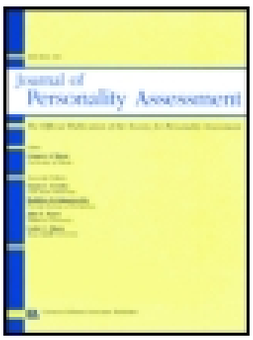

Routledge
Taylor & Francis Group

Journal of Personality Assessment

ISSN: 0022-3891 (Print) 1532-7752 (Online) Journal homepage: http://www.tandfonline.com/loi/hjpa20

Expanding the Validity of the Level of Personality Functioning Scale Observer Report and Self-Report Versions Across Psychodynamic and Interpersonal Paradigms

Michael J. Roche, Nicholas C. Jacobson & Jennifer J. Phillips

To cite this article: Michael J. Roche, Nicholas C. Jacobson & Jennifer J. Phillips (2018): Expanding the Validity of the Level of Personality Functioning Scale Observer Report and Self-Report Versions Across Psychodynamic and Interpersonal Paradigms, Journal of Personality Assessment

To link to this article: https://doi.org/10.1080/00223891.2018.1475394

JOURNAL OF PERSONALITY ASSESSMENT
https://doi.org/10.1080/00223891.2018.1475394

Expanding the Validity of the Level of Personality Functioning Scale Observer Report and Self-Report Versions Across Psychodynamic and Interpersonal Paradigms

Michael J. Roche,[1] Nicholas C. Jacobson,[2] and Jennifer J. Phillips[2]

[1]Department of Psychology, Penn State Altoona; [2]Department of Psychology, The Pennsylvania State University

ABSTRACT

The Level of Personality Functioning Scale (LPFS) operationalizes Criterion A of the *DSM–5* alternative model for personality disorders. Yet, research on this measure has been slow to accumulate and questions remain regarding its reliability and validity. This study examined the LPFS observer-rated (OR) and self-report (SRA) versions of Criterion A in a sample of 240 students who provided psychological life history data and a variety of self-report measures. The results suggested the LPFS OR could be reliably coded, and the LPFS OR and LPFS SRA were significantly associated with outcome variables across psychodynamic and interpersonal paradigms. We discuss the implications of assessing personality dysfunction using the LPFS and the importance of expanding the research base for the AMPD model.

ARTICLE HISTORY
Received 27 October 2017
Revised 13 April 2018

The *Diagnostic and Statistical Manual of Mental Disorders* (5th ed. [*DSM–5*]; American Psychiatric Association, 2013) includes an alternative model for personality disorders (AMPD). In brief, the AMPD defines personality disorder (PD) through criteria that capture the severity of personality dysfunction (Criterion A) as well as individual differences in pathological personality traits (Criterion B). Additionally, a variety of other criteria ensure the PD is inflexible and stable (Criteria C-D), and not better explained by other factors such as another mental disorder, substance use, medical conditions, normal development, or sociocultural environment (Criteria E-G). The AMPD model was placed into Section III of the *DSM–5*, in part, due to it needing more research support. Criterion B arguably has had more sustained research support compared to Criterion A. This could be because the pathological traits defined by Criterion B are similar to other pathological trait models (Personality Psychopathology Five; Harkness, Reynolds, & Lilienfeld, 2014) and normal trait models (Thomas et al., 2013), which means that there are decades of literature forming a research base for examining pathological traits. Additionally, more research has accumulated on Criterion B since the AMPD model was advanced because Criterion B can be captured through self-report and is thus more amenable to large-scale research studies. In contrast, Criterion A was advanced as a clinician-rated measure, with comparably fewer studies examining its reliability, validity, and nomological net. Therefore, research focusing on Criterion A is crucial in developing research support for the AMPD.

Criterion A was advanced to represent the core of what PDs share in common (Bender, Morey, & Skodol, 2011; Waugh et al., 2017). Many categorical PDs have extensive diagnostic comorbidity (Widiger & Trull, 2007), which can be explained if many PDs share a common dysfunctional core. Research has also suggested that a common factor of personality dysfunction explains more variance in current and future functioning compared to diagnostic categories (Hopwood et al., 2011). After careful review of theoretical frameworks and research, Bender et al. (2011) advanced a model where the core of personality dysfunction is represented through self (identity, self-direction) and interpersonal (empathy, intimacy) impairments. Criterion A is operationalized in the *DSM–5* by the Level of Personality Functioning Scale (LPFS; American Psychiatric Association, 2013). Some organize the LPFS into 12 items (Roche, Jacobson, & Pincus, 2016) that describe self-functioning failures in identity (three items) and self-direction (three items), as well as interpersonal functioning failures in empathy (three items) and intimacy (three items). For each item, unique descriptions are given across the five levels of impairment ranging from 0 (*little or no*) to 4 (*extreme*). The LPFS was designed as a clinician-rated measure, although some have used it as a self-report (Roche et al., 2016), finding that the LPFS is associated with self and interpersonal dysfunction in daily life.

Research employing the LPFS as an observer-rated measure

Morey, Bender, and Skodol (2013) asked 337 psychiatrists and psychologists to rate a patient of theirs on the LPFS. They found that the LPFS demonstrated 85% sensitivity and 73% specificity for identifying patients meeting at least one *DSM–IV* PD diagnosis. The LPFS was also related to other markers of life functioning, and frequently incremented the prediction of outcomes over and above *DSM–IV* PD diagnoses.

Few et al. (2013) had graduate student interviewers rate 109 patients on the LPFS after conducting the Structured Clinical Interview for DSM–IV Axis II Personality Disorders (SCID–II; First, Spitzer, Gibbon, Williams, & Benjamin, 1997). Interrater

CONTACT Michael J. Roche mjr386@psu.edu Psychology Department, Penn State Altoona, C128 Smith Building, 3000 Ivyside Park, Altoona, PA 16601.
Supplemental data for this article can be accessed on the publisher's website.

reliability between the original interviewer and a second interviewer (reviewing the videotaped interview) ranged from .47 to .49 for the four aspects of self (identity, self-direction) and interpersonal (empathy, intimacy) impairment. They also found that the LPFS correlated with anxiety, depression, several PD criteria, and that the LPFS was positively correlated (*r*s = .29–.69) with the Criterion B measure of pathological personality traits (Personality Inventory for DSM–5 [PID–5]; American Psychiatric Association, 2013). However, they also found that DSM–IV PDs were not incrementally associated with the LPFS once pathological traits (Criterion B) were included in the model.

Morey (2017) asked 195 student participants to choose an acquaintance to rate along LPFS criteria, and to also rate this target on Five-Factor Model traits using the International Personality Item Pool (Goldberg, 1999). He found that the four components of the LPFS were rated reliably (αs = .86–.92) and again found associations for the five factor traits (*r*s ranging from –.55 to –.77 for four of the five factors, and neuroticism at .79).

Zimmermann et al. (2014) asked 22 students to rate 10 patients who were videotaped while being interviewed using the Operationalized Psychodynamic Diagnosis (OPD) system (Benecke et al., 2009; OPD Task Force, 2008), which is a 1- to 1.5-hr interview covering the topics of symptoms, significant relationships, issues of intimacy, psychosexual development, and childhood experiences. They found that students with little training were able to rate these cases reliably, with the average single Intraclass Correlation Coefficient (ICC) of .51 and the average aggregated Intraclass Correlation Coefficient (ICC) of .96. Furthermore, they found that the LPFS was correlated with PD diagnosis and Axis I disorders. They also examined rater accuracy by comparing student ratings to the scores suggested by experts. They found that students tended to underestimate personality dysfunction by around 1 point (absolute difference), although students showed good rank order consistency where the 10 patients were often correctly ordered from less severe to more severe.

