


## Research

# A new performance-based measure of personality functioning impairment: development and preliminary evaluation of reliability and validity

Adam P. Natoli<sup>1</sup>  · Chloe M. Rodriguez<sup>1</sup> 

Received: 5 September 2023 / Accepted: 31 January 2024

Published online: 22 February 2024

© The Author(s) 2024 

## Abstract

Personality functioning impairment is at the center of many dimensional models of personality. Available measures of personality functioning impairment are limited to self-report, clinician-/informant-rated, and interview methods. Although researchers have begun investigating established performance-based instruments' potential for assessing personality functioning impairment, administration and scoring of these instruments is complex and the latent variables they measure diverge from personality functioning impairment as described in the ICD-11 and the Alternative Model for Personality Disorders (AMPD) of the DSM. We address this absence by developing and psychometrically evaluating the Level of Personality Functioning Scale-Questionnaire-based Implicit Association Test (LPFS-qiAT). The LPFS-qiAT's psychometric properties were evaluated across four studies, producing initial evidence supporting the new instrument's reliability as well as its convergent, discriminant, and criterion-related validity. As the first performance-based measure of personality functioning impairment that aligns with the AMPD and, to a degree the ICD-11, that is easily administered, scored, and interpreted, the LPFS-qiAT shows potential to become a valuable tool in both research and clinical practice.

**Keywords** Personality functioning impairment · AMPD · Implicit association test · Psychometrics

The global prevalence rate of personality disorders is fairly high (ranging from 7.8% to 12.16%), with personality disorders showing links to high health service utilization [1], economic burden [2, 3], medical morbidity and mortality [4], double the odds of poor treatment outcome [5], multiple public health issues [6, 7], and societal costs [8]. Although some mental health professionals and researchers still use a categorical model of personality pathology, recent shifts in diagnostic systems and the growing body of evidence supporting the scientific [9] and practical [10] superiority of dimensional models will inevitably lead to the adoption of the latter. Most dimensional models conceptualize personality in terms of functioning and style and consider personality functioning impairment the core of personality pathology. Consequently, there will be an increased need to assess personality functioning impairment accurately and easily as the field continues to shift toward the use of dimensional models. Numerous measures of personality functioning impairment have been developed (see [9, 11]). Available instruments overwhelmingly use self-report, clinician-/informant-rating, or interview methods of measurement. Some researchers have begun investigating different performance-based

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s44192-024-00059-4>.

✉ Adam P. Natoli, apn017@shsu.edu; Chloe M. Rodriguez, cmr082@shsu.edu | <sup>1</sup>Department of Psychology and Philosophy, Sam Houston State University, Huntsville, TX, USA.



instruments' potential for assessing personality functioning impairment [12]; however, the administration and scoring of these instruments is habitually complex, time intensive, and the latent variable(s) measured diverge from personality functioning impairment as described in the chapter on personality disorders and related traits in the 11th edition of the *International Classification of Diseases (ICD-11)* [13] and in the *Diagnostic and Statistical Manual of Mental Disorders' (DSM) Alternative Model for Personality Disorders (AMPD)* [14, 15]. To address the prominent absence of an easily administered and scored performance-based measure of personality functioning impairment that aligns with these diagnostic systems, we developed a novel instrument that meets these criteria and directly parallels the AMPD's Level of Personality Functioning Scale (LPFS) [14]. The new instrument's internal and retest reliability and several aspects of validity were then evaluated through four separate studies.

## 1 Background

Over the last decade, classification systems of personality disorders have refined their diagnostic frameworks by introducing dimensional models intended to address the many limitations of longstanding categorical models [16–19]. Perhaps the clearest examples of this shift are the ICD-11's [13] removal of all personality disorder categories and the introduction of the AMPD in DSM-5 [14]. Consideration of severity—level of personality functioning impairment—is essential to any dimensional model of personality [20, 21] and research has revealed generalized severity of functioning to be one of the most robust and important single predictors of personality pathology [22]. Accordingly, the dimensional models of the ICD-11 and DSM both emphasize the importance of assessing personality functioning impairment. Yet, measures of personality functioning are informative in matters beyond diagnosis; personality functioning spans a continuum from healthy, adaptive functioning to severe impairment and its measurement is relevant whenever one wishes to understand an individual's psychology. Fortunately, numerous instruments have been developed for measuring personality functioning.

### 1.1 Measures of personality functioning impairment

Most measures of personality functioning impairment make use of self-report methods (see [9, 11, 23, 24] for general and psychometric reviews). The collection of self-report measures include the Levels of Personality Functioning-Brief Form 2.0 (LPFA-BF 2.0) [25], Levels of Personality Functioning Scale-Self-Report (LPFS-SR) [26], DSM-5 Levels of Personality Functioning Questionnaire (DLOPFQ) [27] and a corresponding short-form (DLOPFQ-SF) [28], Self and Interpersonal Functioning Scale (SIFS) [29], Levels of Personality Functioning Scale-Self-Report of Criterion A (LPFS-SRA) [30], the Personality Disorder Severity ICD-11 scale (PDS-ICD-11) [31], and Levels of Personality Functioning Questionnaire for Adolescents from 12 to 18 years (LoPF-Q 12–18) [32]. These instruments range in length (from 12 to 132 items), in number of scales, and in the amount of psychometric evidence supporting their use. Largely, psychometric investigations of self-report measures of personality functioning impairment have supported the validity, internal consistency, factor structure, and measurement invariance of these instruments in several diverse populations and for an assortment of external criteria. To the authors' knowledge, the only<sup>1</sup> study to date to investigate the retest reliability of a self-report measure of personality functioning impairment demonstrated good to excellent reliability of the LPFS-SR over approximately 2 weeks ( $r_s = 0.81–0.91$ ) [33].

Complementing self-report tests of personality functioning impairment are a small group of (semi-)structured interviews and informant- and clinician-rated measures (see [9, 11] for reviews). Many of the above self-report instruments were based on the original LPFS [14], which was introduced in the AMPD as a clinician-rated scale designed to help assess personality disorder severity in a dimensional manner [34]. The LPFS has proven to be a valid measure of personality functioning impairment [11, 34] with interrater reliability ranging from adequate to good ( $ICC = 0.7–0.9$ ) [24]. Prevailing interview-based measures of personality functioning impairment include the Semi-Structured Interview for Personality Functioning DSM-5 (STiP-5.1) [35], the Structured Clinical Interview for the Level of Personality Functioning Scale

<sup>1</sup> Although not an investigation of test–retest reliability (temporal stability), it is worth mentioning here that Weekers et al. [25] demonstrated the LPFS-BF 2.0's ability to detect expected changes over time in response to treatment by administering the instrument to treatment-seeking adults at intake and again at the end of treatment, with a mean time between administrations of 92.13 days. In doing so, they also found a pre–post correlation of  $r = 0.49$ .

(SCID-AMPD Module 1) [36], the Clinical Assessment of the Level of Personality Functioning Scale (CALF) [37], and the PDS-ICD-11 Clinician-Rating Form [38]. Designed as (semi-)structured interviews, these instruments range in duration, from ~ 30 min to upwards of 45–90 min. Multiple studies have demonstrated their criterion-related and predictive validity, while internal consistency is regularly good and interrater reliability has ranged from 0.59 to 0.96 across the various instruments and their subscales.

As shown by this brief summary, there are myriad self-report, interview, and clinician-/informant-rated measures of personality functioning impairment. There has yet to be a valid performance-based instrument specifically developed to measure personality functioning impairment as defined in the AMPD or ICD-11. Although researchers have demonstrated the potential for using the Rorschach [39, 40] and Thematic Apperception Test [41, 42] to assess personality functioning impairment, reliable and valid administration and scoring of these tests is time-intensive, requires extensive training, and is often incompatible with the type of remote administration required for online studies. Moreover, the constructs (currently) assessed via the Rorschach and Thematic Apperception Test conceptually diverge from personality functioning impairment as operationalized in the AMPD and ICD-11. There is a discernable absence of a performance-based measure of personality functioning impairment that is psychometrically robust, aligns with the AMPD and ICD-11, and can be administered and scored just as quickly, easily, and remotely as self-report tests. The lack of such an instrument is deleterious for both research and clinical practice because overreliance on a single method of measurement can lead to flawed empirical and clinical conclusions based on biased or incomplete representations of personality (see [43–46] for discussions). There is a clear need for a psychometrically sound performance-based measure of personality functioning impairment that aligns with the AMPD and ICD-11 and that can be easily, quickly, and remotely administered and scored.

