multi-source Assessment of Personality Pathology (MAPP) were examined. Each questionnaire was administered three times in an alternating sequence over nine consecutive weekdays to a sample of college students. There was some degree of convergence among the three instruments, but there were also substantial empirical differences between them. The data suggest three related conclusions: (1) in general, the self-report version of the MAPP is more conservative than the other two questionnaires, (2) these questionnaires should not be considered interchangeable measures of the same constructs, and (3) the breadth of measurement provided varies as a function of both the questionnaire and the specific personality disorder being measured.

A wide range of instruments, mostly in the form of questionnaires and interviews, are now available for measuring personality disorders (PDs). Although they differ in format, most PD measures rely heavily on the use of self-report because it offers an efficient way to collect information. Questionnaires in particular tend to be quick, inexpensive, and easy to administer (Bersoff & Bersoff, 2000).

There are also disadvantages associated with the use of self-report instruments. One is that they are vulnerable to the effects of social desirability biases. People tend to present themselves in a favorable manner, especially when they are asked to make judgments about attitudes and traits that are negatively valued. Another limitation of self-report instruments is that they necessarily rely on information that is consciously accessible to the individual. This problem, known as the introspective limitation, has a significant impact on the validity of information obtained using self-report instruments (Greenwald, Pickrell & Farnham, 2002).

Some authors have argued that introspective limits may play a larger role among people with personality disorders, who are frequently unable to view their behaviors in a realistic manner (Oltmanns & Turkheimer, 2006). They often believe that problems that they encounter, especially interpersonal conflicts, are not due to any fault of their own. This blindness or lack of insight that is characteristic of individuals with personality disorders makes it even more difficult to obtain accurate personality information using self-report measures. In order to obtain a more comprehensive perspective on an individual's interpersonal difficulties, several researchers have turned towards gathering additional information from informants (Klein, 2003; McDermut & Zimmerman, 2005; Westen, 1997).

In an attempt to understand how people's perceptions of themselves may differ from ways in which they are viewed by others, a new assessment tool was developed using a peer nomination

procedure (Oltmanns, Turkheimer, & Strauss, 1998; Thomas, Turkheimer, & Oltmanns, 2003). The *Multi-source Assessment of Personality Pathology (MAPP)* is composed of 105 items, including 81 items that refer to the features of ten personality disorders listed in *DSM-IV* as well as 24 supplementary items that describe additional personality traits, most of which are positive. The PD items are lay translations of the diagnostic criteria from *DSM-IV*. The original MAPP items were worded to refer to others because it was developed to collect data from informants. For example, the item that corresponds to the first criterion for Narcissistic PD reads “thinks he/she is better than other people (without good reason).”

A self-report version of the MAPP was subsequently developed using the same wording of items that was contained in the informant version. The only difference between the self- and informant-versions of the MAPP is that the self-report questions are worded to refer to the self, rather than another person. This was done so that inconsistencies between sources could be attributed specifically to a difference of opinion and not to the fact that different sets of questions were used.

The self and peer versions of the MAPP have been used to collect data from two relatively large non-clinical samples: army recruits and college students. In both samples, participants were members of existing groups and had lived together in close proximity for a standard period of time (six weeks for the recruits and five to seven months for the college students). The assessment process involved completing the peer-report and the self-report versions of the MAPP (further details are provided in Oltmanns & Turkheimer, 2006). In both samples, peer nominations produced reliable information that is largely independent of that obtained from self-report. Agreement between the two sources of information (self-report and informant-report) was low to moderate ($r = .11$ to $.35$). It was also demonstrated that people do have some awareness of the ways in which they are viewed by others, but are unlikely to report this information unless asked to do so. Finally, data from the military sample also confirmed that, in some instances, peer-reported information is more strongly associated with certain aspects of the person's future social adjustment.

To further understand the implications of the data from the Peer Nomination Project, it is important to understand how the MAPP compares with other widely used self-report instruments. The self-report version of the MAPP has never been compared with more frequently used and better validated self-report instruments that assess personality traits and pathology. In this paper, data comparing the MAPP to the Personality Diagnostic Questionnaire-4+ (PDQ-4+; Hyler, 1994) and the Structured Clinical Interview for DSM-IV Axis II Personality Disorder Questionnaire (SCID-IIQ; Spitzer, Williams, Gibbons & First, 1990) are presented.