These studies represent a good start toward building a research base for the LPFS and Criterion A. There are also opportunities to expand this research base as it relates to stimuli used for coding, reliability analyses, criterion validity, and the integration of observer and self-report methods. In terms of stimuli to code, some studies relied on previous experience with the person (e.g., rate a patient, rate an acquaintance), whereas others employed a videotaped interview (e.g., SCID–II, OPD). Videotaped interviews likely produce rich clinical information that could be lost using other formats, and yet the complexity involved in coordinating videotaped interviews of clinical patients might represent a barrier to entry for researchers who are interested in contributing to this literature, but do not have the resources needed to conduct that type of study. Reliability was evaluated in many of these studies, perhaps most comprehensively by Zimmermann et al. (2014) in demonstrating that untrained students can reliably use the LPFS rating scale. It will be important to replicate this finding and continue to demonstrate the reliability of the LPFS, as some researchers feel the descriptions in the LPFS are too vague and could place high requirements on raters (Tyrer, 2012).

The criterion validity of the LPFS has been evaluated in predicting PD diagnosis, pathological traits, and normal traits. However, there appear to be few if any studies examining the association with psychodynamic or interpersonal constructs, which would be important as Criterion A is meant to be pantheoretical in capturing information relevant to each of these paradigms (Bender et al., 2011; Pincus, in press).

Finally, there are studies employing both self-report and observer ratings of Criterion B (Few et al., 2013) but no study to date has examined the intersection between self-report and observer-rated scores of the LPFS. This is an important extension, especially because self-report research tends to accumulate at a quicker pace. Additionally, it would be useful to know whether the LPFS self-report and observer report produce similar nomological nets, and whether each LPFS measure can explain unique variance in outcome measures (incremental validity).

This study

This study builds on existing research by examining observer-rated and self-reports of the LPFS in a student sample. The stimuli used to rate LPFS impairments come from a psychological life history that the students complete after reflecting about their life. This is contrasted from videotaped interviews, which likely have richer content, but are more time intensive to collect. Reliability will be assessed similarly to Zimmermann et al. (2014) using a smaller coding team (this research employed three teams of five coders vs. 22 in Zimmermann's research), but a much larger sample of subjects to code (present research codes more than 200 cases, in comparison to 10 cases in Zimmermann's research). Our first hypothesis was that raters would be able to reliably code life history data for LPFS themes.

We next examined the associations among the LPFS observer report (LPFS OR) and LPFS self-report of Criterion A (LPFS SRA). Our second hypothesis was that the LPFS OR and LPFS SRA would be significantly associated, and that the subscales should correlate strongest with their corresponding subscales across methods (e.g., identity subscale for LPFS OR should correlate strongest with identity subscale for LPFS SRA, etc.).

We then examined the LPFS OR and SRA associations with pathological traits (the other major part of the AMPD). We expected positive associations consistent with prior research. We next examined the LPFS OR and SRA across psychodynamic and interpersonal paradigms, comparing the strength of correlations across methods, as well as examining the incremental validity.

The psychodynamic paradigm suggests that self and other impairments are paramount in the understanding of personality dysfunction. For instance, the *Psychodynamic Diagnostic Manual* (*PDM–2;* Lingiardi & McWilliams, 2017) provides a psychodiagnostic chart that summarizes one's level of personality organization from healthy to severe. The components of this include one's view of self (e.g., identity), capacity with others (maintaining satisfying relationships), level of defenses organized from the more immature (e.g., projection, splitting) to mature (suppression, humor), and reality testing (see also transference-focused psychotherapy structural diagnosis; Clarkin, Yeomans, & Kernberg, 2006). The attachment literature is often integrated into psychodynamic conceptualizations

(Crawford et al., 2007), because both perspectives assume that a template of self–other relationship patterns forms based on past experiences and shapes how a person approaches future relationships (termed internal working models in the attachment literature, object relations in the psychodynamic literature). To capture these elements, we employed self-report measures of identity diffusion and psychological defenses. We further used a brief measure of attachment style that describes one's approach to connecting with a romantic partner, and we also asked participants to rate the personality dysfunction of their parents. Consistent with psychodynamic theory, we hypothesized that the LPFS measures would be positively related to identity diffusion, immature defenses, insecure attachment, and viewing their parents as also having personality dysfunction, along with negative associations with mature defenses.

The interpersonal paradigm (Hopwood, Wright, Ansell, & Pincus, 2013) organizes dysfunction around the themes of agency (achievement, power, self-differentiation) and communion (solidarity, connectedness, friendship), with pathology represented as the unsuccessful navigation of the motives to get ahead (agency) and get along (communion). This parallels the focus of self and interpersonal dysfunction seen in the description of Criterion A. Previous research coded Thematic Apperception Test (TAT) stories on the need for connectedness and power and found that the need for connectedness was associated with communal motives, and that the need for power was associated with low communion and high agency motives (Locke, 2000). Thus, there is some research support linking observer-coded stories to this interpersonal measure. We hypothesized the LPFS OR and SRA would be associated with low agency and low communion motives.

Due to the concerns about the complicated LPFS wording, we examined whether participants were able to use the LPFS to accurately rate several vignettes. We further examined whether the tendency to under- or overreport personality dysfunction was correlated with LPFS OR, LPFS SRA, or any of the criterion measures. We hypothesized that participants would be within 1 point of the accurate response but might evince underreporting biases consistent with the findings in Zimmermann et al. (2014).

Method

Participants and procedure

Participants for Samples 1 and 2 were recruited from an introduction to personality course given across multiple semesters. In Sample 1, participants gave consent for their data to be used for research purposes and completed other measures beyond the scope of this article in exchange for $5 compensation (participants could opt out of their data being used for research purposes and still earn full compensation). In Sample 2, participants consented for their data to be used for research purposes in exchange for extra credit (an alternative extra credit assignment provided) but did not have the additional project or financial compensation. In both samples, the instructor of the course was blind to who participated in the study until after the semester ended. All data collection was subject to review and approval by an institutional review board. Participants were asked to complete measures throughout their semester in the course, with the instructors of the course using online surveys as well as in-class assignments for data collection (see Roche, Jacobson, & Roche, 2017 for details). Due to this data collection structure, sample sizes vary slightly across measures. Although 70 students gave initial consent to participate in Sample 1, only 50 completed an online survey where demographic information was collected. These participants were mostly female (80%), college aged ($M = 20.52$, $SD = 0.98$), and White (76%), with smaller portions identifying as Asian (22%), African American (10%), Hispanic or Latino (8%), American Indian (4%), Native Hawaiian or Pacific Islander (4%), and other (4%). Sample 2 included 170 participants, which were mostly female, (81%), college aged ($M = 19.92$, $SD = 1.52$), and White (66%), with smaller portions identifying as Asian (17%), African American (8%), Hispanic or Latino (6%), Native Hawaiian or Pacific Islander (1%), and other (2%). Samples 1 and 2 were combined to simplify the presentation of results and because there were few differences in the strength of correlations across samples.

