

Integrative Dimensional Personality Inventory for ICD-11: Development and Evaluation in the Peruvian Correctional Setting

Inventario Integrativo de la Personalidad Dimensional para la CIE-11: desarrollo y evaluación en el sistema penitenciario peruano

Luis Miguel Hualparuca-Olivera^{a,*}, Dayana Nicol Ramos Campos^a,
Paola Andrea Arauco Vivas^a, Rocío Milagros Coz Apumayta^a

^aUniversidad Continental, Perú

Recibido: 06 de diciembre de 2021

Aceptado: 05 de mayo de 2022

Abstract

Background: Personality disorder (PD) is an important predictor of the commission of crimes; however, there is a lack of clinical instruments adjusted to the characteristics of Peruvian convicts. **Objective:** To develop a reliable and valid comprehensive personality measurement instrument, the Integrative Dimensional Personality Inventory, ICD-11 version (IDPI-11), according to the standards of the International Classification of Diseases (ICD-11). **Method:** A stratified sample of prisoners from the Huancayo Penitentiary (HP) was selected ($n_{\text{study 1}} = 60$; $n_{\text{study 2}} = 1095$). **Results:** High reliability indices (McDonald's $\omega \geq .73$) and adequate levels of content validity (CVI-S $\geq .87$), construct validity, and criterion validity of the scales were found. This could explain the probability (a) of belonging to the group of inmates with instrumental or impulsive crimes ($R^2_N \geq .52$, $OR \geq 1.02$, $p \leq .021$), and (b) that recidivism, designated by the prison security level imposed, increases ($R^2 \geq .53$, $\beta \geq 1.16$, $p \leq .008$). **Conclusions:** The instrument is a valid and reliable measure that allows a dimensional and integrative assessment of the personality of convicts of the HP, according to ICD-11 standards.

Keywords: Test Development; Personality; ICD-11; Prisoners; Integrative Dimensional Assessment; IDPI-11.

Resumen

Antecedentes: el trastorno de la personalidad (TP) es un predictor importante en la comisión de delitos; sin embargo, existe una ausencia de instrumentos clínicos para las características del convicto peruano. **Objetivo:** desarrollar un instrumento de medición integral de la personalidad confiable y válido, el Inventario Integrativo de Personalidad Dimensional versión CIE-11 (IDPI-11), según los estándares de la Clasificación Internacional de Enfermedades (CIE-11). **Método:** se utilizó una muestra estratificada de reclusos del Establecimiento Penitenciario de Huancayo (EPH) ($n_{\text{estudio 1}} = 60$; $n_{\text{estudio 2}} = 1095$). **Resultados:** se encontraron altos índices de confiabilidad (ω de McDonald $\geq .73$) y niveles adecuados de validez de contenido (CVI-S $\geq .87$), constructo y criterio de sus escalas, pudiendo explicar la probabilidad de: (a) pertenecer al grupo de internos con delitos instrumentales o impulsivos ($R^2_N \geq .52$, $OR \geq 1.02$, $p \leq .021$); y (b) que la tendencia a reincidir, designada por el grado de seguridad penitenciaria impuesto, aumente ($R^2 \geq .53$, $\beta \geq 1,16$, $p \leq .008$). **Conclusiones:** el instrumento construido es una medida válida y confiable que permite una evaluación dimensional e integrada de la personalidad del convicto de la EP de Huancayo, de acuerdo con los estándares de la CIE-11.

Palabras clave: desarrollo de pruebas; personalidad; CIE-11; reclusos; evaluación dimensional integrada; IDPI-11.

Para citar este artículo:

Hualparuca-Olivera, L. M., Ramos, D. N., Arauco, P. A., & Coz, R. M. (2022). Integrative Dimensional Personality Inventory for ICD-11: Development and Evaluation in the Peruvian Correctional Setting. *Liberabit*, 28(1), e540. <https://doi.org/10.24265/liberabit.2022.v28n1.05>

Este es un artículo Open Access publicado bajo la licencia Creative Commons Atribución 4.0 Internacional. (CC-BY 4.0)



Introduction

The new ICD-11 dimensional model for the diagnosis of personality disorder

Personality refers to the characteristic way in which individuals behave; experience life; and perceive and interpret themselves, other people, events, and situations. On the other hand, personality disorder (PD), according to the new ICD-11 model, which is ready to be implemented from the year 2022 (Bach et al., 2022; Mulder, 2021), is a substantial and marked alteration in personal and social functioning (World Health Organization [WHO], 2022a, 2022b).

In the ICD-11, the diagnosis of PD has taken a radical turn concerning the previous nosological model, since it represented a diffuse conceptualization causing problems in the theoretical, practical, and methodological aspects (Bach et al., 2022; Mulder, 2021; Tyrer et al., 2019; Widiger & Oltmanns, 2021), which present in five important deficits. (a) The most frequent diagnoses corresponded to two of the 10 specific diagnostic categories, borderline personality disorder (BPD), antisocial personality disorder (ASPD), and personality disorders not otherwise specified (PD NOS). (b) In addition, the complexity of the nosology of PD meant that any interest was limited to a few specialists, and inexperienced general practitioners or psychologists avoided getting involved (Tyrer et al., 2019; Watts, 2019). (c) Likewise, a strict categorical PD model could not explain the relationship between abnormal and normal personalities. (d) Moreover, there is consistent evidence that severity, rather than specification of personality pathology, is the main predictor of individual suffering and dysfunction (Tyrer et al., 2019). (e) Furthermore, the constant overlap (comorbidity) of specific personality types added to the categorical diagnostic criterion caused deficits in the construct and diagnostic validity of the taxonomic model criteria of the previous classification system.

The diagnosis of PD in the ICD-11 is mainly defined according to its severity with the categories «mild», «moderate,» or «severe» (Mulder, 2021; WHO, 2022a,

2022b) –located in Chapter 06: Mental, behavioral and neurodevelopmental disorders–, and the category «difficulty» (included in Chapter 24: Factors influencing health status or contact with health services). Optionally, personality can be classified according to five trait domains: (a) negative affectivity, (b) detachment, (c) dissociality, (d) disinhibition, (e) anankastia, and the additional qualifier borderline pattern (previously called BPD). With this new diagnostic model, from a theoretical point of view, the conceptualization of psychopathology, the understanding of semiology (Bach et al., 2022), and the epigenetics of PD (Ramos, 2022) are improved; and from a methodological point of view, it allows the development of new procedures, instruments, and evaluation and intervention devices with multidisciplinary work that the new times demand (see, e.g., Christensen et al., 2020; Martin-Key et al., 2022).

From a practical point of view, according to Bach et al. (2022), the communication of non-stigmatizing clinical information between mental health professionals, health service administrators, researchers, patients, and their families is favored; and the differential diagnosis and assignment of the most relevant mental disorders predicting future treatment needs (such as suicide, assault, and morbidity), and the expected degree of disability leave. This is because the main diagnosis, based on severity, allows the detection and early intervention of personality difficulty or the corresponding referral by the primary care psychologist if the anomaly exceeds the clinical threshold (Bach, 2019; Bach et al., 2022; Bach, Somma, et al., 2021; Tyrer & Mulder, 2022). Diagnosis based on optional qualifiers allows mental health specialists to focus on the type of intervention performed (Bach, 2019; Bach et al., 2022; Bach & Simonsen, 2021).