Similar studies that have examined concordance rates between two or more diagnostic methods have found that agreements between instruments are generally low (Perry, 1992). Several studies have compared the PDQ-4+ with other personality measures, namely semi-structured interviews, and have found low levels of concordance between the PDQ-4+ and other PD measures. Additionally, past studies have found that the PDQ-4+ has a tendency to over-diagnose PDs (Zimmerman & Coryell, 1990; Hyler, Skodol, Kellman, Oldham & Rosnick, 1990). Similarly, studies that have compared the SCID-IIQ with PD interviews have also found that the SCID-IIQ produced rates of diagnoses that are substantially higher than those determined by PD interviews (Ekselius, Lindstrom, von Knorring, Bodlund & Kullgren, 1994; Jacobsberg, Perry & Frances, 1995).

Given that both the PDQ-4+ and the SCID-IIQ were both developed with the same DSM criteria in mind and that they tend to be over-inclusive, it was hypothesized that the level of agreement between these two measures would be moderately high. Also, it was hypothesized that because

the MAPP was constructed using direct lay translations of the DSM-IV criteria for personality disorders, it would be more conservative than either the PDQ-4+ or the SCID-IIQ.

The purpose of this paper is not to encourage other investigators to use the MAPP rather than these other measures. It is simply to document the psychometric properties of the self-report version of the MAPP. This information is important in the sense that interpretations of findings from the Peer-Nomination Project depend on a better understanding of the nature of the MAPP relative to other self-report measures.

Method

Sample

Participants were 203 undergraduate students (63 males, 140 females) from Washington University, all of whom gave informed consent to participate for either course credit or pay. The average age of the participants was 20 years ($SD = 1.1$ years). Of those who reported their racial or ethnic background, 132 (64%) were Caucasian, 8 (4%) were African-American, 22 (11%) were Asian or Pacific Islander, and 5 (3%) were of Hispanic or Latino decent.

Procedure

Participants completed the *Personality Diagnostic Questionnaire—4+ (PDQ-4+; Hyler, 1994)*, the *Structured Clinical Interview for DSM-IV Axis II Personality Disorders Questionnaire (SCID-IIQ; Spitzer et al., 1990)* and the self-report MAPP over nine consecutive weekdays. Each measure was given three times, spaced at least two days apart. Prior to consenting to the experiment, every participant was told that he or she would be filling out the same measures three times during the course of the study. They completed one PD instrument each day in our laboratory. At the end of this nine-day study, participants were asked to fill out a demographic data sheet, debriefed and given either research credit or a \$30 voucher for participating.

For the first 103 subjects, test order of the three personality questionnaires (PDQ-4+, SCID-IIQ, and MAPP) was varied. Thirty-nine participants followed the PDQ-4+, SCID-IIQ, MAPP sequence; 34 followed the SCID-IIQ, MAPP, PDQ-4+ sequence and 30 participants completed the MAPP first, the PDQ-4+ second, and the SCID-IIQ last. These test orders were repeated until their last day of participation. After data collection for the 103 participants was completed, a mixed model ANOVA for repeated measures was used to determine if test-order had an effect on response patterns. Three separate 3×3 models (one for each instrument) were analyzed. Results of the models suggest that for all three instruments, test-order did not have a significant effect on response patterns (all F s less than or equal to 1) nor were there any significant interactions between tests and test-order (PDQ-4+: $F(4,198) = 1.241$, $p = .295$; SCID: $F(4,200) = 1.886$, $p = .114$; MAPP: $F(4,200) = .792$, $p = .532$). Significant within group effects were found (PDQ-4+: $F(2,198) = 27.43$, $p < 0.001$; SCID-IIQ: $F(2,200) = 27.32$, $p < 0.001$; MAPP: $F(2,200) = 5.516$, $p < 0.005$) suggesting that regardless of the test or the order, participants tended to endorse more items the first time that they saw that measure compared to subsequent times (i.e. SCID-IIQ Time 1 > SCID-IIQ Time 2). Since there were no order effects, the remainder of the subjects completed the questionnaires in this order: PDQ-4+, SCID-IIQ, MAPP, and the sequence was repeated three times.

Analyses were conducted using both categorical and dimensional scores. For analyses using categorical scores, the presence of PDs was determined based on the PDQ-4+, SCID-IIQ, and MAPP according to the categorical diagnostic decisions outlined in the DSM. Scoring procedures for the PDQ-4+ and the SCID-IIQ were adopted directly from Hyler (1994) and Spitzer et al. (1990), respectively.

For the MAPP, each item corresponds directly to a single DSM diagnostic criterion, with the exception of Schizotypal PD criterion six and Narcissistic PD criterion eight. In these instances, there are two questions that correspond to a single DSM criterion. Responses of either “2” (I am often like this) or “3” (I am always like this) within a PD category were considered to indicate a pathological response. If the number of pathological responses within a PD category reached or exceeded the threshold (as a score of four or more would indicate for paranoid PD), the diagnosis was recorded¹.