Measures

The stimulus and coding measures are first described, and then criterion measures organized by paradigm. The criterion measures are exclusively self-report. Descriptive statistics and reliability for measures are reported in Table 1.

Table 1. Descriptive statistics for criterion measures.

Variable	*M*	*SD*	Alpha	*N*
LPFS self-report				
LPFS average	0.86	0.54	0.79	170
LPFS identity (ID)	1.22	0.77	0.75	169
LPFS self-direction (SD)	0.94	0.63	0.59	170
LPFS empathy (EPY)	0.53	0.54	0.62	168
LPFS intimacy (INT)	0.73	0.69	0.79	167
Trait				
PID–5 average	0.86	0.46	0.75	163
Negative affectivity	1.24	0.73	0.77	163
Detachment	0.73	0.58	0.71	163
Antagonism	0.57	0.50	0.67	163
Disinhibition	0.80	0.68	0.83	163
Psychoticism	0.96	0.70	0.77	163
Psychodynamic				
Identity diffusion	2.61	0.60	0.92	238
Mature defenses	4.67	1.12	0.67	237
Neurotic defenses	4.12	1.01	0.60	237
Immature defenses	3.46	0.84	0.77	237
Attachment anxiety	4.12	1.15	0.75	237
Attachment avoidance	2.92	1.13	0.80	237
Rating mother on LPFS	0.72	0.50	0.80	56
Rating father on LPFS	0.85	0.62	0.89	48
Interpersonal				
A+	2.09	0.78	0.52	240
A+, C–	0.88	0.75	0.71	240
C–	1.13	0.80	0.62	240
A–, C–	1.57	0.98	0.75	240
A–	1.63	0.89	0.67	240
A–, C+	2.89	0.85	0.71	240
C+	2.84	0.85	0.74	240
A+, C+	2.92	0.75	0.69	240

Note. LPFS = Level of Personality Functioning Scale; PID–5 = Personality Inventory for DSM–5; A = agency; C = communion. Some criterion measures only available in Sample 1 (rating mother and father on LPFS) or Sample 2 (neuroticism, extraversion, openness, agreeableness, conscientiousness).

Stimulus and coding measures for personality dysfunction

Psychological life history. The self-written psychological life history was guided by an abbreviated version of the Life Stories Interview (LSI; McAdams, 2008). Participants were asked to divide their life into chapters and then describe what the chapters were about. Then, participants selected four key autobiographical life episodes to write about from a list of options (a high point, a low point, a turning point, the earliest memory, an important child or adolescent episode, an important adult episode, a wisdom event, or one other important episode). Next, participants described their biggest challenge or problem faced in life, described one character with the most positive influence in their life, one character with the most negative influence in their life, and then participants looked back for a central theme of their life story. In addition to this, participants in Sample 1 were asked to imagine themselves in 5 years, describing what might change or stay the same and what challenges lie ahead. Participants also described four theories they encountered in class that affected them and wrote about how that theory helps them better understand themselves (list of theories from class in Table S2). This was the information used to code for personality dysfunction using the LPFS.

LPFS OR. The LPFS was published in 2013 by the American Psychiatric Association as a clinician rating measure for personality dysfunction. It can be divided into 12 items (three items per subscale of identity, self-direction, empathy, and intimacy), which combine to form a single score of personality dysfunction along a 0 to 4 scale with specific anchoring descriptions along each point in the scale. The LPFS has been used previously by clinically inexperienced student raters to code for personality functioning after watching video recordings of clinical patients (Zimmermann et al., 2014).

LPFS SRA. The same measure can be used as a self-report by asking the participants to rate their own level of personality dysfunction using the same 12-item scales (see Table S1 for an example). The subscales of identity, self-direction, empathy, and intimacy were averaged to arrive at a single dimension of personality dysfunction. Recently, researchers have begun to create self-report measures of the LPFS that depart from the exact language used in the LPFS published in the DSM–5 (e.g., LPFS SR, Morey, 2017; DLOPFQ, Huprich et al., in press; LPFS-BF, Hutsebaut, Feenstra, & Kamphuis, 2016). As such, we use the label LPFS SRA to represent the LPFS self-report version for Criterion A, which distinguishes this self-report measure (using the LPFS scales verbatim from the DSM–5) from other self-report measures.

Pathological personality trait outcome measure

Pathological personality traits were completed using the PID–5–BF (American Psychiatric Association, 2013). This is a 25-item measure of the five pathological personality traits of negative affectivity, detachment, antagonism, disinhibition, and psychoticism, using a 4-point scale ranging from 0 (*very false or often false*) to 3 (*very true or often true*).

Psychodynamic outcome measures

Identity diffusion was measured using 32 items from the Inventory of Personality Organization (IPO; Lenzenweger, Clarkin, Kernberg, & Foelsch, 2001), which were averaged to represent instability of self and others (Ellison & Levy, 2012). The IPO employs a 5-point scale ranging from 1 (*never true*) to 5 (*always true*).

The Defensiveness Style Questionnaire (Andrews, Singh, & Bond, 1993) is a 40-item self-report measure of psychological defenses using a 9-point scale ranging from 1 (*strongly disagree*) to 9 (*strongly agree*). Two items comprise each scale, and the 20 scales were further averaged to form scales for mature defenses (sublimation, humor, anticipation, suppression), neurotic defenses (undoing, pseudo-altruism, idealization, reaction formation), and immature defenses (projection, passive aggression, acting out, isolation, devaluation, autistic fantasy, denial, displacement, dissociation, splitting, rationalization, and somatization).

Attachment styles were measured using the Experiences in Close Relationships–Short Form inventory (ECR–SF; Wei, Russell, Mallinckrodt, & Vogel, 2007). This 12-item measure assesses for attachment anxiety (e.g., "I need a lot of reassurance that I am loved by my partner") and attachment avoidance (e.g., "I want to get close to my partner, but I keep pulling back") in adult romantic relationships, using a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Participants in Sample 1 rated their parents on the LPFS using the same 12-item scale to form a single score of LPFS severity. However, this rating was completed as an in-class assignment, with students free to choose any family members to rate (although most chose both parents), resulting in a smaller sample size for ratings of mother ($n = 53$) and father ($n = 45$) compared to the total sample.

Interpersonal outcome measure

The Circumplex Scales of Interpersonal Values (CSIV; Locke, 2000) measures interpersonal motives using a 32-item measure with a 5-point scale ranging from 0 (*not important to me*) to 4 (*extremely important to me*). This measure includes eight scales (termed octants), that correlate with each other in a specific pattern that conforms to circumplex properties (Wright, Pincus, Conroy, & Hilsenroth, 2009). Specifically, low to high agency (A) is represented along a bipolar vertical axis, and low to high communion (C) is represented along a bipolar horizontal axis. Of the eight scales, four capture these poles (A+, A–, C–, C+), whereas the other four capture blends of these poles (A+C–, A–C–, A–C+, A+C+).