Current measures of the ICD-11 personality disorder model

Existing measures for the evaluation of PD according to ICD-11, –as also detailed in Mulder & Bach (2022)–, are based on self/hetero-informed

instruments and semi-structured interviews as following, see Appendix A (<https://doi.org/10.24265/liberabit.2022.v28n1.05>) for more details:

(1) Psychometric studies of item pool development, based on ICD-11 Clinical Descriptions and Diagnostic Requirements (CDDR), to generate measures include: (a) for the severity of personality dysfunction, the ICD-11 Personality Disorder Severity Scale (PDS-ICD-11; Bach, Brown, et al., 2021) and two domain scales from the ICD-11 Personality Disorder Model (ICD-11 PD Model scales; Clark et al., 2021); (b) for the trait domain qualifiers, five domain scales from the ICD-11 PD Model (Clark et al., 2021) and the Personality Inventory for ICD-11 in its self-report version (PiCD; Oltmanns & Widiger, 2018) and its hetero-informed version (Informant-Personality Inventory for ICD-11 [IPiC]; Oltmanns & Widiger, 2021); and finally (c) for the additional borderline pattern qualifier, the Borderline Pattern Scale (BPS; Oltmanns & Widiger, 2019).

(2) Research that consider the assembly of items from other measures to develop scales include the following: (a) for the severity of personality dysfunction, the Standardized Assessment of Severity of Personality Disorder (SASPD; Olajide et al., 2018); (b) for the trait domain qualifiers, the Personality Inventory for DSM-5-Brief Form-Plus (original) (PID-5BF+ [original]; Kerber et al., 2022), the Personality Assessment Questionnaire for ICD-11 (PAQ-11; Kim et al., 2021), the Personality Inventory for DSM-5-Brief Form scored for ICD-11 (PID-5-BF for ICD-11; Bach & El Abiddine, 2020), the Five-Factor Inventory for ICD-11 (FFiCD; Oltmanns & Widiger, 2020), the Personality Inventory for DSM-5 scored for ICD-11 (revised) (PID-5 for ICD-11 [revised]; Sellbom et al., 2020), the Personality Inventory for DSM-5 scored for ICD-11 (original) (PID-5 for ICD-11

[original]; Bach et al., 2017), and the Personality Assessment Schedule for ICD-11 (PAS-ICD-11; Tyrer, 2017); and (c) for the additional borderline pattern qualifier, an 8-item scale based on an algorithm from other PAQ-11 scales (Kim et al., 2021).

(3) Finally, studies that consider the validation of a measure of Criterion A (level of personality functioning) in the Alternative Model of Personality Disorders of the DSM-5-TR (DSM-5 AMPD; American Psychiatric Association, 2022), taking as criterion the severity of personality dysfunction proposed in the ICD-11, include the Self and Interpersonal Functioning Scale (SIFS; Gamache et al., 2021), the Semi-Structured Interview for Personality Functioning DSM-5 (STiP 5.1; Hutsebaut et al., 2021), and the Level of Personality Functioning Questionnaire Screener (LoPF-Q Screener; Zimmermann et al., 2022).

According to Morgado et al. (2017), there are 10 main limitations of test development studies: (a) *sample characteristic limitations* that include convenience sampling OR small sample size (< 1:10); (b) *methodological limitations* that include cross-sectional methodology OR self-reporting methodology OR web-based survey; (c) *psychometric limitations* that include lack of a more robust demonstration of the construct validity/reliability OR inadequate choice of the instruments/variables to be correlated with the variable of the study OR factor analysis errors; (d) *qualitative research limitations* that include lack of deductive approach to scale development OR lack of a more robust literature review OR subjective analysis OR content validity was not formally assessed OR lack of recruitment and training of a qualified number of interviewers; (e) *missing data* that includes the absence of enough information about grouped OR added OR rounded OR censored OR truncated numbers; (f) *lack of detection of invalid responses*¹

¹ It was originally mentioned as «social desirability bias» (Morgado et al., 2017); however, the term has been modified to encompass the entire spectrum of distorted responses.

that refers to not administering and measuring validity scales; (g) *item limitations* that include ambiguous OR not reversed items; (h) *brevity of the scale* that refers to include a few quantity of items if Cronbach alpha is used to test reliability; (i) *difficulty to control all variables* that refers to those important variables that were not considered within the construct; and finally (j) *lack of manualized instructions* that includes not to consider instructions on application conditions OR administration mode OR response mode OR rating mode OR giving an example.

Taking into account the above, it has been reviewed that the three main limitations for the development of the ICD-11 PD nosology measures are, first of all, *sample characteristic limitations* (94.1%); secondly, *methodological limitations* (94.1%); and, thirdly, *lack of manualized instructions* (88.2%). Even though cross-cultural validation and adaptation studies of these instruments carried out to date have mostly supported the applicability of the new ICD-11 PD diagnostic nosology (Ayinde & Gureje, 2021), as mentioned by Sleep et al. (2021), the current severity instruments of personality functioning show fragile factorial structures with low discriminant validity. In addition, it has been shown that all these instruments contain ordinal response formats with four or more response options, which increases the resistance and fatigue of examinees, and reduces the possibility of controlling and/or measuring their biased answers, thereby increasing systematic error (Moran, 2021). In addition, it has been noted that these instruments have been developed and/or validated mainly in European and North American populations, which have characteristics that are very different from those of underdeveloped countries. These limitations could affect the validity and replicability of the psychometric results found in different settings and in culturally diverse groups, in which their usefulness may be questioned (Sleep et al., 2021).

The need for a comprehensive measure of ICD-11 PD in correctional settings: a science-practice gap

Personality assessment for mental health purposes in prison settings has four main functions: (1) *detection and diagnosis*, for example, to find out if an inmate is depressed and at risk of self-harm; (2) *prognosis*, e.g., to predict whether inmates will present a significant risk to others once they are released; (3) *case conceptualization and treatment selection*, e.g., to understand the severity of the PD and its management; and (4) *treatment monitoring and follow-up*, e.g., to assess whether the rehabilitation program –whether group or individual– is being, or has been, effective in reducing violence risk (Day & Cook, 2019).

The evaluation of PD in prisons also relies on five legal purposes: a) competency to stand a trial, (b) criminal responsibility, (c) dangerousness, (d) pre-sentence, and (e) risk and recidivism assessments. The competency to stand a trial assessment addresses inmates' current state of mind and whether they can understand their charges to assist their attorney in defense (Ben-Porath et al., 2022; Butcher et al., 2015). Criminal responsibility assessments address the prisoner's mental state at the time of committing an offense with intent, recklessness or negligence (Sellbom et al., 2022). Dangerousness assessments are conducted on processed inmates after dismissing criminal liability but accepting a civil one; they seek to determine whether to transfer the inmate to a forensic unit (i.e., a secure environment), a civilian psychiatric unit in the community, or if outpatient treatment is necessary (Butcher et al., 2015; McGrath & Turvey, 2013). Pre-sentence assessments are performed in inmates between the period of adjudication of guilt but before sentencing so that the judge may consider the mitigating factors when deciding on a sentence and/or incorporating mental health needs in the prison treatment plans (Ben-Porath et al., 2022). Risk and recidivism assessments are conducted in inmates to predict future violent behavior –e.g., self and hetero-

aggression— and address their security needs, often aimed at prison classification and reclassification (Cirlugea et al., 2013; Mia et al., 2020; Toop et al., 2019).

Given the above, it is understandable that mental health in prison involves psycho-legal issues within criminal law, since PD, added to the current psychopathology of addiction or other issues, is a predictor of the commission of crimes, current behavior in prison, and possible recidivism of the prisoner once free. These components (moreover, the last one) are the main prognostic foci in the evaluation of the prisoner (Day & Cook, 2019). This is because the average incarcerated person has been widely shown to have a severe personality disorder, with actual and lifetime prevalence rates of PD diagnoses of 40-88% (Hutsebaut et al., 2021).

With the ICD-11 PD model, to date, only two personality measurement instruments have been validated in a few participants from prison populations (85 and 30 inmates, respectively; see Bach & Anderson, 2020; Hutsebaut et al., 2021). However, because the characteristics of this population are not considered in the design of these instruments (Meaux et al., 2021), several factors can affect the reliable evaluation of personality functioning in prison settings. Lack of morality and capacity for reflection are internal factors related to the severity and type of predominant personality pathologies, and social desirability and deception are external factors related to the specific context (Hutsebaut et al., 2021).