## 1.2 Performance-based measurement methodology and the questionnaire-based IAT

Performance-based measures are a broad class of instruments wherein outcomes are derived from an individual's unrehearsed performance on a (semi-)structured task designed to tap behavior and patterns of responding that reflect—or are influenced by—the latent construct being measured. Information-processing measures are a subtype of the performance-based class. These tests require individuals to complete a reaction-time task by which latent constructs are assessed with the understanding that attributes of the latent construct influence the individual's task performance in measurable ways via cognitive and other psychological processes largely occurring outside their conscious awareness. The Implicit Association Test (IAT) [46] is one such instrument. The IAT and its variants measure the strengths of implicit—relatively automatic—associations between concepts using a double-categorization reaction-time task (see [46] for further explanation). Intraindividual differences in performance speeds (latencies) between critical test blocks are the basis of IAT scores, with scores representing the strength of automatic associations between the measured concepts in the individual's mind. With regard to measuring personality functioning impairment, IAT scores would be interpreted as the degree to which the individual associates their concept of themselves with the latent variable of personality dysfunction.

Findings from studies of the IAT's psychometric properties are fairly variable. Meta-analysis [47] produced an average internal consistency of  $\alpha = 0.80$  for the original IAT method, with average alphas of IAT variants ranging from 0.60 to 0.79. Test–retest reliabilities for the IAT averaged  $r = 0.50$  across 58 studies, a result Greenwald and Lai [47] partly attributed to reliability reductions caused by operational requirements of certain research situations, such as those seen with online data collection. Regarding construct validity, meta-analytic studies have found aggregate correlations between the IAT and self-reports measuring analogous constructs to range from  $r = 0.116$  to  $r = 0.361$  and aggregate correlations between IAT measures and external criteria to range from  $r = 0.097$  to  $r = 0.274$  [48–52]. These effect sizes are comparable to those commonly observed between well-validated performance-based measures (e.g., Rorschach) and scores obtained via self-report or informant-rating scales (correlations typically in the  $r = 0.20$ – $0.30$  range) [53–56]. Taken together, psychometric benchmarks for a new performance-based measure of personality functioning impairment based on the IAT methodology can be judiciously set at (a) internal reliability coefficients of approximately  $\alpha = 0.80$  or better, (b) test–retest reliability better than  $r = 0.50$ ,<sup>2</sup> and (c) correlation coefficients representing convergence with validated self-report and informant-rating tests within the  $r = 0.20$ – $0.30$  range or stronger, in addition to (d) evidence of discriminant validity.

<sup>2</sup> Although poor by most standards, test–retest reliability coefficients of  $r = 0.50$  can be considered adequate for use in correlational and group differences research but would be insufficient to justify interpreting a single individual's score.

Although a personality functioning impairment IAT might meet the psychometric benchmarks enumerated above, many of the IAT's psychometric shortcomings are worsened when it is used to measure self-concepts due to various limitations (e.g., problematic response stimuli, vague and not easily translatable interpretations of scores, and restrictions on the complexity of the constructs that can be measured) [57]. The Questionnaire-based IAT (qIAT) [58, 59] was developed to address these specific barriers. The qIAT procedure is consistent with that of the original IAT but supplements the weaknesses listed above in several ways. Most notably, the qIAT uses items of validated self-report tests as stimuli and creates pairing combinations that parallel self-report procedures (e.g., true/false). Rather than single words or images, qIAT stimuli are statements resembling those found on self-report tests (e.g., *"I often think very negatively about myself"*, *"My emotions are usually well regulated and stable"*). Respondents are instructed to classify these statements as quickly and accurately as possible into relevant target categories (e.g., *"impaired"* versus *"not impaired"*, respectively) while concurrently classifying objective self-statements (e.g., *"I am doing a psychology experiment"*, *"I am playing football on the grass"*) into the categories of *"true"* or *"false"*. Despite unmistakable similarities with self-report measures, scores derived from the qIAT reflect an individual's automatic association between themselves and the concept being measured whereas self-report scores reflect the degree to which an individual thoughtfully attributes characteristics of a given trait, feeling, thought pattern, motive, behavior, level of functioning, or experience to themselves. That is, self-report scores quantify how the person views and (chooses to) presents themselves whereas qIAT scores would quantify how strongly the latent construct is connected to the individual's implicit self-conceptualization.

Like self-report tests, the qIAT is particularly well-suited for both research and clinical practice due to its brief administration (~ 5–10 min), simple administration and automatic scoring procedures, direct interpretability of scores (an important feature for clinical utility), and use of validated and standardized stimuli. It is for these reasons that the collection of studies reported herein aimed to develop and psychometrically evaluate a novel, performance-based measure of personality functioning impairment using the qIAT methodology.

### 1.3 Current study

The overarching goal of the reported series of studies was to develop the Level of Personality Functioning Scale-qIAT (LPFS-qIAT) and conduct an initial psychometric evaluation of this new instrument. Following development of the LPFS-qIAT (see Sect. 8 below for details), four studies were conducted to examine the LPFS-qIAT's internal and retest reliability and the convergent, discriminant, and criterion-related validity of this new instrument.

## 2 Method

### 2.1 Participants

All results reported below were obtained using participants recruited through four separate online studies conducted over the course of approximately two years at a medium-sized public university in the Southern United States. Participation was restricted to students who were at least 18 years of age and English-speaking. All participants were required to complete the online studies on a device with a keyboard, which was necessary for administration of the IAT and qIAT measures. Participants were notified at the outset of their participation if they attempted a study on an incompatible device (e.g., touchscreen phone) and were asked to switch to a compatible device, at which point they were permitted to begin the study. No other restrictions were placed upon participation.

Similar procedures were used in each of the four studies: Following informed consent, participants were asked to complete a series of self-report and performance-based measures in a randomized order before completing an extended demographic questionnaire. For those studies including informant ratings and a second participation session (Samples 3 and 4), participants were asked to provide their own email address and the email addresses of two (Sample 3) or three (Sample 4) individuals who know them reasonably well. Informants were automatically emailed a study invitation immediately following a participant's provision of their email address. Participants were automatically emailed a Time 2 study invitation exactly one (Sample 3) or two (Sample 4) weeks following their survey completion at Time 1. Each study received approval from the Institutional Review Board at the authors' university and all participants received course credit for participating. The following describes the conventional data processing procedures that were followed for each study and descriptive statistics of each sample are reported in Table 1. Results of group comparisons between those who

**Table 1** Descriptive statistics of all samples

	Sample 1 (N=806)	Sample 2 (N=497)	Sample 3 (time 1) (N=209)	Sample 3 (time 2) (N=104)	Sample 4 (time 1) (N=236)	Sample 4 (time 2) (N=48)
Age						
μ (SD)	21.19 (4.83)	21.12 (4.91)	21.82 (6.88)	21.58 (7.23)	20.60 (4.58)	21.17 (5.22)
Range	18–57	18–57	18–56	18–56	18–58	18–44
Sex (%)						
Female	82.98%	85.71%	86.60%	87.13%	86.44%	89.58%
Male	16.52%	13.88%	12.92%	11.88%	12.71%	10.42%
Other/I wish not to respond	0.50%	0.40%	0.48%	0.99%	0.85%	0.00%
Ethnicity (%)						
Hispanic/Latine	31.39%	34.00%	29.19%	29.70%	36.44%	41.67%
Non-Hispanic/Latine	68.61%	66.00%	70.34%	70.30%	63.56%	58.33%
Race (%)						
White or Caucasian	68.11%	69.62%	64.33%	65.48%	62.50%	70.83%
Black or African American	14.02%	12.07%	17.20%	16.67%	16.18%	20.83%
Asian or Asian American	3.10%	2.41%	5.10%	5.95%	1.47%	0.00%
American Indian or Alaska Native	2.23%	3.22%	4.46%	3.57%	4.41%	0.00%
Native Hawaiian or Other Pacific Islander	0.25%	0.81%	0.00%	0.00%	0.74%	0.00%
Mixed race	8.31%	5.43%	5.73%	3.57%	9.56%	0.00%
Other/I wish not to respond	3.97%	6.44%	1.91%	4.76%	4.41%	8.33%

As Time 2 data were only used in the examination of test–retest reliability, the Time 2 samples above only describe those participants who provided usable data for both Time 1 and Time 2. Percentage of sample reporting Psychology as (one of) their college major(s) or minor(s): Sample 1 (54.0%), Sample 2 (49.7%), Sample 3 (53.6%), and Sample 4 (47.2%)

**Table 2** LPFS-qiAT test stimuli

True	False
<i>Self-related logical categories and statements</i>	
– I am looking at a computer screen	– I am sunbathing at the beach
– I am participating in an experiment on the internet	– I am climbing a steep mountain
– I am doing a psychology experiment	– I am currently playing an electric guitar
– I am putting my fingers on the keyboard	– I am buying groceries in the local grocery store
– I am in front of the computer	– I am playing football on the grass
Impaired	Not impaired
<i>Personality functioning impairment categories and stimuli</i>	
– I often do not know who I really am	– I generally know who I really am
– I often think very negatively about myself	– I rarely think very negatively about myself
– My emotions change without me having a grip on them	– My emotions are usually well regulated and stable
– I have no sense of where I want to go in my life	– I have a good sense of where I want to go in my life
– I often do not understand my own thoughts and feelings	– I often understand my own thoughts and feelings
– I often make unrealistic demands of myself	– I often make realistic demands of myself
– I often have difficulty understanding the thoughts and feelings of others	– I often understand the thoughts and feelings of others
– I can't stand it when others have a different opinion than me	– I can appreciate others' opinions, even when we disagree
– I often don't understand the effect my behavior has on others	– I often understand the effect my behavior has on others
– My relationships and friendships usually don't last long	– My relationships and friendships usually last long
– I often feel uncomfortable when relationships become more intimate	– I often feel comfortable when relationships become more intimate
– I often have trouble cooperating with others	– I often cooperate with others well

completed the full study and those who dropped out or were removed due to excessively fast responding on the IAT or qiAT can be found in the Supplementary Material of this paper.