Participant vignette ratings

Five short written vignettes were created by members of the *DSM–5* work group (Bender, 2012) to represent prototypical cases for some (Level 1), moderate (Level 2), severe (Level 3), and extreme (Level 4) impairment in personality functioning. Participants from Sample 1 rated these vignettes using the 12-item LPFS. First, these 12 items were averaged to obtain a single score for each vignette (five scores per participant). Then, the scores were subtracted from the gold standard expert ratings to create residual scores for each participant (five scores per participant). Finally, an average of these five residual scores was taken for each participant. This score (one per participant, referred to

as rounding) represents to what extent the participant underendorses (i.e., negative number, rounding down) or overendorses (i.e., positive number, rounding up) personality dysfunction.

Data analysis

LPFS coding

Three research assistant teams (five members per team) were given the deidentified psychological life history data of participants to code for LPFS impairment (Team A coded 70 cases all from Sample 1; Teams B and C coded 85 cases each from Sample 2). Team A was given a brief introduction to the *DSM–5* AMPD model, along with reading the Zimmermann et al. (2014) paper, and then participating in a group discussion of LPFS impairment. Then, Team A coded the same five vignettes described earlier, and the first author led a research lab discussion about the principles of coding.[1] From there, Team A coded the 70 cases. Teams B and C were also provided with a brief introduction and were given the LPFS rating scales but were not given the other training experiences.

Reliability

Reliability was evaluated using an ICC calculation across the three teams of coders separately. A two-way random single measures ICC [2,1] represented the reliability of a single rater, whereas a two-way random average measures ICC [2,5] represented the reliability of the mean across the five raters.

Criterion validity

Criterion validity for the LPFS observer report and self-report was evaluated by comparing correlation magnitude strength using Steiger's *Z* test for dependent correlations (Steiger, 1980). Additionally, both LPFS measures were entered simultaneously into a linear multiple regression to examine the variance explained for each of the criterion measures separately.

The interpersonal measure contains eight scales (termed octants) that are designed to be intercorrelated in a circular pattern. This allows for a more sophisticated analysis of correlation patterns with these eight scales. Specifically, the LPFS OR and LPFS SRA were correlated with the eight scales, and then a Structural Summary Method (SSM; Gurtman, 1992; Gurtman & Pincus, 2003) was applied to these patterns of correlations. The SSM decomposes correlation patterns into parameters capturing the overall correlation strength (e.g., average correlation across the eight scales, termed elevation) and the difference between the average correlation and the peak correlation (referred to as the amplitude). Although no clear cutoff exists, amplitudes lower than .1 might suggest that there is little difference in the magnitude between the eight correlations, and in those cases the elevation score is the only score interpreted (Williams & Simms, 2016; Zimmerman & Wright, 2017).

Another parameter is R2, which measures the extent to which the correlation magnitudes conform to a circular pattern. For instance, if a scale is strongly correlated with C+, then it should have similar (although slightly weaker) associations with adjacent scales (e.g., A+C+, A–C+), and negative associations with the opposite scale (C–). R2 values above .7 suggest that the correlation pattern is sufficiently circular.

If the amplitude score suggests the correlations have enough differentiation (differences between stronger and weaker magnitudes), and this pattern of differentiation conforms to a circular pattern, then the interpersonal angle can be interpreted. The interpersonal angle calculates the exact angle (0–360°) where the correlation is strongest. Conceptually, this is akin to examining the eight octants and noting which scale has the highest association, but the interpersonal angle is more precise and accounts for the information contained in all of the correlation patterns simultaneously. As such, this interpersonal measure can capture the severity (elevation) and style (interpersonal angle) of the LPFS measures.

Vignettes

The participant ratings of LPFS for each of the vignettes were subtracted from the gold standard score, calculating a residual where positive scores indicate a participant who overendorsed symptoms in the vignette (rounded up), and a negative score indicating a participant who underendorsed symptoms in the vignette (rounded down). These residual scores were evaluated across vignettes using an analysis of variance to examine whether the severity of a vignette was associated with over- or underendorsement. Additionally, participant rounding scores were correlated with criterion variables.

Results

Can personality dysfunction be coded reliably using psychological life history data?

The reliability results across the three coding teams are presented in Table 2. Considering the average LPFS score, reliability of individual raters ranged from .36 to .58 across the three teams. The reliability averaged across raters ranged from .74 to .88 across the three teams. In each of the three teams, reliability was highest for the identity domain and weakest for the empathy domain. The LPFS is rated along a 0 to 4 scale, and the average level of LPFS impairment in this sample ranged from 0.18 to 0.38, suggesting that there was limited personality dysfunction in this relatively healthy student participant sample. After replacing one outlier value (LPFS OR score of 1.92) with the next highest value in the data (1.23), the data demonstrated adequate values for skewness (LPFS OR = 1.20, LPFS SRA = 0.60) and kurtosis (LPFS OR = 1.02, LPFS SRA = –0.07) to proceed with linear models.

How are LPFS observer reports and self-reports related to each other?

The correlations between observer and self-report are presented in Table 3. The average score for the LPFS OR is significantly correlated with the LPFS SRA, although the magnitude (r = .27) was smaller than anticipated. The columns of Table 3 represent how each LPFS OR subscale correlates with the self-

[1]Team A coded these five vignettes with good reliability (ICC single rater = .84, 95% CI [.55, .98]; ICC average rater = .96, 95% CI [.86, .99]), and the correspondence to the gold standard ratings of these cases was within one point of the expected score (M = –0.71, SD = 0.11).

Table 2. Descriptive statistics and reliability estimates for the LPFS observer ratings.

	Descriptive		ICCs [95% CI]	
LPFS scale	*M*	*SD*	Single	Average
Team A (Sample 1)				
Average	0.18	0.18	0.58 [0.48, 0.69]	0.88 [0.82, 0.92]
ID	0.34	0.34	0.49 [0.38, 0.61]	0.83 [0.76, 0.89]
SD	0.13	0.21	0.44 [0.33, 0.56]	0.80 [0.71, 0.87]
EPY	0.07	0.13	0.26 [0.16, 0.38]	0.64 [0.49, 0.76]
INT	0.19	0.18	0.29 [0.18, 0.41]	0.67 [0.53, 0.78]
Team B (Sample 2a)				
Average	0.36	0.32	0.42 [0.32, 0.53]	0.78 [0.70, 0.85]
ID	0.55	0.42	0.41 [0.31, 0.52]	0.78 [0.70, 0.85]
SD	0.29	0.31	0.29 [0.19, 0.40]	0.67 [0.55, 0.77]
EPY	0.26	0.29	0.23 [0.13, 0.33]	0.59 [0.44, 0.71]
INT	0.37	0.39	0.31 [0.21, 0.42]	0.69 [0.57, 0.78]
Team C (Sample 2b)				
Average	0.38	0.23	0.36 [0.26, 0.47]	0.74 [0.64, 0.82]
ID	0.60	0.36	0.41 [0.31, 0.52]	0.78 [0.69, 0.85]
SD	0.34	0.25	0.21 [0.12, 0.32]	0.57 [0.41, 0.70]
EPY	0.27	0.20	0.14 [0.06, 0.24]	0.44 [0.23, 0.61]
INT	0.27	0.25	0.23 [0.13, 0.33]	0.59 [0.43, 0.71]

Note. LPFS = Level of Personality Functioning Scale; ICC = Intraclass Correlation Coefficient; ID = identity; SD = self-direction; EPY = empathy; INT = intimacy. Sample 1 data included 70 participants. Sample 2 data included 170 participants. Three research teams (5 research assistants each) coded participant content for LPFS impairment (Sample 1 coded by research team A, 85 participants in Sample 2 coded by research team B, 85 participants in Sample 2 coded by research team C).

report subscales. As hypothesized, the strongest correlations were observed between the observer-report subscale and its corresponding self-report subscale for identity ($r = .31$),[2] self-direction ($r = .22$), and intimacy ($r = .30$). In contrast, LPFS OR empathy was not correlated with any of the LPFS SRA subscales.