In addition, test-retest reliability of the three instruments was calculated by analyzing the number of diagnostic criteria that were identified by each of the measures on the first (Time 1) and second (Time 2) administrations, second and third (Time 3) administrations, and first and third administrations. Agreement among the three personality measures was calculated at the criterion level as well as at the diagnostic level using data from one time point (Time 2) and average data. Time 2 data, as opposed to Time 1 data, were analyzed because of concerns that Time 1 estimates for the SCID-IIQ and the MAPP were not accurate estimates of the actual use of tests because most participants completed these measures after completing the PDQ-4 +.

Measures—All of the instruments used in this study were diagnostically based measures consisting of items that are direct translations of the DSM criteria for each PD category.

Personality Diagnostic Questionnaire 4+ (PDQ-4+): This personality measure, which was developed by Hyler (1994), consists of 99 true/false items. The current version of the PDQ-4 + has not been studied extensively, but there is some evidence that it has good screening properties when screening for the presence or absence of a personality disorder (Davison, Leese & Taylor, 2001). Its predecessor, the Personality Diagnostic Questionnaire-Revised (PDQ-R; Hyler & Rieder, 1987), had high sensitivity, moderate specificity, and high negative predictive power when compared with an interview based diagnosis (Hyler et al., 1990; Hyler, Skodol, Oldham, Kellman, & Doidge, 1992).

Structured Clinical Interview for DSM-IV Axis II Personality Disorders Questionnaire (SCID-IIQ): This measure, which consists of 119 yes/no questions, was developed by Spitzer, Williams, Gibbons and First (1990) to be used in conjunction with the SCID interview in the assessment of personality disorders. Previous research has shown that the SCID-IIQ has high rates of false positives (Jacobsberg et al., 1995). However, given the nature of its design as a screening instrument and its low rates of false negative, there is support for its use.

Multi-source Assessment of Personality Pathology (MAPP): This measure, which was developed by Oltmanns et al. (1998), consists of 105 items, including 81 items based on the features of ten personality disorders listed in the DSM-IV, and 24 items based on additional personality traits, most of which are positive. Each item is rated on a 4-point Likert scale, ranging from 0 (I am never like this) to 3 (I am always like this). It has been shown to have high reliability ($r = .54$ to $.74$) when using the median coefficient alpha for each PD criterion (Thomas et al., 2003).

¹ A personality disorder diagnosis requires extensive historical information regarding issues such as the consistency of the behavior across settings, the stability of the behavior over time, and the impact of the behavior on the person's social and occupational adjustment. This information cannot be gleaned easily from a questionnaire. In this paper, the term “diagnosis” is used simply to note that the person endorsed at least enough of the features that would be needed to make a diagnosis for any particular PD category.

Results

Test-retest reliability of the three scales

Table 1 displays the test-retest reliability values for the three instruments. These values were obtained by correlating the number of diagnostic criteria that were identified by each of the measures on two different test days. As is shown in this table, test-retest reliability values based on Time 1 and Time 2 for the PDQ-4+ are lower (median = .60) than that of the SCID-IIQ (median = .84) or the MAPP (median = .83). A paired sample t-test was performed to determine if the test-retest reliability difference between the PDQ-4+ and the other two scales was statistically significant. The t-test showed that differences between the PDQ-4+ and the SCID-IIQ and between the PDQ-4+ and the MAPP were statistically significant $t(9) = -6.76, p < .01$ and $t(9) = -21.78, p < .01$, respectively. Furthermore, the increase in test-retest reliability values is statistically significant only for the PDQ-4+ when the values are calculated using Time 2 and Time 3 data, $t(9) = -17.46, p < .01$.

Table 2 displays the average PD scores for each instrument across time. In general, Time 1 scores were slightly higher than Time 2 scores and Time 3 scores.

Agreement between the three scales

Table 3 shows the level of agreement between the three instruments using scaled scores at one time point (Time 2) and also using average scores. Agreement between any two instruments using scaled Time 2 data was higher than expected. The highest Pearson correlation coefficient of 0.77 was found for Borderline PD between the PDQ-4+ and the SCID-IIQ and the SCID-IIQ and the MAPP. The worst agreement was found for Histrionic PD between the SCID-IIQ and the MAPP ($r = .54$). The level of agreement among these instruments increases slightly when convergence is calculated using average scores. This result was expected because average data are less affected by random errors. The highest level of agreement ($r = .82$) was found for Paranoid PD between the SCID-IIQ and the PDQ-4+, and Avoidant PD between the PDQ-4+ and the MAPP. The lowest level of agreement ($r = .58$) was found for Histrionic PD between the PDQ-4+ and the SCID-IIQ.