As some authors state (see Abdalla-Filho, 2022; Bartlett, 2011), the ICD-11 PD nosology and therefore its measures, developed from the semiology of clinical patients, do not apply to the characteristics of patients with PD who are involved in legal issues since the very present deficits and cognitive deficiencies of inmates with this diagnosis are not taken into account. Likewise, importance is not given to the comorbidity (differential diagnosis), etiology, and natural life course of the disorder if it is not

treated, including the commitment to security within the penitentiary (Abdalla-Filho, 2022). Another disadvantage is that, in contrast to what Bach et al. (2022) stated and as Halliwell (2021) already warned, the ICD-11 PD nosology does not exempt the risk of stigmatization. This has just been demonstrated in the jurisprudence, in a recent study for criminal responsibility assessment, in which the borderline pattern of the ICD-11, qualified with serious severity, is seen as «the mad» by the juries (Baker et al., 2021). Due to the aforementioned, there is a need for measures for the diagnosis of PD in legal settings to be convincing, adequately specify the severity of the PD, establish if there is a connection with a crime, and provide guidelines for the adequate management of the diagnosis, all of which considering the perspectives of rehabilitation and community protection (Carroll et al., 2022).

The measures that have been used to assess the personality of prison populations show that, internationally, approximately 65% of male convicts have a PD and conviction rates increase the severity of PDs, especially psychopathy. This specific construct is configured within PDs and appears to be an important predictor of subsequent violence, crime, and recidivism (Hengartner et al., 2018). In the national context, although research is limited, it was found that 17.1% of convicted sex offenders at the Lurigancho Penitentiary have a mental disorder, including PD (Sindeev & Guzmán-Negrón, 2019), while more than 50% of the inmates of the Arequipa Penitentiary show psychopathic deviation and 38.4%, antisocial disorder (Arias et al., 2016; Arosquipa & Gutiérrez, 2016). Moreover, at the local level, antisocial disorder (55.8%) and narcissism (51.2%) are the most prevalent PD diagnoses among the inmates of the DeVida program at the HP (Gomez, 2018). With the new ICD-11 PD nosology and its developed measures, these numbers will likely increase, given its greater specificity (Halliwell, 2021).

In Peru, it is difficult to use imported clinical instruments that measure PD —e.g., the Minnesota

Multiphasic Personality Inventory (MMPI-3; Ben-Porath & Tellegen, 2020b), the Personality Assessment Inventory (PAI; Morey, 2007), or the Multiaxial Clinical Inventory of Millon (MCMI-IV; Millon et al., 2015)– because they do not fit to the special features of the level of education, culture, and/or statistical disposition of the prison population (Carlson, 1981). Despite the fact that the authors of these measures have elaborated interpretation guides for the evaluation of personality in prison environments, and although the psychometric evidence –questionably– supports said measures (Neal et al., 2022), the items have been formulated based on the specific characteristics of clinical populations so that they can hardly be adjusted to those of the inmates of the Peruvian correctional system.

In summary, from a practical perspective, although the current clinical personality measurement instruments are relatively complete, they do not fit the educational, cultural, and statistical characteristics of Peruvian inmates. In addition, since the implementation of the WHO nosological classification is near in all state systems in which health professionals work, there is no adequate instrument for the ICD-11 PD model, which was designed from and for the evaluation of convicts. Moreover, from a theoretical perspective, the current personality assessment measures do not have a broad theoretical and substantive basis that understands the origin, current state, or the very sense of being of the personality from a continuous spectrum between health and disease. In this sense, it is necessary to develop a comprehensive measure of personality that meets the standards proposed in the ICD-11, and that, in order to extend its practical potential, incorporates solid theoretical support on said construct as a complex and singular entity.

Purpose of this research

To address this gap, this research aimed to develop and evaluate a comprehensive measure of personality from a mental health perspective in a

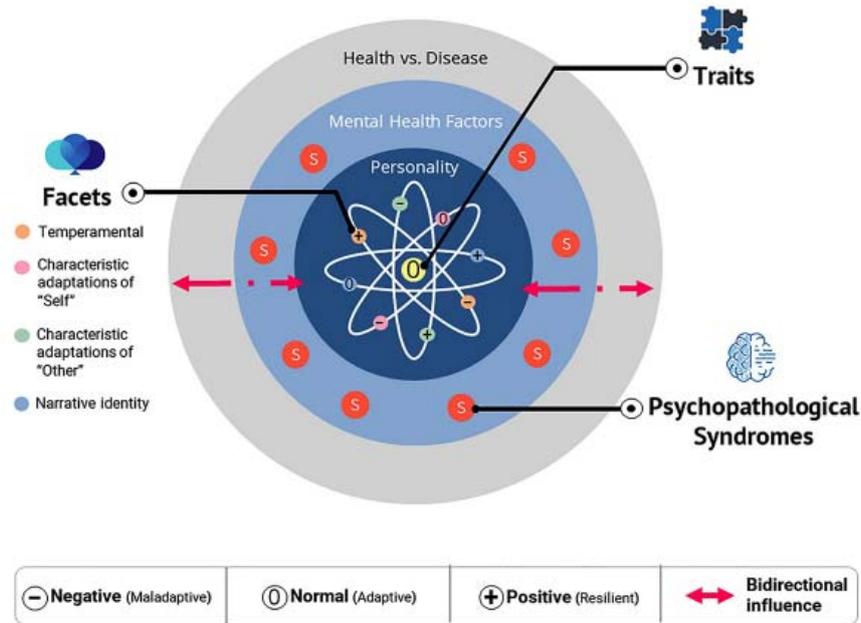
correctional setting, which had to be adjusted to the most up-to-date standard proposed by the WHO (ICD-11; WHO, 2022a) for the diagnosis of PD.

This measure is based on an integrating paradigm of seven personality approaches with the greatest scientific impact throughout history: (a) *evolutionary*, focused on biopsychosocial adaptation (neurodevelopment); (b) *multivariate*, focused on heritable and stable dispositional traits; (c) *psychodynamic*, focused on the dynamism of the unconscious; (d) *interpersonal*, focused on meaningful social interaction; (e) *personological*, focused on narrative identity; (f) *empirical*, focused on correlates with personality disorders; and (g) *salutogenic*, focused on the healthy spectrum of personality. From these paradigms, the theory, five explanatory and evaluation models of the IDPI-11 were generated. One of these models, the *integrative dimensional assessment of personality model*, is shown in Figure 1.

Integrating the theoretical contributions of McAdams and Pals (2006) and De la Iglesia and Castro (2018), personality is defined as «an individual's unique variation on the general evolutionary design for human nature, expressed as a developing pattern of dispositional traits, characteristic adaptations, and self-defining life narratives, complexly and differentially situated in culture and social context». Based on this definition, the integrative dimensional theory of personality is elaborated.

To identify individual differences, one must first understand the similarities among all living things. Thus, from the evolutionary paradigm, three evolutionary strategies are established for living beings: *survival*, *adaptation*, and *replication*. These strategies are homologous to those of human beings which include *existence*, *agency*, and *transcendence*. Other paradigms complement the understanding of personality by focusing on one or more evolutionary strategies.

Figure 1
Integrative dimensional assessment of personality model



Note: Both facets and traits can be positive, normal, or maladaptive. The red arrow indicates a bidirectional influence between layers.

The *multivariate paradigm* focuses on the existence to try to explain the personality of individuals from the question «What am I?». Thus, the *actor self* responds, «I am what I do», referring to the biological and environmental influence traits that qualify the concrete and stable trend of behavior. The *interpersonal paradigm* focuses on the agency to explain the personality from the question «How am I?». And the response from the *agent self*, «I'm how I do», describes that the motivations (composed of goals, strategies, and interpretations), determine the behavior of the individual in specific significant situations of the historical-cultural context. Thus, several people with the same trait can act differently because of their personal motivations, or the same person can act in a similar situation differently because of new motivations.