Criterion validity for pathological traits

The associations with the LPFS OR and SRA are reported in Table 4. As expected, the LPFS OR and SRA were correlated with the average of pathological traits and all five pathological trait domains (although the magnitude of correlation ranged from .2–.5, which is somewhat smaller than the magnitudes observed in Few et al., 2013). The LPFS SRA evinced a stronger correlation than the LPFS OR for detachment, but otherwise the magnitudes were comparable. Entering both LPFS OR and SRA into a multiple regression, both OR and SRA incremented for negative affectivity, disinhibition, and psychoticism, whereas only LPFS SRA was significant for detachment and antagonism.

Criterion validity across the psychodynamic paradigm

Identity diffusion was moderately correlated with the LPFS OR and correlated more strongly with the LPFS SRA, although both LPFS measures incremented this outcome. Mature defenses were negatively correlated with the LPFS OR, and the

[2]Another way to examine this is to compare the correlation between self-report and observer report identity ($r = .31$) to the average of all of the other correlations with identity self-report and identity observer report (*r* average = .15). Doing so, we see higher associations for identity, self-direction ($r = .22$, *r* average = .18), and intimacy ($r = .30$, *r* average = .16), whereas empathy is weaker than its average correlations ($r = .03$, *r* average = .09).

Table 3. LPFS associations among self-reports and observer reports.

		Observer report				
		LPFS average	LPFS ID	LPFS SD	LPFS EPY	LPFS INT
Self-report	LPFS average	.27***				
	LPFS ID		.31***	.17*	.04	.08
	LPFS SD		.30***	.22**	.09	.24**
	LPFS EPY		.08	.11	.03	.06
	LPFS INT		.22**	.19*	.14	.30***

Note. LPFS = Level of Personality Functioning Scale; ID = identity; SD = self-direction; EPY = empathy; INT = intimacy. $N = 167$.
*$p < .05$. **$p < .01$. ***$p < .001$.

LPFS SRA did not increment this association. In contrast, immature defenses were positively correlated with the LPFS SRA, and the LPFS OR did not increment this association. Neurotic defenses were not significantly associated at the zero-order level, but the regression analysis found that the LPFS OR was negatively associated, and the LPFS SRA was positively associated with neurotic defenses (although the variance explained was small). Both LPFS OR and SRA were positively associated with insecure attachment at the zero-order level. In the linear multiple regression analysis, only the LPFS SRA was significantly associated with insecure attachment. Both LPFS OR and SRA were associated with participants rating their mother as having more significant personality dysfunction, and the magnitude was stronger for the LPFS SRA. Only the LPFS SRA was associated with participants rating their father as having more significant personality dysfunction at zero-order and regression analysis.

Criterion validity across the interpersonal paradigm

Recall that the R2 should be above .7 and the amplitude above .1 to meaningfully interpret the circular parameters presented

Table 4. LPFS associations with criterion variables.

Outcome	LPFS OR *r*	LPFS SR *r*	*Z* diff	LPFS OR *b*	LPFS SR *b*	Adj. R^2
Trait						
PID–5 average	.37***	.52***	−1.54	.29***	.44***	.34***
Negative affectivity	.28***	.48***	−1.95	.22**	.42***	.26***
Detachment	.22**	.46***	−2.45*	.10	.43***	.21***
Antagonism	.20*	.25**	−0.49	.14	.22*	.07**
Disinhibition	.33***	.28**	0.44	.34***	.19*	.17***
Psychoticism	.28***	.36***	−0.76	.22*	.30**	.16***
Psychodynamic						
Identity diffusion	.37***	.56***	−2.45*	.22***	.51***	.35***
Mature defenses	−.26***	−.12	−1.45	−.32***	−.05	.11***
Neurotic defenses	−.09	.11	−2.13*	−.21**	.16*	.04*
Immature defenses	.01	.34***	−3.65***	−.14	.38***	.12***
Attachment anxiety	.24***	.40***	−1.84	.12	.36***	.16***
Attachment avoidance	.18**	.25**	−0.77	.13	.22**	.07**
Rating mother–LPFS	.33*	.62***	−2.18*	.09	.59***	.37***
Rating father– LPFS	.24	.43**	−1.12	.07	.40*	.15**

Note. LPFS = Level of Personality Functioning Scale; OR = observer report; SR = self-report; PID–5 = PID–5 = Personality Inventory for DSM–5. Sample sizes vary due to the nature of data collection throughout a semester (see Table 1 for sample size details). *Z* diff is test of differences among dependent correlations with one variable in common using Steiger *Z* calculation (Steiger, 1980). LPFS SR and OR regressed onto criterion variables, with standardized betas reported and adjusted R^2 values.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5. LPFS associations with interpersonal criterion variables.

Scale	Elevation	A	C	Amp	Angle (°)	R2
Observer report						
LPFS average	.00	−.02	−.05	.06	210	.54
LPFS ID	.06	< .01	.02	.02	353	.03
LPFS SD	< .01	−.02	−.06	.06	202	.70
LPFS EPY	−.04	−.01	−.02	.02	200	.16
LPFS INT	< .01	< .01	−.13	.13	179	.81
Self-report						
LPFS average	.13	−.01	−.30	.30	182	.97
LPFS ID	.15	−.10	−.11	.15	222	.94
LPFS SD	.12	.02	−.23	.23	175	.96
LPFS EPY	.08	.04	−.25	.25	171	.95
LPFS INT	.06	.03	−.34	.34	175	.98

Note. LPFS = Level of Personality Functioning Scale; A = agentic dimension; C = communal dimension; ID = identity; SD = self-direction; EPY = empathy; INT = intimacy. Observer report $N = 240$. Self-report $N = 170$. Elevation refers to the average correlation across interpersonal scales. Amp = amplitude. Angle refers to angular displacement in degrees (out of 360°), with 0°/360° = C+, 90° = A+, 180° = C–, 270° = A–, and blends of these poles represented through different degrees around the circle. R2 refers to prototypicality in scale correlations conforming to a circular pattern.

in Table 5. The LPFS OR had an R2 of .54 and amplitude of .06, meaning that the LPFS OR correlations with the interpersonal motives scales do not form a prototypical circular pattern. Moreover, the elevation was zero, which means that the LPFS OR did not have strong associations with interpersonal motives generally. To further inspect this result, we used the LPFS OR subscales to determine whether there are some prototypical patterns obfuscated by the total score. We found that LPFS OR self-direction had a prototypical circular profile (R2 = .70), suggesting the theme of motives were low communion and low agency (angle = 202, agency score = –0.02, communion score = –0.06), although the amplitude was weak (amplitude = 0.06). The LPFS OR intimacy had a prototypical circular profile (R2 = .81) and good differentiation (amplitude = .13), suggesting the theme of motives were low communion (angle = 179, communion score = –0.13).