**Sample 1.** A total of 1043 participants completed the first study and endorsed taking part in the study seriously and honestly. Three participants did not complete the LPFS-qiAT and 234 participants were removed due to excessively fast responses (see “Data processing” section below), a drop rate within ranges commonly seen for anonymous online data collections [60]. Thus, the final sample included 806 participants.

**Sample 2.** A total of 708 students participated in the second study and endorsed taking part in the study seriously and honestly. The final sample included 497 participants after removing 211 participants due to excessively fast responding on the qiAT measure.

**Sample 3.** A total of 287 participants completed Time 1 of the third study and 156 (54%) completed Time 2 and endorsed taking part in the studies seriously and honestly. An additional 78 participants from Time 1 and 41 participants from Time 2 were removed due to excessively fast responding. This resulted in final samples of 209 participants for Time 1 and 115 participants for Time 2. Usable data for both Time 1 and Time 2 was provided by 104 participants.

**Sample 4.** A total of 320 participants completed Time 1 of the fourth study and 78 (24%) completed Time 2, and endorsed taking part in the studies seriously and honestly. Eighty-four participants from Time 1 and 22 participants from Time 2 were removed due to excessively fast responding. This resulted in final samples of 236 participants for Time 1 and 56 participants for Time 2. Forty-eight participants provided usable data for both Time 1 and Time 2.

## 2.2 Measures

### 2.2.1 Level of personality functioning scale-qiAT (LPFS-qiAT; introduced here)

The qiAT can be viewed as an amalgamation of self-report and IAT procedures that enables indirect measurement of a latent factor. Development of the LPFS-qiAT made use of the self-related logical categories presented by Friedman et al. [59] and items from the LPFS-BF 2.0 [25]. The qiAT requires opposing item sets; LPFS-BF 2.0 items reflect impaired

personality functioning and a complementary set of items was created by reversing the meaning of the original LPFS-BF 2.0 items. The full list of LPFS-qIAT items is presented in Table 2. The LPFS-qIAT was built in R using the *iatgen* tool [60] with code modifications made to accommodate the qIAT method. The qIAT involves a procedure comparable to the IAT with minor differences. Specifically, participants complete multiple, brief double-categorization reaction-time tasks wherein a single statement (i.e., test item) is presented on the center of a computer screen along with category labels at the top right and top left corners of the screen. Participants are instructed to classify each statement using the presented category labels as quickly and as accurately as possible by pressing either the “E” or “I” key on their keyboard. The full procedure (see Table 3) consists of seven blocks, including two learning blocks (1, 2), three practice blocks (3, 5, 6), and two critical test blocks (4, 7). There are multiple strategies for handling incorrect responses (erroneous classifications), and the present project made use of two well-established methods for the purpose of generalizability: Sample 1 made use of a 600 ms error message with no response correction required and a forced error correction method was used with Samples 2, 3, and 4 [61]. A copy of the LPFS-qIAT for administration via Qualtrics and scoring code are available at [https://osf.io/8bfka/?view\\_only=c1d5214c613547c1a07a0be548109d36](https://osf.io/8bfka/?view_only=c1d5214c613547c1a07a0be548109d36). Details on data processing, scoring, and psychometric properties of the LPFS-qIAT are reported in the Results section below.

### 2.2.2 Level of personality functioning scale-brief form 2.0 (LPFS-BF 2.0)

The LPFS-BF 2.0 [25] is a 12-item self- and informant-rating measure of self-functioning and interpersonal functioning, producing two corresponding scale scores and a total personality functioning impairment score. This measure uses a 4-point Likert scale, ranging from 1 (*completely untrue*) to 4 (*completely true*). Internal consistencies were reported to be satisfactory for the total scale, as well as the self and interpersonal functioning scales, and meaningful associations were reported between the LPFS-BF 2.0 and other measures of severity of PDs [25]. The LPFS-BF 2.0's total personality functioning impairment scale demonstrated acceptable internal consistency in the present studies for each sample, with reliability coefficients ranging from  $\alpha = 0.88$  and  $\omega = 0.88$  (Sample 1) to  $\alpha = 0.90$  and  $\omega = 0.90$  (Sample 4). Reliability coefficients of the informant-rating version were  $\alpha = 0.88$  and  $\omega = 0.88$  in both Sample 3 and Sample 4. It should be noted that the LPFS-BF 2.0 was replaced with the Level of Personality Functioning Scale-Self-Report (LPFS-SR) [26] for Samples 2 and 3. Study time constraints prevented administration of multiple self-report tests of personality functioning impairment and the authors opined that use of a criterion measure that does not share items with the LPFS-qIAT would make for a more robust scrutiny of the LPFS-qIAT's validity. With respect to Sample 3 specifically, this replacement also allowed for a direct comparison of test–retest reliability results to previous external retest reliability findings (see [33], for a study on the retest reliability of the LPFS-SR). Due to the shorter length of the LPFS-BF 2.0, it replaced the LPFS-SR in the final study (Sample 4) to permit administration of a large assortment of measures used to examine the criterion and discriminant validity of the LPFS-qIAT.

### 2.2.3 Level of personality functioning scale-self report (LPFS-SR)

The LPFS-SR [26] is an 80-item self-report questionnaire designed to assess the severity of personality dysfunction by capturing the aspects of the LPFS in the AMPD. This measure uses a 4-point scale, ranging from *totally false* to *very true*. The LPFS-SR was reported to have good internal consistency and positive associations with other measures of similar constructs [26], as well as strong test–retest reliability over multiple weeks ( $r_s = 0.81–0.91$ ) [34]. The LPFS-SR total scale demonstrated acceptable internal consistency in each of the three studies including the LPFS-SR, with coefficients ranging from  $\alpha = 0.94$  and  $\omega = 0.95$  (Sample 3) to  $\alpha = 0.95$  and  $\omega = 0.96$  (Sample 2). Due to an already lengthy battery of tests being administered to Sample 4, the LPFS-SR was not included in the final study to avoid exhausting study participants.

### 2.2.4 DSM-5-TR self-rated level 1 cross-cutting symptom measure—adult (DSM-5-TR CC)

The DSM-5-TR CC [15] was developed to help clinicians assess all major areas of psychiatric functioning (e.g., mood, psychosis, cognition, personality, sleep) and identify additional areas for inquiry by revealing possible disorders, atypical presentations, subsyndromal conditions, and coexistent pathologies. The DSM-5-TR CC is endorsed by the American Psychiatric Association [15] as a necessary first step in identifying and addressing the heterogeneity of symptoms across

**Table 3** Task sequence for the LPFS-qIAT

Block	Trials	Task	Purpose	Left key	Right key
1	40	Discrimination of personality functioning impairment	Introduction to classification of personality functioning impairment and practice	Impaired	Not impaired
2	20	Discrimination of self-related logical statements	Introduction to classification of self-related logical statements and practice	True	False
3	20	Initial combined task (practice block)	Introduction to combined discrimination task and practice	Impaired or true	Not impaired or false
4	40	Initial combined task (test block)	Critical combined discrimination task	Impaired or true	Not impaired or false
5	40	Discrimination of personality functioning impairment-reversed	Practice classification of personality functioning impairment with keys reversed	Not impaired	Impaired
6	20	Reversed combined task (practice block)	Critical combined discrimination task with keys reversed	Not impaired or true	Impaired or false
7	40	Reversed combined task (test block)		Not impaired or true	Impaired or false

Response key designation and presentation of compatible or incompatible block were randomized in all studies. This example is one of four possible combinations



diagnostic categories. The present study demonstrated acceptable internal consistency of the DSM-5-TR CC's personality functioning domain scale ( $\alpha=0.80$ ,  $\omega=0.80$ ).