The *narrative paradigm* focuses on the transcendence to explain the personality from the question «What am I for?» and the response from

the *author self*, «I am to give value to my existence», refers to the fact that people value and narrate their history and project it into the future to give consistency to their identity and meaning of life. The *psychodynamic paradigm* focuses on the agency and transcendence to explain, from defense mechanisms and object relations, the dynamism underlying the motivations and meaning of life. Likewise, the *salutogenic* and *empirical paradigms* are integrated into each of these three strategies to explain personality in a continuum of health and disease.

With the models and methods extracted from these paradigms that make up the core of the IDPI-11, personality can also be assessed from its positive range (see De la Iglesia & Castro, 2018), and within the context of mental health, including factors associated with or in conjunction with the most prevalent psychopathology in the Peruvian context (see National Institute of Mental Health, 2019).

Methodological design

This research is a psychometric study because it includes the development of new instruments and appropriate methods for scoring, establishment of reliability and validity analyses for the measures, examination of the properties of the items and scales and/or their dimensions, and evaluation of the differential functioning of the items between subgroups (VandenBos, 2015). In psychometrics it is known that one of the approaches for developing measures uses a deductive process followed by an inductive one (Boateng et al., 2018). The deductive process is based on an iterative literature review of paradigms, theories, and models to formulate and delimit the construct (or constructs), dimensions, indicators and items, which, after being evaluated semantically and theoretically, follow an inductive process based on statistical refinement and the subsequent estimation of their properties so that their norms can be generalized (Kuhn, 2012).

In this sense, to address the objective of the development and evaluation of the measure, this research was divided into a pre-empirical phase and two phases of empirical studies, which together consisted of 13 steps. *Preliminary phase* included (1) identification of constructs and generation of items, and (2) item revision by a panel of five experts. *Study 1: item refinement* included (3) item calibration (selection) with classical test theory (CTT) and item response theory (IRT). *Study 2: development, evaluation and normalization of the IDPI-11 scales* included (4) item designation for the floating scales, (5) formulation of differentiated item scores for scales according to their prototypicality, (6) content validity of scales, (7) item-level analysis from CTT and IRT², (8) structural validity of scales of the Traits subgroup,

(9) convergent and discriminant validity of the scales within their own groups, (10) validity of differentiation by groups of the scales, (11) criterion validity of Personality group scales, (12) reliability of scales, and (13) development of norms for the scales raw scores³. All the phases of this research address the development purpose and, in the last phase, the evaluation aim of the measure. Likewise, the sequential steps represent the specific objectives of each study (see Table D.1 in Appendix D for full details).

To develop the IDPI-11 scales and their corresponding items, a combination of theoretical and statistical test development methods was used: (a) the rational theoretical method for all the IDPI-11 scales, (b) the item response theory (IRT) for item calibration, (c) a sequential system for the development of construct-oriented scales for the elaboration of the scales of the IDPI-11 Traits subgroup, (d) content grouping with statistical refinement, and (e) a priori designation (Ben-Porath & Tellegen, 2020a; Williams et al., 2019). The *rational theoretical method* is based on the judgment of the creators of the instrument, and their understanding of psychopathology, for both the elaboration and selection of items. This method assumes a theory-based relationship between item content and the assessed personality attributes, giving each item face validity.

Likewise, the *sequential system for the development of construct-oriented scales* uses a modification of the exploratory factor analysis (EFA) to find, based on the exclusion of some items in the analysis, well-differentiated latent constructs which are adjusted to what is required by the theory. The *IRT*, on the other hand, allows refinement of the scales through calibration (selection of items)

² The factor analysis approach from the CTT and IRT approach is used not only for the refinement of the measure but also for complementing the structural validity (internal structure) of the scales to report the individual properties of each item in the evaluation phase (Bach, Brown, et al., 2021; Nguyen et al., 2014).

³ Some authors consider norms development as part of the evaluation phase of the measure (see, e.g., Boateng et al., 2018). In study 2 it is placed as an independent section for its better understanding.

according to the level of the trait to be measured and the probability of supporting each item. Moreover, *content grouping with statistical refinement*, involves grouping items according to similar theoretical content, verifying the belonging of the item to the scale or the interassociation between individual items using correlational statistical methods (e.g., Spearman's rho, polychoric correlations [ρ], tetrachoric correlations [r_{tc}], Cronbach's α or CFA). Finally, *a priori designation* is a method for generating a scale without defined scoring content (Ben-Porath & Tellegen, 2020a). Table D.1 of Appendix D shows the methods used to develop each of the IDPI-11 scales by steps and phases.

Preliminary development of the IDPI-11 scales

Identification of constructs and generation of items

For the identification of the construct, a vast literature was reviewed concerning the generation of new diagnostic models of PD based on the DSM-5 AMPD (see Cain & Mulay, 2022) and the ICD-11 (see Bach & Roger, 2022), focusing more on the last one because it is a global, intersectoral and multidisciplinary standard. Issues concerning the reasons for the new versions, theories that supported them, discussions and scientific debates around these classification systems were also reviewed. As a result, it was intended to incorporate mental health constructs to enable a better understanding of personality since, for example, Morey (2019) explained the importance of having instruments that are not only parsimonious but also complete and practical, as commercial instruments do well for clinical evaluation of personality. Also important was the inspiration from the work of Widiger (2016) and De la Iglesia and Castro (2018) who focused on the positive variants of maladaptive personality traits. In this sense, some authors were asked to share their instruments to thoroughly review the method of writing the items and their organization to measure the personality constructs. General psychopathology

instruments were also reviewed, focusing on the constructs of psychopathology and mental health factors more frequently in the Peruvian population. Similarly, upon identifying the lack of test development (not validation) studies in the correctional population, we decided to implement our study in that setting.

By applying the *rational theoretical method* with some considerations –as explained in Table D.2 of Appendix D– the constructs and items were generated and constantly revised, eliminated, added, and reformulated as new relevant information was obtained. In this way, the items were ready and organized with their respective indicators and dimensions: IDPI-11 «solid scales» which included R3, R4, F1, F2, F3, F4, F5, F6, F7, F8, S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, a.1, a.2, a.3, a.4, a.5, a.6, a.7, b.1, b.2, b.3, c.1, c.2, c.3, d.1, d.2, d.3, d.4, e.1, e.2, and e.3. These were organized into groups and subgroups to facilitate the interpretation of the instrument in its final version. Likewise, taking a dimensional perspective of the personality constructs and integrating them into the salutogenic model, items with a positive or resilient connotation –strengths and virtues of personality (Peterson & Seligman, 2004)– were generated and accompanied the items with a maladaptive meaning to compose each of the scales of the Facets of Traits group. Then, those items were chosen and, due to their meaning, could belong to new scales without their own items. IDPI-11 «Floating scales» which included R1, R2, PF, A, B, C, D, E, and PL. These scales could not only be part of the personality construct but also part of others that may help to evaluate the biased responses of those evaluated as instruments already used in the prison system (see Table 4 to view the names of each of these scale codes). It should be noted that the floating scales PF A, B, C, D, and E were also composed of positive meaning items and maladaptive meaning items. The PL scale only included maladaptive items from the trait facet scales, given its categorical nature, similar to other scales of the Psychopathology group.

Appendix C provides a brief description of each of the IDPI-11 scales.

Finally, the items were read and reviewed by a 68-year-old male who had only completed the third grade of primary school. Thanks to his comments, these items were restructured so that they were clear and understandable for the target prison population. Also, the pertinent administrative, ethical and logistical procedures were already being carried out to be able to evaluate said population.