Diagnostic agreement between any two instruments based on Time 2 data and average data are reflected in Table 4. These values are lower than those obtained using scaled scores. The highest level of agreement ($r = .61$) was found for Dependent PD between the PDQ-4+ and the SCID-IIQ. The lowest level of agreement was found for Schizotypal PD and Borderline PD between the PDQ-4+ and the MAPP ($r = -0.1$ for both). As with the scaled scores, the level of diagnostic agreement increases when average data are used. The highest level of agreement using average data ($r = .81$) was found for Narcissistic PD using the SCID-IIQ and the MAPP. The lowest level of agreement ($r = 0$) was found for Borderline PD using the PDQ-4+ and the MAPP.

Prevalence and Frequency of DSM PD Diagnoses Based on the Three Scales

Table 5 presents the number of subjects who met or exceeded the diagnostic criteria for the DSM-IV personality disorders based on the three instruments at each assessment occasion and also based on average scores. Both the PDQ-4+ and the SCID-IIQ identified more subjects as meeting criteria for each of the personality disorders compared to the MAPP, with the exception of Obsessive-Compulsive PD. In comparing the PDQ-4+ and the SCID-IIQ, no consistent statements can be made about the over-inclusiveness of one scale over the other. Depending upon the PD that is being measured and on whether diagnostic decisions are based on scores at one time point or average scores, there are cases in which the PDQ-4+ is more inclusive compared to the SCID-IIQ and vice versa.

Average number of diagnoses for each of the three instruments

Average data were used to calculate the proportion of individuals who met diagnostic criteria for one or more PDs based on the three measures. Similar proportions were found for the PDQ-4+ (N = 94; 46.3%) and the SCID-IIQ (N = 109; 53.7%). The MAPP identified only 21 individuals (10.3%) who met diagnostic criteria for one or more PDs. Those who met criteria for at least one personality disorder according to the PDQ-4+, on average, met criteria for 1.7 PDs. Those who met criteria for at least one PD based on the SCID-IIQ met criteria for 1.6 PDs on average. The 21 individuals who met criteria for at least one PD according to the MAPP met criteria for 1.5 PDs on average. In the sample of 203 subjects, the PDQ-4+ identified 111 PDs, the SCID-IIQ identified 125 PDs, and the MAPP identified 76 PDs.

Comparison of DSM PD diagnoses

To compare diagnoses across instruments, average data were used. This comparison was done in three steps. First, we determined how many subjects were identified by each of the scales as meeting diagnostic criteria for each of the nine PDs. For instance, the PDQ-4+ had identified 21 participants who met the diagnostic criteria for paranoid PD. Of the 21 individuals, 10 also met the diagnostic criteria for paranoid PD according to one other scale (either the SCID-IIQ or the MAPP). Of the 10 participants, only 2 met diagnostic criteria for paranoid PD according to all three scales. Refer to Appendix 1 and its caption notes for more details.

In general, convergence between the PDQ-4+ and the MAPP and that between the SCID-IIQ and the MAPP was infrequent (agreement on 26 and 29 cases respectively) compared to the convergence between the PDQ-4+ and the SCID-IIQ (agreement on 92 cases). Additionally, with the exception of Obsessive-Compulsive PD, the MAPP always identified people that were also identified by at least one other measure. In the sample of 203 undergraduate students, 365 PDs were identified. Of the 365 identified PDs, convergence across the three instruments was only observed for 24 cases (6.6%).

Discussion

The primary objective of this research was to examine the degree to which the PDQ-4+, the SCID-IIQ, and the MAPP, three self-report measures of personality pathology, measured the same constructs of personality pathology. This is the first study that we are aware of that has evaluated levels of convergence among instruments based on the average of multiple assessment days. With scaled scores, the levels of agreement between measures based on data from one test day (Time 2) were higher than those reported in past studies. The level of diagnostic agreement based on one test day (Time 2), however, was consistent with those reported in past studies (Hyler et al., 1990; Hyler et al., 1992; Skodol et al., 1991). That is to say, the level of agreement was modest. When data from multiple days were considered, however, diagnostic agreement increased to a more acceptable level, probably because average data are less affected by random errors.

Regarding the prevalence of PDs, this study produced results that are consistent with the literature on the psychometric properties of the PDQ-4+ and the SCID-IIQ. It confirmed that these measures tend to produce prevalence rates that are higher than we would expect based on past interview-based epidemiological studies (Torgersen, Kringlen, & Cramer, 2001). This is expected and perhaps even desirable given that the PDQ-4+ and the SCID-IIQ were designed as screening instruments.