### 2.2.5 Personality inventory for DSM-5-brief form plus modified (PID5BF + M)

The PID5BF + M [62] is a 34-item self-report measure designed to assess the pathological trait domains described in Criterion B of the DSM-5's AMPD and the chapter on personality disorders and related traits in the ICD-11: Negative Affectivity, Detachment, Disinhibition, Antagonism/Dissociality, Anankastia, and Psychoticism. The PID5BF + M is a modified version of the PID5BF + [63], which was derived from the Personality Inventory for DSM-5 [64] via ant colony optimization and demonstrated acceptable model fit, good reliability, and criterion-related validity across multiple samples [63]. The PID5BF + M was created to better capture the ICD-11 domain of anankastia and its three facets. The PID5BF + M has been shown to be a psychometrically sound instrument for assessing the six combined DSM-5 and ICD-11 personality trait domains [62]. Internal consistency reliability for the six trait domain scales used in the present studies ranged from  $\alpha=0.76$  and  $\omega=0.76$  (Detachment, Sample 3) to  $\alpha=0.88$  and  $\omega=0.88$  (Anankastia, Samples 3 and 4).

### 2.2.6 Schwartz outcome scale-10 (SOS-10)

The SOS-10 [65] is a 10-item self-report created to measure a broad domain of psychological health and well-being. Blais et al. [65] demonstrated the SOS-10's convergent and divergent validity, ability to differentiate clinical and non-clinical samples, acceptable test-retest reliability, and good internal reliability ( $\alpha=0.96$ ). The present study produced acceptable internal consistency of the SOS-10 ( $\alpha=0.94$ ,  $\omega=0.94$ ).

### 2.2.7 World Health Organization's Quality of Life Brief Scale (WHOQOL-BREF)

The WHOQOL-BREF [66] is a 26-item self-report measure of four quality of life domains: Physical Health, Psychological, Social Relationships, and Environment. Derived from the WHOQOL-100, the WHOQOL-BREF was reported to correlate with WHOQOL-100 domain scores at approximately  $r=0.90$ . Discriminant validity, content validity, and test-retest reliability were reported to be good for the WHOQOL-BREF [66]. The present study demonstrated acceptable internal consistency of domain scales, with reliability coefficients ranging from  $\alpha=0.67$  and  $\omega=0.69$  (Social Relationships) to  $\alpha=0.85$  and  $\omega=0.85$  (Psychological).

### 2.2.8 Inventory of depression and anxiety symptoms-second version (IDAS-II)

The IDAS-II [67] is a 99-item self-report questionnaire composed of 19 scales designed to assess a broad range of depression, anxiety, and bipolar symptoms. This measure uses a 5-point scale ranging from *not at all* to *extremely*. The IDAS-II scales have shown good convergent, discriminant, and criterion validity [67]. The present study demonstrated acceptable internal consistency with domain scale reliability coefficients of  $\alpha=0.80$ – $0.91$  and  $\omega=0.81$ – $0.92$ .

### 2.2.9 Personality implicit association test-extraversion (extraversion IAT)

Using the methodology developed by Carpenter et al. [60], the extraversion IAT (see [68]) was administered via participants' personal computers through an online data-collection platform (Qualtrics). Stimuli from the categories of self (*me, my, own, I, self*) and others (*they, your, them, you, others*) and items from the categories of extraversion (*sociable, talkative, active, impulsive, outgoing*) and introversion (*shy, withdrawn, passive, deliberate, reserved*) were presented [68]. Testing consisted of seven blocks and took approximately five minutes to complete. Relative to IATs measuring other self-concept constructs (e.g., racial or political biases), personality trait IATs have routinely demonstrated better reliability and construct validity; extraversion and neuroticism IATs appear to be the most robust [68]. Data processing procedures used for the Extraversion IAT are reported in the results section.

### 2.2.10 Social desirability-gamma short scale (KSE-G)

The KSE-G [69] is a 6-item self-report instrument that measures two aspects of socially desirable responding: Exaggeration of positive qualities (PQ+) and Minimization of negative qualities (NQ-). Originally developed in German (Kurzsкала Soziale

Erwünschtheit–Gamma) [70], the English-language adaptation has demonstrated reliability and validity coefficients comparable to those of the original German version, as well as metric measurement invariance [69]. Convergent and discriminant validity of the KSE-G supports the instrument's construct validity, and estimates of internal consistency ( $\alpha$ ) have ranged between 0.65 and 0.72 for PQ+ and between 0.64 and 0.79 for NQ– [69]. The present study obtained approximately comparable levels of internal consistency for the PQ+ ( $\alpha=0.58$ ,  $\omega=0.58$ ) and the NQ– ( $\alpha=0.65$ ,  $\omega=0.65$ ).

### 2.2.11 Indecisiveness Scale (IS)

The IS [71] is a 22-item self-report measure of general indecisiveness, which pertains to difficulties in making decisions regardless of whether they are of great or little significance. The IS is composed of 11 features, each of which include one positively formulated and one negatively formulated item. Psychometric study of the IS has found support for a unifactor structure, acceptable levels of test–retest reliability, and predictive validity in terms of various decision-making situations [71]. The Present study demonstrated acceptable internal consistency ( $\alpha=0.92$ ,  $\omega=0.92$ ).

## 2.3 Statistical analyses

Following development of the LPFS-qIAT, four studies were conducted to examine the LPFS-qIAT's internal and retest reliability and the convergent, discriminant, and criterion-related validity of this new instrument. The LPFS-qIAT and Extraversion IAT data for all samples were processed using a *D*-score data cleaning and scoring algorithm (see [60]). First, participants with more than 10 percent of responses faster than 300 ms (i.e., excessively fast responders) were removed. Then, in Sample 1, participant response errors were addressed by replacing participants' response errors with their block means of correct trials plus 600 ms (i.e., the  $D_{600}$  procedure). The built-in-error-penalty method, wherein participants are required to self-correct an error before proceeding, is a preferred method [61, 72] and was used in Samples 2, 3, and 4. The use of different methods across studies was done for the purpose of generalizability. Finally, LPFS-qIAT and Extraversion IAT responses were *D*-scored, with positive scores indicating greater levels of personality functioning impairment or extraversion. Internal consistency reliability of the LPFS-qIAT was assessed using a variant of Cronbach's alpha [73], and Pearson correlations using a bootstrapping procedure with 5000 replicates were used to evaluate the instrument's test–retest reliability. Correlation coefficients and associated 95% confidence intervals based on 5000 bootstrap replicates were also calculated to evaluate the convergent, discriminant, and criterion-related validity of the LPFS-qIAT. In all cases, the statistical significance of an effect size was determined by examining the 95% confidence intervals computed around the estimated effect size and deemed significant at  $p < 0.05$  when a value of zero (null) did not fall between the upper and lower intervals.

## 3 Results

### 3.1 Data processing

For all samples, the LPFS-qIAT and Extraversion IAT data were processed using a *D*-score data cleaning and scoring algorithm (see [60]). The percent of dropped trials (trials  $> 10$  s) was low, ranging from 0.0043% (Time 2 Sample 4) to 0.0057% (Time 1 Sample 4). The percentage of participants removed due to excessively fast responding, which was expected to be elevated in our student convenience samples, ranged from 22.7% (Sample 1) to 29.8% (Sample 2) and was within the ranges commonly seen for anonymous online data collections [60]. As further detailed in the Supplementary Material, removed participants (i.e., excessively fast responders) were significantly younger and self-reported higher levels of psychopathology than retained participants in our two large samples (Samples 1 and 2). This pattern was not observed in Samples 3 or 4, nor were removed participants in Samples 3 and 4 found to differ from retained participants with respect to informant ratings. Participants in Samples 3 and 4 who completed Time 2 did not significantly differ from those who dropped out on any measure of personality functioning impairment, but drop-out participants did endorse significantly higher levels of disinhibition and psychoticism at Time 1.

Error rates on the LPFS-qIAT ranged from 11.4% (Sample 2) to 14.3% (Time 1 Sample 3) and were handled using two different methods across studies. For Sample 1, participants' response errors were replaced with their block means of correct trials plus 600 ms (i.e., the  $D_{600}$  procedure). The built-in-error-penalty method, wherein participants are required

to self-correct an error before proceeding, is a preferred method [61, 72] and was used in Samples 2, 3, and 4. The use of different methods across studies was done for the purpose of generalizability. LPFS-qIAT and Extraversion IAT responses were then *D*-scored, with positive scores indicating greater levels of personality functioning impairment or extraversion.

## 3.2 Reliability

### 3.2.1 Internal consistency reliability

Internal consistency reliability of the LPFS-qIAT, assessed using a variant of Cronbach's alpha [73], was found to approach, and in most cases, exceed that typically observed for the IAT (averaged  $\alpha = 0.80$ ) [47]. Except for Sample 2 ( $\alpha = 0.792$ ) and Sample 3 Time 2 ( $\alpha = 0.792$ ), internal consistency reliability of the LPFS-qIAT was above 0.800 and ranged from 0.820 to 0.875.