Item revision by a panel of experts

The whole instrument was validated by a panel of five experts: the number of experts recommended according to Zamanzadeh et al. (2015). One of these experts is currently working in the penitentiary system, with approximately 10 years of practice, and four of them have similar average experience in the field of clinical psychology. The instrument developed is especially focused on clinical and mental health constructs, which is why most of the panel of experts belong to this area. However, as the instrument was standardized in a prison sample, it was pertinent to complement it, at least, with the validation of experts in that area to obtain a valid and reliable instrument. The experts were asked to review the items regarding the criteria of «clarity». After this review, and after an analysis of the content validity index of each item, satisfactory results were obtained, since 222 items reached item content validity indices (CVI-I) in the range of .80 to 1; therefore, 59 items were reformulated taking into account the observations provided. Likewise, it is important to mention that no item was eliminated in this step, since all 222 items had a CVI-I equal to or greater than .80 and item content validity ratios (CVR-I) greater than .60. Although the cut-off criteria may vary depending on the size of the panel of reviewers, in general, Almanasreh et al. (2019) indicated the following: CVI-I = .78 to 1, to be retained; CVI-I < .78, to be eliminated; and CVR-I = .60 to 1, to be retained, and CVR-I < .60, to be eliminated. These 222 items,

which were already reviewed, were randomly reordered to be administered to participants of study 1.

Study 1: item refinement

Methods

The main objective of the first study was to select parsimonious items from the CCT and IRT to be kept in the composition of the IDPI-11 scales. Although brief in time and austere in sample size, the first application of the instrument is important to predict and meet psychometric needs in a subsequent larger study. It also provides an opportunity to learn about the potential limitations in the examinee's performance in the test and the logistical complications.

Participants

Through a random sampling stratified by type of crime, 70 inmates were selected from the HP, a male penitentiary located east of the city of Huancayo, assuming a loss rate of 10% after evaluating the eligibility criteria, which excluded the cases with 10 or more blank or null answers and/or some disability that limited the willingness to give consent at the time of the evaluation. The final sample of this study consisted of 60 inmates with an average age of 38.3 years and an age range between 26 and 49 years (81.7%). In addition, it was shown that most inmates had a cohabiting partner (20%), did not finish secondary school (35%), came from the Peruvian highlands (65%), were located in pavilion P2 (33.3%), were sentenced to a minimum security facility (38.3%), were detained for crimes against public administration (38.7%), have spent from 4 to 12 years in the HP (23.3%), and were serving a prison sentence also from 4 to 12 years (58.3%). See Table 1 for more details on this sample.

Measures

Integrative Dimensional Personality Inventory-II (IDPI-II). This measure was developed for the

present study and consisted of 222 false/true (F/T) initial items, distributed in 49 scales⁴: 4 Response Styles scales, 8 Mental Health Factors scales, 7 Personality scales, 10 Psychopathology scales, and 20 Trait Facets scales. The instrument had a booklet of questions with relevant instructions. It also had an answer sheet that included a section for sociodemographic data (date of birth, marital status, level of education, and place of origin) and another one to fill in the answers to the items.

Procedure

The data collection was performed during the first week of April 2021, meeting the respective ethical and logistical procedures. To carry out the data collection, permission from the Ethics and Research Committee of the Faculty of Humanities (CEI-DD-HH) of the Continental University, as well as the Directorate of Penitentiary Treatment of the National Penitentiary Institute of Peru - Junín Region, was obtained. Then, the head of the psychology department was trained in the procedures for administering the instrument and he, in turn, trained his personnel, since contact was restricted in the prison area due to the COVID-19 pandemic. Inmates sat at a (personal) desk with a pencil, blue or black pen, and eraser.

The informed consent and instrument sheets were available on the desk. Next, the psychologist monitored the entire administration procedure of the instrument. In addition, he was in charge of excluding the inmates who did not have the decision-making capacity to sign the informed consent and complete the questionnaire. Likewise, the psychologist signed the informed consent form as a witness for each of the examinees who participated in the study. The completed study 1 questionnaires and informed consent forms were collected from the prison on the day after the last day

of the week. Data collection for study 1 was carried out in two steps in an average total time of 36 minutes to avoid examinees to get tired.

Furthermore, the techniques used for data collection were: (a) a survey that consisted of the application of the instrument and (b) a review of secondary data from a database with information on the type of crime, prison security level imposed, and other data (age, National Identity Card number, pavilion, time served in prison, and length of sentence). Although not all were analyzed inferentially, they are presented in Table 1 for a better understanding of the characteristics of the sample in this study. In addition, Microsoft Excel v. 2019, IBM SPSS Statistics 28.0.1.0, and RStudio v. 2021.09.1+372 were used for data processing and analysis. Furthermore, missing data were imputed with the abovementioned second software using the multiple imputation method.

Data analysis

For the descriptive statistics of the sociodemographic characteristics of the sample, the mean age and mode of the rest of these characteristics were calculated using IBM SPSS Statistics. To select the items from the classical test theory (CTT) approach, Spearman's rho statistics were analyzed to relate each item to its original scale using IBM SPSS Statistics. The cut-off point of rho $\geq .70$ was used to retain the item, since it implies a strong correlation (Akoglu, 2018). In addition, using Microsoft Excel, the items difficulty index (p) and discrimination index (D) were analyzed, following the algorithm of Reynolds and Livingston (2019). The cut-off points used were: (i) for items difficult to endorse by the participants –including all «solid scales,» except for scales F1, F2, F3, and F4– the cut-off points were $p < .50$ and $D \geq .30$ to retain the item

⁴ It is important to emphasize that the IDPI-11 has two types of scales: (a) «solid scales,» which have their own items (prototype items) and items from other scales (non-prototype items), and (b) «floating scales,» which only have items from other scales (non-prototype items). The responses to the items are recorded in a dichotomous format F = 0 and T = 1 which, at the time of scoring, are configured in two types of items (prototype item = 2 points, non-prototype item = 1 point), as they correspond to each of the IDPI-11 scales.

and (ii) for items theoretically easy for participants to endorse –including items of scales F1, F2, F3, and F4– $p = .30$ to $.70$ and $D \geq .30$ were used to retain the item (Reynolds & Livingston, 2019).

Furthermore, parameters b and a of the item characteristic curve from the IRT approach were also used in this step with the RStudio «psych» package. For items that were theoretically difficult to ratify –including items of all solid scales, except for scales F1, F2, F3, and F4– the cut-off point of

$b \geq 0$ was used to retain these items because they measure psychopathological constructs (Nguyen et al., 2014; Yang & Kao, 2014). In addition, another threshold of $b = -2$ to 2 was used to retain the item because these items were theoretically not difficult or easy for participants to endorse –including items of scales F1, F2, F3, and F4–. For the item discrimination parameter across all items of «solid scales,» the cut-off point of $a \geq 0$ was used to retain them (Nguyen et al., 2014; Yang & Kao, 2014).

Table 1
Sociodemographic characteristics of study 1 and study 2 samples

Sociodemographic Characteristics	Study 1		Study 2	
	<i>n</i>	%	<i>n</i>	%
Age ^a				
18-25 years	4	6.7	110	10
26-49 years	49	81.7	866	79.1
50-85 years	7	11.7	119	10.9
			<i>M</i> (study 1) = 38.3; <i>M</i> (study 2) = 38.4	
Marital status				
Single	5	8.3	160	14.6
Married (first marriage)	2	3.3	50	4.6
Married (second marriage or more)	6	10	118	10.8
Separated	9	15	110	10
Facing a divorce process	7	11.7	129	11.8
Divorced	10	16.7	180	16.4
Cohabiting	12	20	249	22.7
Widowed	9	15	99	9
			<i>Mo</i> (study 1) = «Cohabiting»; <i>Mo</i> (study 2) = «Cohabiting»	
Education				
Complete primary education or below	20	33.3	419	38.3
Incomplete secondary education	21	35	427	39
Complete secondary education	9	15	150	13.7
Technical higher education	7	11.7	59	5.4
University undergraduate or graduate program	3	5	40	3.7
			<i>Mo</i> (study 1) = «Incomplete secondary education»; <i>Mo</i> (study 2) = «Incomplete secondary education»	