In contrast, the LPFS SRA average score had excellent prototypicality (R2 = .97) and differentiation (amplitude = .30), with the theme of motives as low communion (angle = 182, communion score = –0.30). All subscales also had excellent R2 values, acceptable differentiation, and self-direction, empathy, and intimacy all fell within 5° of each other along the theme of low communion. The identity subscale had a blend of low communion and low agency (angle = 222, agentic score = –0.10, communal score = –0.11).

Can participants accurately rate vignettes using the LPFS?

Participant ratings of five vignettes are presented in Table 6. For ease of interpretation, Table 6 organizes the cases in order of severity rather than the order in which participants rated the cases. As hypothesized, across the five cases the tendency was to round down (underreport) personality dysfunction (M = –0.55, SD = 0.37). This magnitude is in line with what Zimmermann had reported. There was a significant effect for the severity of case, where participants were more likely to round down with extreme (4) personality dysfunction (M = –1.82), and severe (3) personality dysfunction (M = –0.90) compared to the less severe cases.

Table 6. Participant ratings of prototypical personality dysfunction cases.

Vignette	*M*	*SD*	Gold standard	Residual *M*	Residual *SD*
Case 4	1.33	0.66	1	0.33[d]	0.66
Case 2	2.05	0.52	2	0.05[d]	0.52
Case 3	1.56	0.52	2	−0.44[c]	0.52
Case 1	2.10	0.62	3	−0.90[b]	0.62
Case 5	2.18	0.67	4	−1.82[a]	0.67

Note. $N = 46$. Residual scores are calculated by subtracting the gold standard score from each participant's score. There was a significant effect of case on residual scores at the $p < .05$ level for the five conditions, $F(4, 225) = 91.13$, $p < .001$. Significant Bonferroni post-hoc contrasts are noted in subscript letters a through d. Participants were presented with these cases in sequential order from Case 1 to Case 5.

Across the five vignettes, a residual score (referred to as rounding) was calculated for each participant such that a positive score reflected overendorsement (rounding up) and a negative score reflected underendorsement (rounding down) of personality dysfunction. This variable was correlated with the available criterion measures from Sample 1 (the vignette coding was only completed in Sample 1). The tendency to round up (overendorse) personality dysfunction in the vignettes was not associated with the LPFS OR (r =.22, p = .144), but was associated with the LPFS SRA (r = .42, p = .004). Rounding up was also significantly correlated with psychoticism (r = .37, p = .014), identity diffusion (r = .47, p = .001), and anxious attachment (r = .41, p = .006), and although it was correlated with interpersonal motive elevation (r = .41, p = .005), the unacceptable R2 (.58) and amplitude (.09) values suggested that it did not conform to a circular theme. To examine whether this tendency to round up or down explained the association between the LPFS SRA and criterion variables, multiple linear regression analyses were examined with the LPFS SRA and rounding variable entered together for the outcome variables that had a zero-order significant correlation with the rounding variable. The LPFS SRA remained significant in all cases.

Discussion

This research examined the validity of the LPFS as an observer-rated measure and self-report measure. Results can be organized into the themes of reliability, observer and self-report integration, criterion validity, and over- and underendorsement.

Reliability

The reliability of a single rater ranged in this study from .36 to .58. In the Zimmermann et al. (2014) research, their reliability was higher at .51 (although the confidence interval ranged from .31–.78). Several factors could account for this discrepancy. First, the stimulus provided was richer in the Zimmermann study, with videos conducted by trained experts for over an hour. Second, subjects were clinical patients who were likely to express more personality dysfunction. In contrast, the life history data were autobiographical, which did not allow for follow-up or probing, and the LPFS means suggested that this participant sample was in the relatively healthy range of personality dysfunction, leading to a range restriction that might have made discriminating cases more difficult. Although

there were not large differences between the teams of coders, Team A was given additional training experiences (e.g., coding vignettes and a discussion), which might have marginally improved their coding (Team A ICC was .58, which is slightly higher than Zimmermanns et al.'s [2014] ICC of .51). Nevertheless, these results suggest caution in using the LPFS with a single rater.

The aggregated ICC scores ranged from .74 to .88, suggesting that the LPFS appears to be a sufficiently reliable measure when aggregating across raters for research purposes. Zimmermann et al. (2014) had a higher ICC of .96, but this might be due to aggregating 22 raters (vs. five). Similar to before, it might also reflect the higher quality stimulus to code compared to the life history. Examining the subscales, identity had the strongest ICCs and empathy had the weakest. It could be that the life history information spoke more directly toward identity functioning, which made that aspect easier to code in this sample. Empathy was also the weakest ICC by far in the Zimmermann et al. (2014) research, which could suggest the stimuli used in both studies are not doing a good job at evoking empathy content, or that it might be a weakness in how the LPFS operationalizes empathy impairment. Future research should gain a better understanding of what LPFS empathy is capturing by examining how LPFS empathy is related to a variety of other empathy measures (ideally across self-report and observational methods).

OR and SRA integration

The LPFS OR and SRA do not appear to be redundant, as their correlation is smaller than .3. There was evidence for convergence across methods, as subscales on the LPFS OR correlated with the corresponding subscales on the LPFS SRA (except for empathy, again highlighting a potential weakness of that subscale). The LPFS SRA had a stronger magnitude of association with several outcomes (detachment, identity diffusion, immature defenses, rating mother), likely due to the shared method variance between the LPFS SRA and the criterion measures that were also self-report. Despite this, several variables exhibited incremental validity where both the LPFS SRA and OR contributed uniquely to explain variance. Thus, it appears both LPFS OR and SRA provide important information. Future research should include criterion measures beyond self-report to examine the incremental validity of the LPFS OR and SRA. For instance, Sample 1 included data on total course grade for the participants ($n = 70$). This outcome is not self-report data and was significantly associated with the LPFS OR ($r = -.28$, $p = .017$), but not the LPFS SRA ($r = -.11$, $p = .353$).

It would also be useful to more closely examine discontinuities between self-reported and observer-reported scores (e.g., an individual perceives himself or herself as much lower or higher on personality dysfunction compared to the observer score), as these discontinuities can produce meaningful patterns of association (e.g., Bornstein, 2002). Other research examining self- and peer reports of PDs finds that both methods can be predictive of important life outcomes (Fiedler, Oltmanns, & Turkheimer, 2004), further highlighting the importance of integration across methods.

LPFS and pathological traits

The LPFS OR and SRA correlated in the expected directions with pathological traits. For the most part the LPFS SRA was a stronger predictor, except for the dimension of disinhibition, where the standardized beta was stronger for the LPFS OR. This might suggest that coding methods can pick up on disinhibition more efficiently than a self-report, which is consistent with the self–other knowledge asymmetry model that suggests observers can be more effective at detecting traits high in evaluativeness (Vazire, 2010).

LPFS and the psychodynamic paradigm

This is among the first studies to empirically link the LPFS with psychodynamic constructs, and several core constructs evidence strong associations to both the LPFS OR and SRA. Identity diffusion is a core concept in psychodynamic thought, and this study found it significantly associated with both LPFS OR and SRA.