### 3.2.2 Test-retest reliability

One hundred four participants from Sample 3 and 48 participants from Sample 4 completed the LPFS-qIAT at both Time 1 and Time 2 of their respective studies and were used to calculate test-retest reliability over the span of one and two weeks, respectively. Pearson correlations using a bootstrapping procedure with 5000 replicates were used to evaluate test-retest reliability. Calculations produced a correlation coefficient of 0.381 ( $n = 104$ ; 95% CI 0.200, 0.552) for test-retest reliability over one week and a coefficient of 0.478 ( $n = 48$ , 95% CI 0.285, 0.661) over two weeks. Surprisingly, the LPFS-qIAT demonstrated slightly better test-retest reliability over a one-week period than the LPFS-SR in the same study ( $r = 0.334$ ;  $n = 102$ ; 95% CI 0.137, 0.516), although this was substantially poorer than that found by Hopwood et al. [33] for the same test over a two-week period ( $r_s = 0.81-0.91$ ). The test-retest reliability of the LPFS-qIAT over a two-week period was poorer than that found for the LPFS-BF 2.0 over the same duration ( $r = 0.658$ ;  $n = 46$ ; 95% CI 0.386, 0.910). To our knowledge, the test-retest reliability of the LPFS-BF 2.0 has not been published or investigated, making this a novel finding.

## 3.3 Validity

### 3.3.1 Convergent and criterion-related validity

Convergent validity of the LPFS-qIAT with self-report measures of personality functioning impairment was evaluated in each of the four samples, as well as with informant-rating measures in Samples 3 and 4. Correlation coefficients and associated 95% confidence intervals based on 5000 bootstrap replicates are reported in Table 4. Overall, the LPFS-qIAT significantly correlated with all self-report and informant-rating measures of personality functioning impairment in the expected direction. Correlation coefficients ranged from  $r = 0.143$  (95% CI 0.006, 0.282) (LPFS-SR in Sample 3) to  $r = 0.227$  (95% CI 0.090, 0.362) (LPFS-BF 2.0 in Sample 4), with all correlations significant and most correlations hovering around the  $r = 0.20$  range.<sup>3</sup> The obtained effect sizes are comparable to those typically observed between performance-based and self-report or informant-rating scores on measures of analogous constructs (typically in the  $r = 0.20-0.30$  range) [53–56].

Also reported in Table 4 are results pertaining to the LPFS-qIAT's criterion-related validity. LPFS-qIAT scores were significantly positively related to Negative Affectivity in all samples, and occasionally associated with other maladaptive personality trait scores; however, correlations between the LPFS-qIAT and trait scores were consistently smaller in magnitude than those observed between the LPFS-qIAT and measures of personality functioning impairment. The LPFS-qIAT was significantly negatively correlated with overall psychological health (SOS-10), quality of social relationships (WHOQOL-BREF Social Relationships), and perceived environmental quality and resources (WHOQOL-BREF Environment). Associations between LPFS-qIAT scores and psychological ( $r = -0.118$ ; 95% CI 0.017,  $-0.247$ ) and physical ( $r = -0.123$ ; 95% CI 0.006,  $-0.250$ ) health were in the expected direction but did not reach a level of statistical significance at  $p < 0.05$ .

<sup>3</sup> As shown in Table S16 of the Supplementary Material, correlations between the LPFS-qIAT and other measures of personality functioning impairment remained statistically significant when participants that had been removed for excessively fast responding were reintroduced. These results help minimize concerns over the statistically significant differences seen on some measures of psychopathology between participants who were retained and those who were removed in Samples 1 and 2.

**Table 4** Correlation coefficients depicting the convergent validity of the LPFS-qIAT with other measures of personality functioning impairment and the LPFS-qIAT's criterion-related validity

Measure	Method	Sample 1 (N=806)	Sample 2 (N=497)	Sample 3 (N=209)	Sample 4 (N=236)
lpfs-bf 2.0 total	Self-report	<b>0.194</b> (0.128, 0.261)	-	-	<b>0.227</b> (0.090, 0.362)
	Informant <sup>a</sup>	-	-	<b>0.148</b> (0.051, 0.244)	-
LPFS-SR total	Self-report	<b>0.189</b> (0.118, 0.257)	<b>0.198</b> (0.112, 0.292)	<b>0.143</b> (0.006, 0.282)	-
DSM-5-TR CC PF	Self-report	-	-	-	<b>0.187</b> (0.054, 0.313)
Negative affectivity	Self-report	<b>0.124</b> (0.054, 0.191)	<b>0.097</b> (0.011, 0.182)	<b>0.139</b> (0.016, 0.261)	<b>0.138</b> (0.020, 0.251)
Detachment	Self-report	<b>0.163</b> (0.097, 0.229)	<b>0.154</b> (0.066, 0.243)	0.059 (-0.068, 0.190)	<b>0.207</b> (0.080, 0.329)
Antagonism	Self-report	<b>0.089</b> (0.019, 0.160)	0.027 (-0.053, 0.109)	0.049 (-0.092, 0.193)	0.090 (-0.052, 0.222)
Disinhibition	Self-report	<b>0.124</b> (0.053, 0.193)	<b>0.114</b> (0.020, 0.207)	0.065 (-0.067, 0.192)	<b>0.142</b> (0.004, 0.272)
Anankastia	Self-report	<b>0.080</b> (0.014, 0.142)	0.081 (-0.006, 0.169)	0.016 (-0.119, 0.155)	-0.010 (-0.134, 0.113)
Psychoticism	Self-report	<b>0.082</b> (0.012, 0.150)	<b>0.091</b> (0.005, 0.177)	0.038 (-0.087, 0.164)	0.113 (-0.025, 0.244)
SOS-10	Self-report	-	-	-	<b>-0.131</b> (-0.003, -0.254)
WHOQOL-BREF physical	Self-report	-	-	-	-0.123 (0.006, -0.250)
WHOQOL-BREF psychological	Self-report	-	-	-	-0.118 (0.017, -0.247)
WHOQOL-BREF social relationships	Self-report	-	-	-	<b>-0.137</b> (-0.007, -0.263)
WHOQOL-BREF environment	Self-report	-	-	-	<b>-0.150</b> (-0.034, -0.266)

Bold indicates statistical significance based on 95% confidence intervals from 5000 bootstrap replicates. Sample 1 personality traits scores were obtained using the PID5BF+ [62] whereas trait scores for all other samples obtained using the PID5BF+M [61]

<sup>a</sup>412 informant ratings of 285 participants

**Table 5** Correlation coefficients depicting the discriminant validity of the LPFS qIAT

Discriminant measure	Sample 1 (N=806)	Sample 2 (N=497)	Sample 3 (N=209)	Sample 4 (N=236)
Age	-0.057 (-0.133, 0.021)	-0.040 (-0.117, 0.039)	-0.095 (-0.234, 0.046)	-0.083 (-0.178, 0.026)
Extraversion IAT	-	0.005 (-0.085, 0.091)	-	-
IDAS-II				
General depression	-	-	-	0.087 (-0.049, 0.214)
Dysphoria	-	-	-	0.085 (-0.049, 0.213)
Lassitude	-	-	-	0.095 (-0.042, 0.226)
Insomnia	-	-	-	0.040 (-0.088, 0.166)
Suicidality	-	-	-	-0.037 (-0.170, 0.104)
Appetite loss	-	-	-	0.014 (-0.118, 0.142)
Appetite gain	-	-	-	0.079 (-0.048, 0.203)
Well-being	-	-	-	-0.107 (-0.240, 0.032)
Ill temper	-	-	-	0.054 (-0.085, 0.195)
Mania	-	-	-	0.066 (-0.068, 0.197)
Euphoria	-	-	-	0.057 (-0.064, 0.177)
Panic	-	-	-	0.044 (-0.089, 0.176)
Social anxiety	-	-	-	0.081 (-0.048, 0.209)
Claustrophobia	-	-	-	0.046 (-0.092, 0.181)
Traumatic intrusions	-	-	-	0.023 (-0.122, 0.169)
Traumatic avoidance	-	-	-	0.091 (-0.049, 0.230)
Checking	-	-	-	0.045 (-0.080, 0.164)
Ordering	-	-	-	0.056 (-0.074, 0.185)
Cleaning	-	-	-	0.053 (-0.081, 0.185)
KSE-G				
Exaggerating positive qualities	-	-	-	-0.024 (-0.150, 0.101)
Minimizing negative qualities	-	-	-	0.063 (-0.082, 0.200)
Indecisiveness Scale				
Total indecisiveness	-	-	-	-0.019 (-0.161, 0.126)
Minimum detectable effect size	0.098	0.127	0.192	0.181

All correlation coefficients were found to be non-significant based on 95% confidence intervals from 5000 bootstrap replicates. Minimum detectable effect sizes derived from sensitivity analysis based on sample size and two-tailed significance test given  $\alpha=0.05$  and power of 0.80

### 3.3.2 Discriminant validity

Discriminant validity of the LPFS-qIAT was investigated through its association with self-report measures of depression, anxiety, indecisiveness, socially desirable responding, and age as well as its association with an IAT measure of extraversion. Sensitivity analysis based on sample size for a two-tailed significance test given  $\alpha=0.05$  and power of 0.80 indicated minimum detectable effect sizes ranging from  $r=0.098$  (Sample 1) to  $r=0.192$  (Sample 3), evidencing that the present study's discriminant validity analysis was powerful enough to detect even small effect sizes. Correlation coefficients and associated 95% CIs based on 5000 bootstrap replicates are reported in Table 5. In all cases, LPFS-qIAT scores minimally and non-significantly correlated with scores representing theoretically unrelated (or minimally related) constructs. Importantly, all effect sizes were smaller than those observed between the LPFS-qIAT and all self-report and informant-rating measures of personality functioning impairment.