The different ways in which these three personality inventories word their questions has an impact on the prevalence rates that are obtained. In general, the PDQ-4+ and the SCID-IIQ questions are worded in a more positive manner, making them easier to endorse compared to

the MAPP questions. Perhaps as a result of wording the items in a positive manner, some of the questions from the PDQ-4+ and the SCID-IIQ do not accurately reflect the DSM criterion being assessed. The second criterion for histrionic PD mentioned above is a good example of how the positively worded PDQ-4+ and SCID-IIQ (#67, Do you flirt a lot?) questions seem to stray beyond the original meaning of the DSM criterion. Neither the PDQ-4+ nor the SCID-IIQ question is focused narrowly on whether the person may interact with others in an *inappropriately* sexually seductive or provocative manner. The corresponding question on the MAPP (#91) states, “I am inappropriately sexually seductive when interacting with other people.” Unlike the PDQ-4+ or the SCID-IIQ, the wording of the MAPP question is specific, guiding the subject to think about his or her interactions with others in determining whether one would consider them to be inappropriately sexually seductive, which is more consistent with the second criterion for histrionic PD. A close examination of the ways in which the different personality disorder instruments explore the phenomenology of PDs may reveal why there are low convergence rates among these assessment devices.

In comparing the frequencies of PD diagnoses based on the PDQ-4+ and the SCID-IIQ, it is clear that while one measure may be more inclusive for one PD, it may not be for another. For example, the SCID-IIQ identified 22 individuals who met diagnostic criteria for histrionic personality disorder, whereas, the PDQ-4+ only identified 12. For obsessive-compulsive personality disorder (OCPD), however, the PDQ-4+ identified 74, whereas the SCID-IIQ only identified 61.

By comparing diagnoses across instruments, it becomes clear that in addition to disagreeing on the prevalence of these disorders, they are also identifying different groups of individuals, thus suggesting that these instruments are not measuring personality pathology in similar ways. For example, 98 people met diagnostic criteria for OCPD (Figure 1–9) by at least one instrument. The PDQ-4+ identified 74 individuals, of whom 36 did not meet diagnostic criteria according to either of the other scales. The SCID-IIQ identified 61 individuals, of whom 21 did not meet criteria according to either of the other instruments. The MAPP identified 12 individuals, of whom 1 did not meet criteria based on the other instruments. The fact that these instruments identified different groups of people meeting criteria for this PD suggests that these instruments are not interchangeable.

Results of the present study showed that the MAPP was more conservative than the PDQ-4+ and/or the SCID-IIQ. This was expected because both the PDQ-4+ and the SCID-IIQ were written to be screening instruments while the MAPP was designed specifically to compare self descriptions with information obtained from peers using the same descriptions of pathological personality features. Additionally, although the present study demonstrated that level of agreement among the three personality inventories increased when convergence was calculated using repeated assessments and averaging scores over a period of several days, the data clearly indicated that the assessment instruments are not interchangeable. Diagnostic impressions derived from one instrument should not be regarded as the only possible options. Additionally, although some screening instruments have been regarded as being generally “over-inclusive,” results from the present study show that it may not be appropriate to think of any particular self-report instrument as being “over-inclusive” compared to another. Prevalence rates of personality disorders seem to depend largely on the personality disorder being measured and on whether diagnostic decisions are based on data from one testing day or from repeated assessments and average scores.

The current study has three limitations. First and foremost, the sample included only college students. This is of concern for the following reasons: (1) the age of the college student sample is young and therefore results may not generalize to older adults; (2) related to the age of the sample, some college students have not fully developed their sense of selves. Their ‘variable’

sense of selves may affect how they respond to these personality questionnaires that ask about enduring habits and behaviors; (3) most college students are high-functioning individuals and may not be representative of the population in this regard. Therefore, it will be useful to repeat this study in a more heterogeneous sample that includes older adults from the community.

Second, without the administration of diagnostic interviews, behavioral measures, or indices of social adjustment against which to compare results from self-report measures, the concurrent validity of the questionnaires cannot be determined. Such data will be especially important in helping to determine if the MAPP is a useful assessment tool in comparison with other questionnaires.