In psychodynamic theory, the level of personality dysfunction can be characterized by both the presence of immature defenses and the absence of mature defenses (Clarkin et al., 2006). Interestingly, the LPFS OR picked up the absence of mature defenses by uniquely finding a negative association with mature defenses. The LPFS SRA, though, picked up on the presence of immature defenses by finding a unique positive association. This might again suggest that both LPFS OR and SRA provide unique information that is useful to integrate. Anxious and avoidant attachment styles are concepts linked to early self–other relationship formations that are often integrated into psychodynamic thinking. Again, both the LPFS OR and SRA were associated with these forms of insecure attachment (although the LPFS SRA was stronger in magnitude). Finally, psychodynamic theory would suggest a person's personality functioning is rooted in experiences with caregivers. This research found support for this, as the LPFS OR and SRA scales were significantly associated with viewing the mother (and less so the father) as having personality dysfunction.

LPFS and the interpersonal paradigm

We hypothesized that the LPFS would be associated with low agency and low communion. This was partially supported, as the LPFS SRA average score and three of the four subscales had themes of low communion (but not low agency in particular). The exception was for identity impairment, which did have aspects of low agency and low communion. Presumably, the subscales of identity and self-direction have the strongest link to low agency, so it makes sense that the identity subscale relates in that way. Consistent with this, the LPFS OR subscale of self-direction had themes of low agency and low communion, whereas the subscale of intimacy impairment had themes of low communion. The total score of the LPFS demonstrated a prototypical theme for the LPFS SRA but not the LPFS OR. Furthermore, in the regression analyses the LPFS SRA was uniquely associated with traits that have strong interpersonal themes (detachment, antagonism). Taken together, it might suggest that the LPFS OR is capturing less interpersonal dysfunction.

The LPFS and over- or underendorsement

Participants appeared able to use the LPFS scale to rate others on personality dysfunction, suggesting that it is plausible for them to use the LPFS to rate themselves. However, there was a tendency to "round down" away from personality dysfunction, particularly as the pathology was more extreme. This could have implications for how participants completed the LPFS SRA, which might contribute to range restriction.

Implications for the LPFS SRA

The LPFS SRA emerged as a possible alternative to the more time-intensive coding for LPFS OR data. The LPFS SRA was almost always more strongly correlated with the criterion measures (perhaps explained by shared method variance), but the pattern of criterion validity was similar across the LPFS SRA and OR. The LPFS SRA and OR appear to have similar functional relationships with most of the criterion measures (although there were differences across some outcomes and the correlation among LPFS measures was small). The LPFS SRA also demonstrated a more prototypical interpersonal profile compared to the LPFS OR. Furthermore, participants were able to accurately identify vignettes across different levels of personality dysfunction, suggesting that participants should be able to use the scale effectively on themselves (although there might be a tendency to underendorse symptoms). Other measures of the LPFS have emerged in recent years, employing self-report (Huprich et al., in press; Hutsebaut et al., 2016; Morey, 2017) and semistructured interviews (Hutsebaut, Kamphuis, Feenstra, Weekers, & De Saeger, 2017), and future research should examine how these measures compare to each other.

Implications for the LPFS OR

This research provides evidence for the reliability and validity of the LPFS as a measure of personality dysfunction. The use of the LPFS aggregated across raters seems to be viable with little training for research assistant coders. However, it is possible more thorough training would be needed if the goal is to use individual coders reliably. The life history data appeared to be an appropriate stimulus to evoke personality dysfunction, although perhaps more targeted questions on empathy and interpersonal dysfunction would be useful. This research used self-report criterion outcomes, and future research should expand this to include other life outcomes where the LPFS OR might be more predictive.

The level of personality functioning is a construct, and this construct was assessed using self-report and observer report methods. Each carries strengths and weaknesses. Importantly, the sample was in the healthier range of personality, and it is unclear whether these association patterns would hold in a more pathological range (e.g., is self-report less accurate the more personality disordered the individual becomes?).

Limitations and future directions

This research has several strengths, including a large sample of data coded for LPFS impairment, the examination of observer report and self-report concurrently, evaluating criterion validity across multiple paradigms, the use of a new stimulus (life history data) to more efficiently capture material to code for personality dysfunction, and the examination of participant accuracy in employing the LPFS to case vignettes. This research also has several limitations, which should be recognized. The participants were drawn from a healthy and demographically homogenous student sample, which potentially limits the generalizability into clinical samples and diverse populations and could contribute to range restriction curtailing the strength of associations. Due to the nature of data collection throughout the semester in a personality course taught by different instructors, there were missing data related to demographics. Several short form versions of personality measures were used and it is possible a fuller picture would have been obtained by employing the full measures. Future research could also code the life history data for other themes (e.g., agency and communion, object relations) to provide multimethod data for criterion outcomes. Despite these limitations, this research takes an important step forward in identifying the validity and reliability of the LPFS and Criterion A of the AMPD. Future research should continue to build on this research base so that the *DSM–5* AMPD can be validated and ultimately moved into the official diagnostic system of the *DSM*.

Funding

Funding was provided by The Pennsylvania State University College of Liberal Arts, Superior Teaching and Research Award (2014).

References

American Psychiatric Association. (2013). *The diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Press.

Andrews, G., Singh, M., & Bond, M. (1993). The defense style questionnaire. *The Journal of Nervous and Mental Disease*, *181*, 246–256. doi:10.1097/00005053-199304000-00006

Bender, D. S. (2012). *Level of personality functioning scale: Training case vignettes*. Unpublished manuscript, Department of Psychiatry, University of Arizona College of Medicine, Tucson, AZ.

Bender, D. S., Morey, L. C., & Skodol, A. E. (2011). Toward a model for assessing level of personality functioning in *DSM-5*, part I: A review of theory and methods. *Journal of Personality Assessment*, *93*, 332–346. doi:10.1080/00223891.2011.583808

Benecke, C., Koschier, A., Peham, D., Bock, A., Dahlbender, R. W., Biebl, W., & Doering, S. (2009). Erste ergebnisse zu reliabilität und validität der OPD-2 strukturachse [First results on the reliability and validity of the OPD-2 axis structure]. *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, *55*, 84–96. doi:10.13109/zptm.2009.55.1.84

Bornstein, R. F. (2002). A process dissociation approach to objective-projective test score interrelationships. *Journal of Personality Assessment*, *78*, 47–68. doi:10.1207/S15327752JPA7801_04

Clarkin, J. F., Yeomans, F. E., & Kernberg, O. F. (2006). *Psychotherapy for borderline personality: Focusing on object relations*. Washington, DC: American Psychiatric.

Crawford, T. N., Livesley, W. J., Jang, K. L., Shaver, P. R., Cohen, P., & Ganiban, J. (2007). Insecure attachment and personality disorder: A twin study of adults. *European Journal of Personality*, *21*, 191–208. doi:10.1002/per.602

Ellison, W. D., & Levy, K. N. (2012). Factor structure of the primary scales of the inventory of personality organization in a nonclinical sample using exploratory structural equation modeling. *Psychological Assessment*, *24*, 503–517. doi:10.1037/a0026264

Few, L. R., Miller, J. D., Rothbaum, A. O., Meller, S., Maples, J., Terry, D. P., … MacKillop, J. (2013). Examination of the Section III DSM-5 diagnostic system for personality disorders in an outpatient clinical sample. *Journal of Abnormal Psychology*, *122*, 1057–1069. doi:10.1037/a0034878

Fiedler, E. R., Oltmanns, T. F., & Turkheimer, E. (2004). Traits associated with personality disorders and adjustment to military life: Predictive validity of self and peer reports. *Military medicine*, *169*, 207–211. doi:10.7205/MILMED.169.3.207

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured clinical interview for DSM-IV Axis II disorders (SCID-II)*. Washington, DC: American Psychiatric Press.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, *7*, 7–28.