## 4 Discussion

Available measures of personality functioning impairment as operationalized in the AMPD and ICD-11 exclusively use self-report, clinician-/informant-rated, or interview methods and there is a problematic absence of performance-based instruments capable of measuring personality functioning impairment. The LPFS-qlAT takes a step in filling this gap as the first performance-based measure of personality functioning impairment that directly parallels the AMPD and, to a degree, the ICD-11. The LPFS-qlAT is brief, with completion taking 5–10 min, is easily administered in person or online, and can be automatically scored in a manner allowing simple interpretation. This initial psychometric evaluation of the LPFS-qlAT produced encouraging evidence across four separate studies.

The LPFS-qlAT's good internal reliability was repeatedly observed, demonstrating internal consistency for this specific instrument, and allowing confidence in its online administration. The latter is a notable strength of the LPFS-qlAT compared to most other performance-based measures of personality. Although poor by self-report standards, one- and two-week test–retest reliability coefficients were approaching those often seen for IAT measures (averaged  $r=0.50$ ) [47], but are at levels insufficient to justify interpreting a single individual's score. Convergent validity of the LPFS-qlAT was established using three different self-report tests of personality functioning impairment and one informant-rated measure. The correlation coefficients obtained across the four studies were comparable in magnitude to those normally observed between performance-based (e.g., Rorschach, TAT) and self-report or informant-rating measures of analogous constructs (typically in the  $r=0.20$ – $0.30$  range) [53–56]. Evidence of the LPFS-qlAT's criterion-related validity was demonstrated by its positive correlation with maladaptive personality traits, particularly Negative Affectivity, and inverse relationships with measures of psychological health and quality of life. Finally, discriminant validity of the LPFS-qlAT was exhibited through small, non-significant correlations with self-report measures of depression, anxiety, indecisiveness, socially desirable responding, and age as well as a negligible and non-significant association with an IAT measure of extraversion. Taken together, the LPFS-qlAT can be cautiously viewed as an internally reliable and valid measure of personality functioning impairment based on this initial, but rather extensive psychometric examination. The LPFS-qlAT's test–retest reliability emerged as a notable limitation that must be addressed in future studies.

### 4.1 Limitations, constraints on Generalizability, and future directions

Although encouraging, the present studies are not without their limitations. First, each of the reported studies were conducted using diverse college student convenience samples, limiting the generalizability of findings. The present samples are more diverse in race and ethnicity than common for college student samples, but the current findings might not generalize to a broader population. Future studies should build upon the somewhat promising findings reported herein through generalizability studies using more diverse samples, including both community and clinical populations. Similarly, differences between retained/continued and removed/dropout participants may also limit the generalizability of results and should be investigated in future studies. However, it is necessary to first address the limitations of the LPFS-qlAT's retest reliability. The poor retest reliability results were obtained via online studies using small convenience samples. Future research may wish to first investigate the LPFS-qlAT's test–retest reliability in larger, more reliable samples. Should reliability coefficients not meaningfully improve, researchers should explore tactics for increasing retest reliability to a level that would allow confidence. One prospective solution to this issue involves averaging results from a repeated administration of the LPFS-qlAT at a single timepoint, a strategy often successfully used in research and clinical practice for other measures (e.g., blood pressure) [74]. Encouragingly, researchers using this strategy recently produced an IAT test–retest reliability of  $r=0.89$  over a period of 2 years [75]. The current study also did not specifically evaluate the LPFS-qlAT's incremental validity over other methods, such as self-report measures, when predicting meaningful outcomes (e.g., actual problems in relationships or at work). Investigating this instrument's incremental validity in future studies is important. When doing so, researchers must be mindful of undesired method effects and not measure outcomes using the same method against which the LPFS-qlAT is being compared (e.g., outcomes measured by self-reports should not be used when the LPFS-qlAT's incremental validity over self-report measures of personality functioning is being tested). Lastly, the LPFS-qlAT possibly fails to account for all components of the ICD-11's conceptualization of personality functioning, which is represented by self- and interpersonal dysfunction *in addition to* emotional, cognitive, and behavioral manifestations of the personality dysfunction [13]. Future studies will need to investigate the degree to which the LPFS-qlAT can capture all components of personality functioning impairment enumerated in the ICD-11. That is, future studies should test the LPFS-qlAT's construct validity using ICD-11-compliant criterion measures. Should the

LPFS-qlAT demonstrate poor construct validity in those studies, researchers might choose to apply the methodology described in this paper to a self-report measure that directly aligns with the ICD-11, such as the PDS-ICD-11 [31], to develop a new performance-based instrument that more closely matches the ICD-11's conceptualization of personality functioning. Despite these limitations and need for further studies, development (and further validation) of a novel, easily administered and automatically scored performance-based measure of personality functioning impairment has several implications.

Professional practice guidelines advocate for the use of multimethod assessment [45, 76, 77]. The availability of a performance-based measure of personality functioning impairment that is psychometrically promising, aligns with the AMPD (and partly with the ICD-11), and is easily administered, scored, and interpreted can prove valuable. Multimethod research designs minimize unwanted measurement effects, and the examination of convergences and divergences of test scores derived from multiple methods for measuring analogous constructs—intertest score relationships—can facilitate a more comprehensive and precise measurement of study variables. When similar strategies are applied in clinical practice, clinicians can gain a greater understanding of the relationship between a client's explicit and implicit personality dysfunction. Hence, the LPFS-qlAT, as the first performance-based measure of personality functioning aligning with contemporary models, holds potential to become a useful tool in both research and clinical practice by facilitating multimethod assessment. However, researchers must first address the LPFS-qlAT's poor retest reliability, investigate the generalizability of the present findings, and evaluate this instrument's incremental validity over other methods when predicting meaningful outcomes before this statement can be made more definitively. The LPFS-qlAT might also benefit research and clinical practice as a novel method for capturing changes in personality dysfunction. The LPFS-qlAT was derived from the LPFS-BF 2.0, and the latter has shown a high sensitivity to change in response to therapy [25]. Should future studies find that the LPFS-qlAT can also detect therapy-related changes in personality functioning, the LPFS-qlAT could be a useful tool for therapists and in therapy outcome research—perhaps especially when paired with the LPFS-BF 2.0. Alternatively, scores derived from performance-based measures such as the LPFS-qlAT are largely based on unrehearsed performance, which could interfere with using the instrument in this way due to the potential for practice effects [78]. The effect of repeated measurement using the LPFS-qlAT over the course of treatment, the potential measurement error due to practice effects, and the instrument's validity and utility for therapy outcome research could be important questions to explore once the other psychometric needs noted above have been met (e.g., improved retest reliability). Nevertheless, the findings of the reported studies offer encouragement for future research and continuing to improve and validate the LPFS-qlAT as a robust, performance-based measure of personality functioning.

**Author contribution** Both authors were involved in data curation, formal analysis, investigation, and writing. APN was lead on conceptualization, data curation, formal analysis, investigation, methodology, project administration, and writing.

**Funding** No funds, grants, or other support was received. The authors have no relevant financial or non-financial interests to disclose.

**Data availability** Deidentified data are only available upon reasonable request made to the first author (per IRB guidelines). Copies of the LPFS-qlAT and analysis code can be found in our online supplementary documents (see <https://osf.io/8bfka/>). This study was not preregistered.