References

- Bersoff, DM.; Bersoff, DN. Ethical issues in the collection of self-report data. In: Stone, AA., et al., editors. *The science of self-report: Implications for research and practice*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2000. p. 9-24.
- Davison S, Leese M, Taylor PJ. Examination of the screening properties of the Personality Diagnostic Questionnaire 4+ (PDQ-4+) in a prison population. *Journal of Personality Disorders* 2001;15:180–194. [PubMed: 11345853]
- Ekselius L, Lindström E, von Knorring L, Bodlund O, Kullgren G. SCID II interviews and the SCID Screen questionnaire as diagnostic tools for personality disorders in DSM-III–R. *Acta Psychiatrica Scandinavica* 1994;90:120–123. [PubMed: 7976457]
- Greenwald AG, Pickrell JE, Farnham SD. Implicit partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology* 2002;83:367–379. [PubMed: 12150234]
- Hyler, SE. *The Personality Diagnostic Questionnaire 4+*. New York: New York State Psychiatric Institute; 1994.
- Hyler, SE.; Rieder, RO. *PDQ-R: Personality diagnostic questionnaire-revised*. New York: New York State Psychiatric Institute; 1987.
- Hyler SE, Skodol AE, Kellman HD, Oldham JM, Rosnick L. The validity of the Personality Diagnostic Questionnaire: A comparison with two structured interviews. *American Journal of Psychiatry* 1990;147:1043–1048. [PubMed: 2197883]
- Hyler SE, Skodol AE, Oldham JM, Kellman HD, Doidge N. Validity of the Personality Diagnostic Questionnaire--Revised: A replication in an outpatient sample. *Comprehensive Psychiatry* 1992;33(2):73–77. [PubMed: 1544299]
- Jacobsberg L, Perry S, Frances A. Diagnostic agreement between the SCID-II screening questionnaire and the Personality Disorder Examination. *Journal of Personality Assessment* 1995;65:428–433. [PubMed: 8609583]
- Klein DN. Patients' versus informants' reports of personality disorders in predicting 7 ½ year outcome in outpatients with depressive disorders. *Psychological Assessment* 2003;15:216–222. [PubMed: 12847782]
- McDermut, W.; Zimmerman, M. Assessment instruments and standardized evaluation. In: Oldham, JM.; Skodol, AE.; Bender, DS., editors. *Textbook of personality disorders*. Washington, D.C: American Psychiatric Press; 2005.
- Oltmanns, TF.; Turkheimer, E. Perceptions of Self and Others Regarding Pathological Personality Traits. In: Krueger, RF.; Tackett, JL., editors. *Personality and psychopathology*. New York, NY, US: Guilford Press; 2006. p. 71-111.
- Oltmanns TF, Turkheimer E, Strauss ME. Peer assessment of personality traits and pathology in female college students. *Assessment* 1988;5:53–65. [PubMed: 9458342]
- Perry JC. Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry* 1992;149:1645–1653. [PubMed: 1443240]
- Skodol AE, Oldham JM, Rosnick L, Kellman HD, SEH. Diagnosis of DSM-III-R personality disorders: A comparison with two structured interviews. *International Journal of Methods in Psychiatric Research* 1991;1:13–26.

- Spitzer, RL.; Williams, JBW.; Gibbon, M.; First, MB. User's guide for the structured clinical interview for DSM-III-R: SCID. Washington, DC: American Psychiatric Association; 1990.
- Thomas C, Turkheimer E, Oltmanns TF. Factorial structure of personality as evaluated by peers. *Journal of Abnormal Psychology* 2003;112:81–91. [PubMed: 12653416]
- Torgersen S, Kringlen E, Cramer V. The prevalence of personality disorders in a community sample. *Archives of General Psychiatry* 2001;58:590–596. [PubMed: 11386989]
- Westen D. Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of Axis II. *American Journal of Psychiatry* 1997;154:895–903. [PubMed: 9210738]
- Zimmerman M, Coryell WH. Diagnosing personality disorders in the community: A comparison of self-report and interview measures. *Archives of General Psychiatry* 1990;47:527–531. [PubMed: 2350205]

Appendix 1: The Number of DSM-IV PD Diagnoses Based on the Instruments

Appendix 1. Appendix 1 consists of ten Venn diagrams, nine of which presents the number of “diagnoses” based on the instrument(s) for each of the nine PDs (antisocial PD was excluded). The tenth Venn diagram represents the total number of diagnoses for all nine PDs.