Sociodemographic Characteristics	Study 1		Study 2	
	<i>n</i>	%	<i>n</i>	%
Population's origin				
Coast	10	16.7	209	19.1
Highlands	39	65	677	61.8
Jungle	10	16.7	169	15.4
Abroad	1	1.7	40	3.7
<i>Mo</i> (study 1) = «Highlands»; <i>Mo</i> (study 2) = «Highlands»				
Pavilion				
P1: Minimum security	17	28.3	313	28.6
P2: Medium security	20	33.3	359	32.8
P3: Maximum security	17	28.3	312	28.5
P4a: CREO program	1	1.7	23	2.1
P4b: PLANPAM	1	1.7	14	1.3
P5: Psychiatric	2	3.3	37	3.4
P6: DeVida Program	2	3.3	37	3.4
<i>Mo</i> (study 1) = «P2: Medium security»; <i>Mo</i> (study 2) = «P2: Medium security»				
Prison security level				
Minimum	23	38.3	389	35.5
Medium	20	33.3	428	39.1
Maximum	14	23.3	248	22.6
Special Close	3	5	30	2.7
<i>Mo</i> (study 1) = «Minimum»; <i>Mo</i> (study 2) = «Medium»				
Type of crime				
Against the family	1	1.6	10	.9
Against public trust	1	1.6	20	1.8
Against production and distribution of counterfeit money	2	3.2	10	.9
Against public administration	24	38.7	10	.9
Against freedom	6	9.7	438	40
Against public tranquility	8	12.9	10	.9
Against public health	17	27.4	169	15.4
Against life, body, and health	1	1.6	110	10
Against property	1	1.6	308	28.1
Against public security	1	1.6	10	.9
<i>Mo</i> (study 1) = «Against public administration»; <i>Mo</i> (study 2) = «Against freedom»				

Sociodemographic Characteristics	Study 1		Study 2	
	<i>n</i>	%	<i>n</i>	%
Time served				
Less than 1 month	4	6.7	50	4.6
1 to 6 months	12	20	208	19
7 months to 1 year	13	21.7	240	21.9
1 to 3 years	9	15	228	20.8
4 to 12 years	14	23.3	259	23.7
13 to 24 years	3	5	40	3.7
25 or more years	5	8.3	70	6.4
<i>Mo</i> (study 1) = «4 to 12 years»; <i>Mo</i> (study 2) = «4 to 12 years»				
Length of sentence				
1 to 3 years	6	10	150	13.7
4 to 12 years	35	58.3	535	48.9
13 to 24 years	7	11.7	199	18.2
25 or more years	5	8.3	70	6.4
Life imprisonment	6	10	121	11.1
Undefined	1	1.7	20	1.8
<i>Mo</i> (study 1) = «4 to 12 years»; <i>Mo</i> (study 2) = «4 to 12 years»				

Note: $N = 2009$ ($n_{\text{study 1}} = 60$; $n_{\text{study 2}} = 1095$); M = mean; Mo = mode. Central tendency statistics were observed: mean in the case of quantitative variables and mode in the case of qualitative variables.

^a This variable was transformed into a qualitative ordinal variable to facilitate its understanding.

Results

After conducting the analyses for the selection of items in study 1, 32 items were eliminated for not reaching adequate levels in all calibration indices (see Table 2). Note, for example, that item 7 «Solo por las mañanas siento que me falta el aire» obtained adequate levels of item-scale correlation ($\rho = .76$) and discrimination from the TCT ($D = .31$) (Akoglu, 2018; Reynolds & Livingston, 2019), but the other calibration indices had inadequate levels ($p = .33$; $b = -8.99$; $a = -.15$) (Nguyen et al., 2014; Reynolds

& Livingston, 2019; Yang & Kao, 2014). Another example is in item 8 «Actualmente, me siento terriblemente deprimido y triste la mayor parte del tiempo», which measures a psychopathological construct. It obtained an adequate level of difficulty indices from the TCT and IRT ($p = .52$; $b = .14$) (Nguyen et al., 2014; Reynolds & Livingston, 2019; Yang & Kao, 2014), but its other indices were inappropriate ($\rho = .67$; $D = .09$; $\alpha = -.01$) (Akoglu, 2018; Nguyen et al., 2014; Reynolds & Livingston, 2019; Yang & Kao, 2014).

Table 2
Statistics of items eliminated from the IDPI-11 after the refinement

N.º	Item	Scale	<i>rho</i>	<i>p</i>	<i>D</i>	<i>b</i>	<i>a</i>
16	No suelo tratar tan mal a las personas.	R3	.23	.22	.14	-9.98	-.47
96	Tengo la solución a todos los problemas.	R3	.42	.83	.23	-6.46	-.25
100	No soy tan malo como dicen los demás.	R3	.52	.13	.16	-11.40	-.37
178	Mis notas siempre fueron mayores a 18.	R3	.65	.76	.25	-12.70	-.18
214	Soy más trabajador que los demás.	R3	.31	.25	.17	-10.90	-.28
7	Solo por las mañanas siento que me falta el aire.	R4	.76	.33	.31	-8.99	-.15
48	En ocasiones no recuerdo mi nombre.	R4	.35	.42	.26	-10.20	-.28
56	Me duele todo el cuerpo, incluso hasta las puntas de mi cabello.	R4	.42	.46	.16	-13.30	-.13
137	Siento desesperación cuando me ignoran.	R4	.31	.21	.15	-12.40	-.25
190	Cuando recuerdo lo sucedido durante el día, tengo pesadillas.	R4	.14	.36	.10	-9.72	-.23
28	Bebo por lo menos 2 litros de agua cada día.	F1	.25	.89	.13	-9.59	-.17
145	<i>Pienso que soy un «bueno para nada».</i>	F2	.41	.45	.21	-8.73	-.15
9	<i>Mis padres me trataban muy bien cuando era niño.</i>	F5	.37	.27	.25	-12.50	-.14
5	Últimamente, me siento sin energía para hacer mis cosas.	F7	.42	.41	.16	-13.30	-.25
150	<i>Cuento con varias personas que me apoyan cuando tengo problemas.</i>	F8	.48	.97	.18	-8.14	-.21
8	Actualmente, me siento terriblemente deprimido y triste la mayor parte del tiempo.	S1	.67	.52	.09	.14	-.01
221	Sudo y tiemblo cuando estoy entre muchas personas.	S6	.54	.26	.15	-.30	-.36
180	Me han dicho muchas veces que no se me entiende cuando hablo.	S7	.43	.43	.26	-6.68	-.25
64	Tengo ideas que circulan tanto en mi cabeza que no me dejan tranquilo.	S10	.69	.35	.16	-10.40	-.14
110	Tengo mucho miedo de que la persona a quien quiero me abandone.	a.1	.31	.24	.19	-16.50	-.17
121	<i>Me siento bien cuando estoy con otras personas.</i>	a.1	.43	.17	.16	-14.10	-.60
134	<i>Me siento seguro de mí mismo a pesar de los defectos que tengo.</i>	a.6	.27	.14	.25	-11.10	-.23
146	Me siento «menos» que los demás porque tengo muchos defectos.	a.6	.53	.93	.48	-10.40	.16
41	Evito contar cosas de mi vida para no tener problemas.	a.7	.68	.34	.23	-3.63	-.17
207	Soy una persona que no expresa alegría.	b.2	.37	.26	.12	-9.21	-.13
44	Si tengo alguna queja, lo digo respetuosamente.	b.3	.39	.78	.19	-15.20	-.15
195	<i>Respeto las decisiones de los demás.</i>	c.1	.58	.36	.18	-12.60	-.19
35	<i>Me agrada abrazar a las personas que quiero.</i>	c.1	.62	.14	.06	-11.90	-.16
167	Suelo actuar de manera impulsiva.	d.1	.48	.24	.13	-8.13	-.41
183	<i>Cumplo con mis trabajos, ya que puedo concentrarme fácilmente.</i>	d.2	.54	.25	.24	-6.68	-.34
216	Se me dificulta cumplir mis trabajos, ya que me distraigo fácilmente.	d.2	.67	.15	.18	-7.43	-.21
101	<i>Es más importante aprender a valorar el esfuerzo que buscar buenos resultados.</i>	e.1	.71	.87	.44	-4.61	.15