## Declarations

**Ethics approval and consent to participate** Approval was obtained from the ethics committee of Sam Houston State University. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

**Consent for publication** Informed consent was obtained from all individual participants included in the study.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Comtois KA, Russo J, Snowden M, Srebniak D, Ries R, Roy-Byrne P. Factors associated with high use of public mental health services by persons with borderline personality disorder. *Psychiatr Serv (Washington, DC)*. 2003;54(8):1149–54. <https://doi.org/10.1176/appi.ps.54.8.1149>.
2. Hastrup LH, Jennum P, Ibsen R, Kjellberg J, Simonsen E. Costs of schizotypal disorder: a matched-controlled nationwide register-based study of patients and spouses. *Acta Psychiatr Scand*. 2021;144:60–71. <https://doi.org/10.1111/acps.13292>.
3. Soeteman DI, Hakkaart-van Roijen L, Verheul R, Busschbach JJ. The economic burden of personality disorders in mental health care. *J Clin Psychiatry*. 2008;69(2):259–65. <https://doi.org/10.4088/jcp.v69n0212>.
4. Moran P, Stewart R, Brugha T, Bebbington P, Bhugra D, Jenkins R, Coid JW. Personality disorder and cardiovascular disease: results from a national household survey. *J Clin Psychiatry*. 2007;68:69–74.
5. Newton-Howes GM, Tyrer P, Johnson T, Mulder R, Kool S, Dekker J, Schoevers R. Influence of personality on the outcome of treatment in depression: systematic review and meta-analysis. *J Pers Disord*. 2014;28:577–93.
6. Hastrup LH, Jennum P, Ibsen R, Kjellberg J, Simonsen E. Welfare consequences of early-onset borderline personality disorder: a nationwide register-based case-control study. *Eur Child Adolesc Psychiatry*. 2020;31(2):253–60.
7. Samuels J. Personality disorders: epidemiology and public health issues. *Int Rev Psychiatry*. 2011;23(3):223–33. <https://doi.org/10.3109/09540261.2011.588200>.
8. Bornstein RF. Illuminating a neglected clinical issue: Societal costs of interpersonal dependency and dependent personality disorder. *J Clin Psychol*. 2012;68:766–81. <https://doi.org/10.1002/jclp.21870>.
9. Zimmermann J, Kerber A, Rek K, Hopwood CJ, Krueger RF. A brief but comprehensive review of research on the Alternative DSM-5 Model for Personality Disorders. *Curr Psychiatry Rep*. 2019;21:92.
10. Bornstein RF, Natoli AP. Clinical utility of categorical and dimensional perspectives on personality pathology: a meta-analytic review. *Personal Disord Theory Res Treat*. 2019;10:479–90. <https://doi.org/10.1037/per0000365>.
11. Birkhölzer M, Schmeck K, Goth K. Assessment of criterion A. *Curr Opin Psychol*. 2021;37:98–103. <https://doi.org/10.1016/j.copsyc.2020.09.009>.
12. Ganellen RJ. Assessing normal and abnormal personality functioning: strengths and weakness of self-report, observer, performance-based methods. *J Pers Assess*. 2007;89(1):30–40. <https://doi.org/10.1080/00223890701356987>.
13. World Health Organization. *International Statistical Classification of Disease and Related Health Problems (11th edn)*. World Health Organization; 2018.
14. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. 5th edn. American Psychiatric Association; 2013.
15. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders-Text Revision: DSM-5-TR*. 5th edn. American Psychiatric Association; 2022.
16. Livesley WJ. Diagnostic dilemmas in classifying personality disorder. In: Phillips KA, First MB, Pincus HA, editors. *Advancing DSM: dilemmas in psychiatric diagnosis*. American Psychiatric Association; 2003. p. 153–90.
17. Rounsaville BJ, Alarcon RD, Andrews G, Jackson JS, Kendell RE, Kendler K. Basic nomenclature issues for DSM-V. In: Kupfer DJ, First MB, Regier DE, editors. *Research agenda for DSM-V*. American Psychiatric Association; 2002. p. 1–29.
18. Skodol AE. Personality disorder classification: Stuck in neutral, how to move forward? *Curr Psychiatry Rep*. 2014;16:480.
19. Widiger TA, Samuel DB. Diagnostic categories or dimensions? A question for the *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition*. *J Abnormal Psychol*. 2005;114:494–504.
20. Parker G, Both L, Olley A, Hadzi-Pavlocic D, Irvine P, Jacobs G. Defining personality disordered functioning. *J Pers Disord*. 2002;16:503–22.
21. Tyrer P. The problem of severity in the classification of personality disorders. *J Pers Disord*. 2005;19:309–14.
22. Hopwood CJ, Malone JC, Ansell EB, Sanislow CA, Grilo CM, Pinto A, Morey LC. Personality assessment in DSM-5: empirical support for rating severity, style, and traits. *J Pers Disord*. 2011;25:305–20.
23. Natoli AP, Bornstein RF. Validating the level of personality functioning scale: we don't use multimethod research designs. 2019; <https://doi.org/10.31234/osf.io/nhrd2>.
24. Roche MJ, Jaweed S. Comparing measures of Criterion A to better understand incremental validity in the alternative model of personality disorders. *Assessment*. 2021. <https://doi.org/10.1177/10731911211059763>.
25. Weekers LC, Hutsebaut J, Kamphuis JH. The Level of Personality Functioning Scale-Brief Form 2.0: update of a brief instrument for assessing level of personality functioning. *Personal Ment Health*. 2019;13:3–14. <https://doi.org/10.1002/pmh.1434>.
26. Morey LC. Development and initial evaluation of a self-report form of the DSM-5 Level of Personality Functioning Scale. *Psychol Assess*. 2017;29:1302–8.
27. Huprich SK, Nelson SM, Meehan KB, Siefert CJ, Haggerty G, Sexton J, Dauphin VB, Macaluso M, Jackson J, Zackula R, Baade L. Introduction of the DSM-5 levels of Personality Functioning Questionnaire. *Personal Disord Theory Res Treat*. 2018;9(6):553–63. <https://doi.org/10.1037/per0000264>.
28. Siefert CJ, Sexton J, Meehan K, Nelson S, Haggerty G, Dauphin B, Huprich S. Development of a short form for the DSM-5 levels of personality functioning questionnaire. *J Pers Assess*. 2020;102(4):516–26.
29. Gamache D, Savard C, Leclerc P, Côté A. Introducing a short self-report for the assessment of DSM-5 level of personality functioning for personality disorders: the Self and Interpersonal Functioning Scale. *Personal Disord Theory Res Treat*. 2019;10(5):438.
30. Roche MJ, Jacobson NC, Pincus AL. 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. *J Abnorm Psychol*. 2016;125(8):1090–102. <https://doi.org/10.1037/abn0000177>.
31. Bach B, Brown TA, Mulder RT, Newton-Howes G, Simonsen E, Sellbom M. Development and initial evaluation of the ICD-11 Personality Disorder Severity Scale: PDISCD-11. *Personal Ment Health*. 2021;15:223–36. <https://doi.org/10.1002/pmh.1510>.