Gurtman, M. B. (1992). Construct validity of interpersonal personality measures: The interpersonal circumplex as a nomological net. *Journal of Personality and Social Psychology*, *63*, 105–118. doi:10.1037/0022-3514.63.1.105

Gurtman, M. B., & Pincus, A. L. (2003). The circumplex model: Methods and research applications. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 407–428). Hoboken, NJ: John Wiley & Sons.

Harkness, A. R., Reynolds, S. M., & Lilienfeld, S. O. (2014). A review of systems for psychology and psychiatry: Adaptive systems personality psychopathology five (PSY-5), and the *DSM*-5. *Journal of Personality Assessment*, *96*, 121–139. doi:10.1080/00223891.2013.823438

Hopwood, C. J., Malone, J. C., Ansell, E. B., Sanislow, C. A., Grilo, C. M., … Gunderson, J. G. (2011). Personality assessment in DSM-5: Empirical support for rating severity, style, and traits. *Journal or Personality Disorders*, *25*, 305–320. doi:10.1521/pedi.2011.25.3.305

Hopwood, C. J., Wright, A. G., Ansell, E. B., & Pincus, A. L. (2013). The interpersonal core of personality pathology. *Journal of Personality Disorders*, *27*, 270–295. doi:10.1521/pedi.2013.27.3.270

Hutsebaut, J., Feenstra, D. J., & Kamphuis, J. H. (2016). Development and preliminary psychometric evaluation of a brief self-report questionnaire for the assessment of the DSM–5 level of personality functioning scale: The LPFS brief form (LPFS-BF). *Personality Disorders: Theory, Research, and Treatment*, *7*(2), 192. doi:10.1037/per0000159

Huprich, S. K., Nelson, S. M., Meehan, K. B., Siefert, C. J., Haggerty, G., Sexton, J., … Baade, L. (in press). Introduction of the DSM-5 levels of personality functioning Questionnaire. *Personality Disorders: Theory, Research, and Treatment*.

Hutsebaut, J., Kamphuis, J. H., Feenstra, D. J., Weekers, L. C., & De Saeger, H. (2017). Assessing *DSM-5*-oriented level of personality functioning: Development and psychometric evaluation of the semi-structured interview for personality functioning *DSM-5* (STiP-5.1). *Personality Disorders: Theory, Research, and Treatment*, *8*, 94–101. doi:10.1037/per0000197

Lenzenweger, M. F., Clarkin, J. F., Kernberg, O. F., & Foelsch, P. A. (2001). The inventory of personality organization: Psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis, proneness, and self-domains in a non-clinical sample. *Psychological Assessment*, *13*, 577–591. doi:10.1037/1040-3590.13.4.577

Lingiardi, V., & McWilliams, N. (Eds.). (2017). *Psychodynamic diagnostic manual: PDM-2*. New York, NY: Guilford Publications.

Locke, K. D. (2000). Circumplex scales of interpersonal values: Reliability, validity, and applicability to interpersonal problems and personality disorders. *Journal of Personality Assessment*, *75*, 249–267. https://doi.org/10.1207/S15327752JPA7502_6. doi:10.1207/S15327752JPA7502_6

McAdams, D. P. (2008). *The life story interview*. Retrieved from http://www.sesp.northwestern.edu/docs/LifeStoryInterview.pdf

Morey, L. C., Bender, D. S., & Skodol, A. E. (2013). Validating the proposed *diagnostic and statistical manual of mental disorders, 5th edition*, severity indicator for personality disorder. *The Journal of Nervous and Mental Disease*, *201*, 729–735. doi:10.1097/NMD.0b013e3182a20ea8

Morey, L. C. (2017). Application of the *DSM-5* level of personality functioning scale by lay raters. *Journal of Personality Disorders*, *31*, 1–12. doi:10.1521/pedi_2017_31_305

Pincus, A. L. (in press). An interpersonal perspective on criterion a of the DSM-5 alternative model for personality disorders. *Contemporary Opinion in Psychology*.

Roche, M. J., Jacobson, N. C., & Pincus, A. L. (2016). Using repeated daily assessments to uncover oscillating patterns and temporally-dynamic triggers in structures of psychopathology: Applications to the *DSM-5* alternative model of personality disorders. *Journal of Abnormal Psychology*, *125*, 1090–1102. doi:10.1037/abn0000177

Roche, M.J., Jacobson, N.C., & Roche, C.A. (2017). Teaching strategies for personality assessment at the undergraduate level. *Journal of Personality Assessment*, *99*, 117–125.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251. doi:10.1037/0033-2909.87.2.245

Task Force, OPD (Eds.). (2008). *Operationalized Psychodynamic Diagnosis OPD-2: Manual of diagnosis and treatment planning*. Cambridge, MA: Hogrefe & Huber.

Thomas, K. M., Yalch, M. M., Krueger, R. F., Wright, A. G., Markon, K. E., & Hopwood, C. J. (2013). The convergent structure of DSM-5 personality trait facets and five-factor model trait domains. *Assessment*, *20*, 308–311. doi:10.1177/1073191112457589

Tyrer, P. (2012). Diagnostic and statistical manual of mental disorders: A classification of personality disorders that has had its day. *Clinical Psychology and Psychotherapy*, *19*, 372–374. doi:10.1002/cpp.1810

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*, 281–300. doi:10.1037/a0017908

Waugh, M. H., Hopwood, C. J., Krueger, R. F., Morey, L. C., Pincus, A. L., & Wright, A. G. (2017). Psychological assessment with the DSM–5 Alternative model for personality disorders: Tradition and innovation. *Professional Psychology: Research and Practice*, *48*, 79. doi:10.1037/pro0000071

Wei, M., Russell, D. W., Mallinckrodt, B., & Vogel, D. L. (2007). The experiences in close relationship scale (ECR)-short form: Reliability, validity, and factor structure. *Journal of Personality Assessment*, *88*, 187–204 https://doi.org/10.1080/00223890701268041. doi:10.1080/00223890701268041

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist*, *62*, 71–83. doi:10.1037/0003-066X.62.2.71

Williams, T. F., & Simms, L. J. (2016). Personality disorder models and their coverage of interpersonal problems. *Personality Disorders: Theory, Research, and Treatment*, *7*, 3–23. doi:10.1037/per0000140

Wright, A. G., Pincus, A. L., Conroy, D. E., & Hilsenroth, M. J. (2009). Integrating methods to optimize circumplex description and comparison of groups. *Journal of Personality Assessment*, *91*, 311–322. doi:10.1080/00223890902935696

Zimmermann, J., Benecke, C., Bender, D. S., Skodol, A. E., Schauenburg, H., Cierpka, M., & Leising, D. (2014). Assessing *DSM-5* level of personality functioning from videotaped clinical interviews: A pilot study with untrained and clinically inexperienced students. *Journal of Personality Assessment*, *96*, 397–409. doi:10.1080/00223891.2013.852563

Zimmermann, J., & Wright, A. G. (2017). Beyond description in interpersonal construct validation: Methodological advances in the circumplex structural summary approach. *Assessment*, *24*, 3–23. doi:10.1177/1073191115621795