Appendix 1-1. PARANOID PD

Appendix 1-2. SCHIZOID PD

Appendix 1-3. SCHIZOTYPAL PD

Appendix 1-4. BORDERLINE PD

Appendix 1-5. HISTRIONIC PD

Appendix 1-6. NARCISSISTIC PD

Appendix 1-7. AVOIDANT PD

Appendix 1-8. DEPENDANT PD

Appendix 1-9. OBSESSIVE-COMPULSIVE PD

Table 1
Test-Retest Reliability of the PDQ-IV, the SCID-IIQ and the MAPP represented as Pearson Correlation coefficients

Personality Disorder	PDQ-IV			SCID-IIQ			MAPP		
	T1 and T2	T2 and T3	T1 and T3	T1 and T2	T2 and T3	T1 and T3	T1 and T2	T2 and T3	T1 and T3
Paranoid (PAR)	0.61	0.84	0.58	0.87	0.86	0.83	0.83	0.86	0.83
Schizoid (SZD)	0.54	0.80	0.59	0.73	0.79	0.73	0.74	0.83	0.77
Schizotypal (SZT)	0.64	0.82	0.63	0.84	0.84	0.80	0.87	0.86	0.81
Antisocial (ANT)	0.56	0.85	0.58	0.87	0.96	0.92	0.78	0.78	0.74
Borderline (BOR)	0.58	0.87	0.47	0.86	0.91	0.85	0.82	0.84	0.79
Histrionic (HIST)	0.65	0.85	0.62	0.86	0.87	0.84	0.82	0.89	0.81
Narcissistic (NARC)	0.65	0.79	0.66	0.84	0.87	0.78	0.84	0.88	0.87
Avoidant (AVD)	0.56	0.81	0.66	0.80	0.86	0.77	0.86	0.89	0.80
Dependent (DEP)	0.60	0.84	0.57	0.66	0.83	0.66	0.83	0.88	0.82
Obsessive-Compulsive (OCPD)	0.59	0.82	0.56	0.82	0.87	0.76	0.76	0.80	0.69
Median	0.60	0.83	0.59	0.84	0.87	0.79	0.83	0.86	0.81

Note. PDQ-IV = Personality Diagnostic Questionnaire; SCID-IIQ = Structured Clinical Interview for DSM-IV Personality Disorders Questionnaire; MAPP = Multi-source Assessment of Personality Pathology. Lowest test-retest reliability coefficient ($r = .45$) was found for Borderline PD as assessed by the PDQ-IV. Highest test-retest reliability coefficient ($r = .90$) was found for Borderline PD as assessed by the SCID-IIQ.

Table 2

Average number of PD criteria endorsed at each administration for all three measures

<i>PDQ</i>	Day 1	Day 2	Day 3
PAR	2	1.5	1.4
SZD	1.5	1	0.9
SZT	2.2	1.5	1.3
ANT	1.4	1.8	1.1
BOR	2	1.3	1.2
HIS	2.5	2.4	2.3
NAR	2.5	1.8	1.6
AVD	2.5	2	1.9
DEP	1.9	1.4	1.3
OC	3.3	3.4	3.3
<i>SCID</i>	Day 1	Day 2	Day 3
PAR	1.6	1.3	1.4
SZD	1.4	1.2	1.2
SZT	1.3	1.2	1.1
ANT	0.4	0.3	0.4
BOR	2.3	2	1.9
HIS	2.7	2.5	2.4
NAR	1.5	1.4	1.4
AVD	2.7	2.5	2.3
DEP	1.8	1.4	1.3
OC	3	3.1	3
<i>MAPP</i>	Day 1	Day 2	Day 3
PAR	0.5	0.5	0.1
SZD	0.7	0.3	0.4
SZT	0.7	0.9	0.5
ANT	0.2	0.3	0.2
BOR	0.5	0.5	0.4
HIS	0.5	0.5	0.5
NAR	0.7	0.5	0.6
AVD	0.8	0.3	0.8
DEP	0.4	0.4	0.3
OC	1.3	1.3	1.5

Note. PDQ-IV = Personality Diagnostic Questionnaire; SCID-IIQ = Structured Clinical Interview for DSM-IV Personality Disorders Questionnaire; MAPP = Multi-source Assessment of Personality Pathology. Numbers in parentheses indicate range of scores.

Table 3

Agreement between PDQ-IV, SCID-IQ, and MAPP scaled scores on DSM-IV Personality Disorders as represented by Pearson correlation coefficients

Personality Disorder	Time 2			Average Scores		
	PDQ-IV vs. SCID-IQ	PDQ-IV vs. MAPP	SCID-IQ vs. MAPP	PDQ-IV vs. SCID-IQ	PDQ-IV vs. MAPP	SCID-IQ vs. MAPP
Paranoid (PAR)	0.74	0.71	0.75	0.82	0.77	0.81
Schizoid (SZD)	0.62	0.55	0.57	0.73	0.70	0.68
Schizotypal (SZT)	0.62	0.66	0.61	0.67	0.73	0.64
Borderline (BOR)	0.77	0.71	0.77	0.78	0.77	0.75
Histrionic (HIST)	0.56	0.69	0.54	0.58	0.69	0.62
Narcissistic (NARC)	0.67	0.65	0.67	0.74	0.76	0.78
Avoidant (AVD)	0.67	0.74	0.75	0.76	0.82	0.78
Dependent (DEP)	0.63	0.65	0.58	0.68	0.70	0.66
Obsessive- Compulsive (OCPD)	0.62	0.62	0.72	0.65	0.65	0.78
Median	0.63	0.66	0.67	0.73	0.73	0.75