32. Goth K, Birkhölzer M, Schmeck K. Assessment of personality functioning in adolescents with the LoPF-Q 12–18 self-report questionnaire. *J Pers Assess.* 2018;100(6):680–90.
33. Hopwood CJ, Good EW, Morey LC. Validity of the DSM-5 Levels of Personality Functioning Scale-Self Report. *J Pers Assess.* 2018;100:650–9.
34. Morey LC, Bender DS, Skodol AE. Validating the proposed Diagnostic and Statistical Manual of Mental Disorders, 5th edition, severity indicator for personality disorder. *J Nervous Mental Dis.* 2013;201(9):729–35.
35. Hutsebaut J, Kamphuis JH, Feenstra DJ, Weekers LC, De Saeger H. 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 Disord Theory Res Treat.* 2017;8(1):94.
36. Bender DS, Skodol AE, First MB, Oldham JM. Structured clinical interview for the DSM-5<sup>®</sup> alternative model for personality disorders (SCID-5-AMPD) module I level of personality functioning scale. APPI. 2018.
37. Thylstrup B, Simonsen S, Nemery C, Simonsen E, Noll JF, Myatt MW, Hesse M. Assessment of personality-related levels of functioning: a pilot study of clinical assessment of the dsm-5 level of personality functioning based on a semi-structured interview. *BMC Psychiatry.* 2016;16(1):1–8. <https://doi.org/10.1186/s12888-016-1011-6>.
38. Sellbom M, Brown TA, Bach B. Development and psychometric evaluation of the Personality Disorder Severity ICD-11 (PDS-ICD-11) Clinician-Rating Form. *Personal Ment Health.* 2023. <https://doi.org/10.1002/pmh.1596>.
39. Graceffo RA, Mihura JL, Meyer GJ. A meta-analysis of an implicit measure of personality functioning: The Mutuality of Autonomy Scale. *J Pers Assess.* 2014;96(6):581–95. <https://doi.org/10.1080/00223891.2014.919299>.
40. Valkonen H, Lindfors O, Knekt P. Association between the Rorschach Ego Impairment Index and the Level of Personality Organization interview assessment in depressive and anxiety disorder patients. *Psychiatry Res.* 2012;200(2–3):849–56.
41. Ruchensky JR, Dowgwillo EA, Kelley SE, Massey C, Slavin-Mulford J, Richardson LA, Blais MA, Stein MB. Exploring the alternative model for personality disorders using SCORS-G ratings on thematic apperception test narratives. *J Pers Disord.* 2022;36:201–16. [https://doi.org/10.1521/pedi\\_2021\\_35\\_535](https://doi.org/10.1521/pedi_2021_35_535).
42. Stein MB, Slavin-Mulford J, Sinclair SJ, Chung WJ, Roche M, Denckla C, Blais MA. Extending the use of the SCORS-G composite ratings in assessing level of personality organization. *J Pers Assess.* 2018;100(2):166–75.
43. Huprich SK, Bornstein RF, Schmitt TA. Self-report methodology is insufficient for improving the assessment and classification of Axis II personality disorders. *J Pers Disord.* 2011;25(5):557–70.
44. Mihura JL. The necessity of multiple test methods in conducting assessments: the role of the Rorschach and self-report. *Psycholog Injury Law.* 2012;5:97–106.
45. Natoli AP. More methods can result in more knowledge: why psychology needs to use multi-method approaches. *PSYPAG Quarterly.* 2019;112:11–3.
46. Greenwald AG, McGhee DE, Schwartz JLK. Measuring individual differences in implicit cognition: the implicit association test. *J Pers Soc Psychol.* 1998;74(6):1464–80. <https://doi.org/10.1037/0022-3514.74.6.1464>.
47. Greenwald AG, Lai CK. Implicit social cognition. *Annu Rev Psychol.* 2020;71:419–45. <https://doi.org/10.1146/annurev-psych-010419-050837>.
48. Greenwald AG, Poehlman TA, Uhlmann EL, Banaji MR. Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *J Personality Soc Psychol.* 2009;97(1):17.
49. Hofmann W, Gawronski B, Gschwendner T, Le H, Schmitt M. A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Pers Soc Psychol Bull.* 2005;31(10):1369–85.
50. Kurdi B, Seitchik AE, Axt JR, Carroll TJ, Karapetyan A, Kaushik N, et al. Relationship between the Implicit Association Test and intergroup behavior: a meta-analysis. *Am Psychol.* 2019;74(5):569.
51. Nosek BA, Greenwald AG, Banaji MR. Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personal Soc Psychol Bull.* 2005;31(2):166–80. <https://doi.org/10.1177/0146167204271418>.
52. Oswald FL, Mitchell G, Blanton H, Jaccard J, Tetlock PE. Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *J Pers Soc Psychol.* 2013;105(2):171.
53. Bornstein RF. A process dissociation approach to objective-projective test score interrelationships. *J Pers Assess.* 2002;78(1):47–68. [https://doi.org/10.1207/S15327752JPA7801\\_04](https://doi.org/10.1207/S15327752JPA7801_04).
54. Meyer GJ, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, et al. Psychological testing and psychological assessment: a review of evidence and issues. *Am Psychol.* 2001;56(2):128.
55. Mihura JL, Meyer GJ, Dumitrascu N, Bombel G. The validity of individual Rorschach variables: systematic reviews and meta-analyses of the comprehensive system. *Psychol Bull.* 2013;139(3):548.
56. Zeigler-Hill V, Fulton JJ, McLemore C. Discrepancies between explicit and implicit self-esteem: implications for mate retention strategies and perceived infidelity. *J Soc Psychol.* 2012;152(6):670–86.
57. Greenwald AG, Farnham SD. Using the implicit association test to measure self-esteem and self-concept. *J Pers Soc Psychol.* 2000;79(6):1022.
58. Yovel I, Friedman A. Bridging the gap between explicit and implicit measurement of personality: the Questionnaire-Based Implicit Association Test. *Personality Individ Differ.* 2013;54(1):76–80. <https://doi.org/10.1016/j.paid.2012.08.015>.
59. Friedman A, Katz BA, Cohen IH, Yovel I. Expanding the scope of implicit personality assessment: an examination of the Questionnaire-Based Implicit Association Test (qiAT). *J Pers Assess.* 2021. <https://doi.org/10.1080/00223891.2020.1754230>.
60. Carpenter TP, Pogacar R, Pullig C, Kouril M, Aguilar S, LaBouff J, Chakroff A. Survey-software implicit association tests: a methodological and empirical analysis. *Behav Res Methods.* 2019;51:2194–208.
61. Greenwald AG, Nosek BA, Banaji MR. Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *J Pers Soc Psychol.* 2003;85:197–216. <https://doi.org/10.1037/0022-3514.85.2.197>.
62. Bach B, Kerber A, Aluja A, Bastiaens T, Keeley JW, Claes L, Fossati A, Gutierrez F, Oliveira SES, Pires R, Riegel KD, Rolland JP, Roskam I, Sellbom M, Somma A, Spanemberg L, Strus W, Thimm JC, Wright AGC, Zimmermann J. International assessment of DSM-5 and ICD-11 personality disorder traits: toward a common nosology in DSM-5.1. *Psychopathology.* 2020;53(3–4):179–88. <https://doi.org/10.1159/000507589>.

63. Kerber A, Schultze M, Müller S, Rühling RM, Wright AGC, Spitzer C, Krueger RF, Knaevelsrud C, Zimmermann J. Development of a short and ICD-11 compatible measure for DSM-5 maladaptive personality traits using ant colony optimization algorithms. *Assessment*. 2022;29(3):467–87. <https://doi.org/10.1177/1073191120971848>.
64. Krueger RF, Derringer J, Markon KE, Watson D, Skodol AE. Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychol Med*. 2012;42(9):1879–90. <https://doi.org/10.1017/S0033291711002674>.
65. Blais MA, Lenderking WR, Baer L, deLorell A, Peets K, Leahy L, Burns C. Development and initial validation of a brief mental health outcome measure. *J Personality Assess*. 1999;73(3):359–73. [https://doi.org/10.1207/S15327752JPA7303\\_5](https://doi.org/10.1207/S15327752JPA7303_5).
66. WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med*. 1998;28(3):551–8.
67. Watson D, O'Hara MW, Naragon-Gainey K, Koffel E, Chmielewski M, Kotov R, Ruggero CJ. Development and validation of new anxiety and bipolar symptom scales for an expanded version of the IDAS (the IDAS-II). *Assessment*. 2012;19(4):399–420.
68. Back MD, Schmukle SC, Egloff B. Predicting actual behavior from the explicit and implicit self-concept of personality. *J Pers Soc Psychol*. 2009;97:533–48. <https://doi.org/10.1037/a0016229>.
69. Nießen D, Partsch MV, Kemper CJ, Rammstedt B. An English-language adaptation of the Social Desirability-Gamma Short Scale (KSE-G). *Measure Instrum Soc Sci*. 2019;2(2):1–10. <https://doi.org/10.1186/s42409-018-0005-1>.
70. Kemper CJ, Beierlein C, Bensch D, Kovaleva A, Rammstedt B. Soziale Erwünschtheit-Gamma (KSE-G) [Social Desirability-Gamma Short Scale (KSE-G)]. Zusammenstellung sozialwissenschaftlicher Items und Skalen. 2014. <https://doi.org/10.6102/zis186>.
71. Germeijs V, De Boeck P. A measurement scale for indecisiveness and its relationship to career indecision and other types of indecision. *Eur J Psychol Assess*. 2002;18(2):113.
72. Rudman LA. *Implicit measures for social and personality psychology*. Sage; 2011.
73. Schnabel K, Asendorpf JB, Greenwald AG. Using implicit association tests for the assessment of implicit personality self-concept. In: *Handbook of personality theory and assessment: personality measurement and testing*. UK: SAGE; 2008. p. 508–28. <https://doi.org/10.4135/9781849200479.n24>.
74. Stergiou GS, et al. Reproducibility of home, ambulatory, and clinic blood pressure: implications for the design of trials for the assessment of antihypertensive drug efficacy. *Am J Hypertens*. 2002;15:101–4.
75. Greenwald AG, Brendl M, Cai H, Cvencek D, Dovidio JF, Frieze M, Wiers R. The Implicit Association Test at age 20: what is known and what is not known about implicit bias. 2020;. <https://doi.org/10.31234/osf.io/bf97c>.
76. Hopwood CJ, Bornstein RF, editors. *Multimethod clinical assessment*. USA: Guilford; 2014.
77. Krishnamurthy R, Hass GA, Natoli AP, Smith BL, Arbisi PA, Gottfried ED. Professional practice guidelines for personality assessment. *J Pers Assess*. 2022;104(1):1–16. <https://doi.org/10.1080/00223891.2021.1942020>.
78. Fiedler K, Bluemke M. Faking the IAT: aided and unaided response control on the Implicit Association Tests. *Basic Appl Soc Psychol*. 2005;27(4):307–16. [https://doi.org/10.1207/s15324834basp2704\\_3](https://doi.org/10.1207/s15324834basp2704_3).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.