Note: *rho* = Spearman's *rho*; *p* = difficulty index (TCT); *D* = discrimination index (TCT); *b* = difficulty parameter (TRI); *a* = discrimination parameter (TRI); R3 = Dissimulation; R4 = Simulation; F1 = Healthy Habits; F2 = Self-Esteem; F5 = Childhood Abuse; F7 = Health Concern; F8 = Lack of Social Support; S1 = Major Depression; S6 = Agoraphobia; S7 = Schizophrenia Spectrum; S10 = Obsessive-Compulsive Disorder; a.1 = Calm vs. Anxiety; a.5 = Humor vs. Depressiveness; a.6 = Initiative vs. Shame; a.7 = Faith in Others vs. Distrust; b.2 = Love vs. Emotional Detachment; b.3 = Assertiveness vs. Lack of Assertiveness; c.1 = Altruism vs. Egocentrism; d.1 = Prudence vs. Temerity; d.2 = Commitment vs. Irresponsibility; d.4 = Emotional Fullness vs. Thrill Seeking; e.1 = Frustration Tolerance vs. Perfectionism. The table shows all the items eliminated for not meeting adequate levels in all calibration indices. Inverted items with respect to their scale are shown in *italics*. Indices with adequate level of calibration are shown in **bold**.

Among the eliminated items, indicators in which no longer any items were eliminated. For example, the Minimization indicator, and the Defensiveness and Positive Impression dimensions of the R3 scale. Other eliminated indicators include Infrequent Symptoms of Post-Traumatic Stress from the R4 scale, Social Support Not Specified from the F8 scale, Intense Sadness from the S1 scale, Conceptual Disorganization from the S7 scale, Social Deterioration Due to Obsessions from the S10 scale, Comfort During Social Relationships vs. Social Anxiety from the a.1 scale, and Ease of Concentration vs. Distractibility from the d.2 scale. After deleting the items, the selected items were randomly sorted and renumbered according to their quantity.

Study 2: development, evaluation and normalization of IDPI-11 scales

Methods

This study has three sections: (a) Development of the IDPI-11 scales with the parsimonious items selected in the previous study. (b) Evaluation of the psychometric properties of the developed measure. (c) Norms derivation for IDPI-11 scales. Eleven specific objectives of this study are included in each of these sections. The first section includes the following objectives: (1) to designate items for the floating scales and (2) to formulate differentiated item scores for scales according to their prototypicality. The second section includes these objectives: (3) to analyze the content validity of scales, (4) to perform an item-level analysis of the measure with CTT and IRT, (5) to analyze the structural validity of the scales of the Trait subgroup, (6) to analyze the convergent and discriminant validity of the scales within their own groups, (7) to analyze the validity of differentiation by groups of the scales, (8) to analyze the criterion validity of Personality group scales, and (9) to analyze the internal reliability of the IDPI-11 scales. The third section includes one objective: (10) to develop norms for the raw scores of the IDPI-11 scales.

Participants

Through a random sampling stratified by type of crime, 1130 inmates from the HP were initially included, assuming a loss rate of approximately 10% as a result of the eligibility criteria. The inclusion criterion was the approval of the consent form for the evaluation by the convict, and the exclusion criteria were cases with 10 or more blank or null answers and/or some disability that limited the willingness to give consent at the time of the evaluation. As a result, a representative sample of 1095 inmates remained eligible for the study: an adequate sample size since it has a CI = 95% and $e = 2%$ (see Bayne, 2018; Dillman et al., 2014). The average age of the inmates was 38.4 years, with an age range between 26 and 49 years (79.1%). In addition, it was shown that most inmates had a cohabiting partner (22.7%), did not finish secondary school (39%), came from the Peruvian highlands (61.8%), were located in pavilion P2 (32.8%), were sentenced to a medium security facility (39.1%), were detained for crimes against freedom (40%), have spent from 4 to 12 years in the HP (23.7%), and were serving a prison sentence from 4 to 12 years (48.9%). See Table 1 for more details on this sample.

Measures

Integrative Dimensional Personality Inventory-II (IDPI-11). This measure was developed for the present study with the calibrated items. The IDPI-11 evaluates the personality of the examinees in a dimensional (health-disease continuum) and integrative manner (integrating related constructs for a differential diagnosis and a complete evaluation) in the context of mental health. It uses a dichotomous format for its answers, which speeds up its application. The response options are «F» or «T» for its 190 items distributed in 49 scales within nine subcategories and five main groups. The instrument includes 4 Response Style scales (Invalidity, Inconsistency, Dissimulation, and Simulation), 8 Mental Health Factors scales, 10 Psychopathology scales, 7 Personality scales, and 20 Trait Facet scales.

In addition, it can be applied individually or in groups in approximately 36 minutes.

Procedure

The data collection was performed from April to July 2021. Like study 1, the test was performed in two steps. It followed the same procedures as the previous study because it was granted permission from the pertinent authorities of the prison and the Department of Ethics of the Continental University. Furthermore, the techniques used for data collection were: (a) a survey that consisted of the application of the instrument and (b) a review of secondary data from a database with information on the type of crime, prison security level imposed, and other data (age, National Identity Card number, pavilion, time served in prison, and length of sentence). On this occasion, some sociodemographic characteristics such as age, type of crime, and prison security level were used to perform inferential statistics and assess the criterion validity of the Personality group scales. In addition, Microsoft Excel v. 2019, IBM SPSS Statistics 28.0.1.0, IBM SPSS Amos 28.0.0, and RStudio v. 2021.09.1+372 were used for data processing and analysis. Likewise, missing data were imputed with the abovementioned second software using the multiple imputation method.

Data analysis

For the descriptive statistics of the sociodemographic characteristics of this sample (step 5), the mean age and mode of the rest of these characteristics were calculated using IBM SPSS Statistics.

Development of scales. To designate items for the floating scales (*objective I*)—taking into account the considerations described in Table D.2 of Appendix D—the *a priori designation* method was used. It consists in selecting all the items of the IDPI-11 to compose Invalidity (R1 scale), thus the raw score of this scale is increased by «1» point each time the examinee answers blank or null. This method is similar

to those used by Ben-Porath and Tellegen (2020a) to compose the Can Not Say (CNS) scale of the MMPI-3—although this procedure did not require statistical analysis and therefore had no results—, it is important to report the answers. Likewise, to compose Inconsistency (R2 scale), the *content grouping with statistical refinement* method was used. It consisted in evaluating the semantics of pairs of IDPI-11 items with similar content and items with opposite content. The pairs of items with similar content were mainly drawn from the same scale (e.g., both Agoraphobia items or Post-Traumatic Stress items). On the other hand, the pairs of items with opposite content were extracted from those items formulated for trait facets that have items of positive meaning and items of maladaptive connotation. Then, the pairs of selected items were evaluated using tetrachoric correlations (r_{tc}) from RStudio «psych» package and cut-off points of $r_{tc} > .70$ to confirm its relevance in the R2 scale (Glen, 2016).

Likewise, the items of the scales of the Facets of Traits group were also designated to compose the scales of the Traits subgroup—Serenity vs. Negative Affectivity (scale A), Humanity vs. Detachment (scale B), Integrity vs. Dissociality (scale C), Moderation vs. Disinhibition (scale D), and Psychological Flexibility vs. Anankastia (scale E)—using the *sequential system of construct-oriented scale development*. This procedure implied that the adjustment of the exploratory factorial analysis (EFA) was verified using IBM SPSS Statistics through the varimax rotation of this group of items. Kaiser-Meyer-Olkin (KMO) and Bartlett's Test of Sphericity (BTS) were used with a cut-off point $> .75$ of KMO and a chi-square p -value [p -value (χ^2)] $< .05$ of BTS as pre-requisites to evaluate the EFA (Ferrando et al., 2022; Indrayan & Holt, 2017). After verifying the adjustment of the EFA, the lambdas (λ) of the items' factor loads were analyzed, and only those with $\lambda > .30$ were considered adequate (Padilla, 2019). Moreover, taking into account the guidelines of Padilla (2019), the subsequent EFA required constant elimination from the analysis, items with low

communalities ($h^2 < .30$)⁵ and those items mainly loaded on the last factors. With this, it was possible to obtain items loaded on the five factors required for the scales of the Traits subgroup.