Note. PDQ-IV = Personality Diagnostic Questionnaire; SCID-IQ = Structured Clinical Interview for DSM-IV Personality Disorders Questionnaire; MAPP = Multi-source Assessment of Personality Pathology.

* All correlation values were significant at $p \leq .001$

Table 4

Diagnostic agreement between PDQ-IV, SCID-IIQ, and MAPP on DSM-IV Personality Disorders as represented by Pearson correlation coefficients

Personality Disorder	Time 2				Average Scores			
	PDQ-IV vs. SCID-IIQ	PDQ-IV vs. MAPP	SCID-IIQ vs. MAPP	PDQ-IV vs. SCID-IIQ	PDQ-IV vs. MAPP	SCID-IIQ vs. MAPP	PDQ-IV vs. SCID-IIQ	
Paranoid (PAR)	0.48	0.26	0.37	0.51	0.30	0.38	0.38	
Schizoid (SZD)	0.13	0.40	0.37	0.57	0.71	0.40	0.40	
Schizotypal (SZT)	--	-0.01	--	--	0.50	--	--	
Borderline (BOR)	0.45	-0.01	0.14	0.29	0	0.26	0.26	
Histrionic (HIST)	0.43	--	--	0.39	0.28	0.20	0.20	
Narcissistic (NARC)	0.48	0.33	0.70	0.65	0.60	0.81	0.81	
Avoidant (AVD)	0.55	0.35	0.33	0.61	0.40	0.40	0.40	
Dependent (DEP)	0.44	0.40	0.37	0.44	0.44	0.44	0.44	
Obsessive-Compulsive (OCPD)	0.61	0.24	0.37	0.59	0.26	0.39	0.39	
Median	0.45	0.26	0.37	0.51	0.40	0.39	0.39	

Note. PDQ-IV = Personality Diagnostic Questionnaire; SCID-IIQ = Structured Clinical Interview for DSM-IV Personality Disorders Questionnaire; MAPP = Multi-source Assessment of Personality Pathology.

-- = No correlations obtained.

* All correlation values were significant at $p \leq .001$

Table 5
Prevalence of DSM-IV Personality Disorders according to the three diagnostic instruments based on average scores

Personality Disorder	PDQ-IV		SCID-IQ		MAPP	
	Time 2	Average	Time 2	Average	Time 2	Average
Paranoid	26 (12.8)	21 (10.3)	14 (6.9)	13 (6.4)	2 (1.0)	2 (1.0)
Schizoid	6 (3.0)	2 (1.0)	7 (3.4)	6 (3.0)	1 (0.5)	1 (0.5)
Schizotypal	10 (4.9)	8 (3.9)	*	*	1 (0.5)	2 (1.0)
Borderline	6 (3.0)	3 (1.5)	19 (9.4)	14 (6.9)	2 (1.0)	1 (0.5)
Histrionic	25 (12.3)	12 (5.9)	28 (13.8)	22 (10.8)	0 (0)	1 (0.5)
Narcissistic	11 (5.4)	6 (3.0)	6 (3.0)	6 (3.0)	3 (1.5)	4 (2.0)
Avoidant	41 (20.2)	33 (16.3)	45 (22.2)	42 (20.7)	6 (3.0)	8 (3.9)
Dependent	6 (3.0)	5 (2.5)	7 (3.4)	5 (2.5)	1 (0.5)	1 (0.5)
Obsessive- Compulsive	90 (44.3)	74 (36.5)	81 (39.9)	61 (30.0)	16 (7.9)	12 (5.9)
Total	221	161	207	169	32	32

Note. PDQ-IV = Personality Diagnostic Questionnaire; SCID-IQ = Structured Clinical Interview for DSM-IV Personality Disorders Questionnaire; MAPP = Multi-source Assessment of Personality Pathology. Numbers outside of parentheses denote the number of people; numbers in parentheses denote percentage.

* Prevalence of schizotypal PD diagnoses based on the SCID-IQ was not assessed because the SCID-IQ does not include questions corresponding to criteria #4–7 of this PD.