Next, the items were designated for the composition of the Functioning (PF) and Borderline Pattern (PL) through the *content grouping with statistical refinement method* to identify possible items of the Facets of Traits group scales to compose such scales. To have a good theoretical basis, the ICD-11 guidelines on the severity of PD were revised, which were mainly based on intra and interpersonal dysfunction, similar to criterion A of the DSM-5 AMPD (Section III). Existing measures of this construct were also reviewed to identify items corresponding to these main characteristics and to include them in the PF scale. The procedure for initially composing items on the PL scale was similar. Once the items were identified, one-factor CFA with the maximum likelihood (ML) estimator from IBM SPSS Amos was used to compose each of these scales. The cut-off points used to assess the goodness of fit of the model were: relative chi-square (χ^2/df) < 5, root mean squared error of approximation (RMSEA) $\geq .05$, Tucker Lewis index (TLI) $\geq .90$, standardized root mean square residual (SRMR) $\leq .08$, and comparative fit index (CFI) $\geq .95$ (Boateng et al., 2018; Collier, 2020). Once the fit of the model was confirmed, the factor loads were assessed, and a cut-off point of $\lambda \geq .60$ was used to retain the items in each scale (Awang, 2014).

To formulate differentiated item scores for scales according to their prototypicality (*objective 2*), the *content grouping with statistical refinement method*⁶ was used. It includes only items from solid scales; therefore, Spearman's rho was calculated using IBM SPSS Statistics to analyze the relationship

between the items of the solid scales and the other solid scales and thus assign non-prototype items to other scales. It was verified that these items provide a *theoretical* and *diagnostic* contribution to the scale in order to finally include it (Grossman, 2019; Millon et al., 2015). The cut-off point used to consider the inclusion of an item in the target scale was a moderate to high degree of correlation, $|\rho| \geq .40$ (Akoglu, 2018).

It should be noted that, although the prototype items of the solid scales Dissimulation (R3) and Simulation (R4) were designated as non-prototype items in other scales of the IDPI-11, prototype items of other scales were not designated as non-prototypes in the R3 or R4 scale. This is due to the possibility that, in an acute episode of psychotic symptomatology (e.g., with a high score on the Schizophrenia Spectrum [S7 scale]), the prisoner may endorse items in the R4 scale, which does not imply a «faking bad» performance (Sellbom et al., 2022). This situation can also occur in the way that an examinee with egocentric characteristics (e.g., with a high score on Altruism vs. Egocentrism [scale c.1]) supports items on the R3 scale, which does not imply a «faking good» performance (Ben-Porath et al., 2022). However, the fact of dissimulating (elevation in the R3 scale) and feigning (elevation in the R4 scale) is often accompanied by similar responses –supported by items in the IDPI-11 scales that measure egocentrism (scale c.1) and psychotic symptomatology (S7 scale)– due to the tendency of the inmates' responses (Ben-Porath et al., 2022; Sellbom et al., 2022).

Thus, the designation of the scores for the scales was relatively simple. A score of «2» was established if the item supported by the examinee was adequate

⁵ We chose different cut-off points for factor loadings at different objectives, either for EFA or CFA, due to development or evaluation purposes and the theoretical importance of the desired latent factors.

⁶ Although this procedure originally corresponded to Wiggins (1973, as cited in Williams et al., 2019), Millon et al. (2015) modified it to give differentiated scores to the items according to their prototypicality within a scale, thus reducing the overlap between these scales and, at the same time, increasing the internal consistency of the scales without the need to add more items to the measure.

to its scale (prototype item), and a score of «1» if the item was not appropriate (non-prototype item). With this, direct non-prototype items correspond to objective scales with the same direction as their original scale –either towards health or disease⁷– and inverse non-prototype items correspond to target scales with a direction opposite to that of their original scale.

Evaluation. To analyze the content validity of scales (*objective 3*), CVI-S was used –which is the average of the CVI-I of the items that compose it, also called Ave-CVI-I– from the CVI-I of the prototype and non-prototype items of each scale, based on the panel of experts’ assessment of the IDPI-11 items conducted in step 2 with a cut-off point of CVI-S > .80 to consider an adequate level of content validity (Almanasreh et al., 2019).

Then, to analyze an item-level analysis of the measure with CTT and IRT (*objective 4*), successive one-factor CFAs were performed to assess the degree of relationship between the items (prototypes and non-prototypes) and the latent factor of each scale using IBM SPSS Amos and the standardized RMR plugin. To do this, first, the fit of the one-factor models was evaluated taking as cut-off point $\chi^2/df < 5$, RMSEA $\geq .05$, TLI $\geq .90$, SRMR $\leq .08$, and CFI $\geq .95$ (Boateng et al., 2018; Collier, 2020). Then, the factor loads were analyzed taking the cut-off point as appropriate: $\lambda \geq .60$ (Awang, 2014).

Likewise, to perform the item-level analysis from IRT –having fulfilled the assumption of unidimensionality of the IRT, evidenced in the adjustment parameters of the CFA– we proceeded to evaluate the assumptions of monotonicity, local independence and invariance of the items with the monotonicity violation criterion (*crit*), item discrimination parameter (*a*), and differential item

functioning (DIF) statistics from RStudio’s «mokken», «psych», and «mirt» packages, respectively. This considered the following as cut-off points: *crit* < .80, to confirm the monotonicity; $a \leq 4$, to confirm local independence; chi square (χ^2) > 3.3; and *p*-value (χ^2) $\geq .05$ of DIF to confirm the invariance of the items according to the age groups and type of crime (Nguyen et al., 2014). Also, using RStudio’s «mirt» package, and as suggested by Nguyen et al. (2014), the adjustment of the appropriate IRT model to the items of each of our scales was evaluated, comparing the models of two (2PL) and three parameters (3PL), as is usually done in the literature for dichotomous items. To do this, the -2 log-likelihood ratio (-2LL) statistic, also called the Likelihood ratio test, was used with the RStudio «mirt» package. In addition, as suggested by these authors, the fit of the selected IRT model was evaluated at the item level (Nguyen et al., 2014), with the Orlando and Thissen’s chi square ($S-\chi^2$) statistic, using the same package. For the last two previous analyses, the cut-off point *p*-value (χ^2) $\geq .05$ of -2LL was used to confirm that the most parsimonious model (2PL) fits better than the most complex and flexible model (3PL), and a *p*-value ($S-\chi^2$) $\geq .05$ was used to confirm the good fit of each of the items for each of our scales (Nguyen et al., 2014).

Finally, and after evaluating these IRT prerequisites, we proceeded to analyze the statistics: item difficulty parameter (*b*) and item discrimination parameter (*a*) with the RStudio «psych» package, using the cut-off points of $b = -2$ to 2 for items not difficult to support (including the items of scales F1, F2, F3, and F4), $b \geq 0$ for construct items theoretically difficult to support (including the items of the rest of the scales), and $a \geq 0$ for all the IDPI-11 items to value a good level of discrimination (Nguyen et al., 2014; Yang & Kao, 2014).

⁷ Dissimulation (scale R3) and the scales of the Protective Factors subgroup (Healthy Habits [scale F1], Self-Esteem [scale F2], Meaningful Activities [scale F3], and Openness to Treatment [scale F4]) have a *direction towards health*. The scales of the Validity Indices subgroup (Invalidity [scale R1] and Inconsistency [scale R2]) *do not have a specific direction*. And the rest of the scales follow a *direction towards psychopathology*.

To analyze the structural validity of scales of the Trait subgroup (*objective 5*), a five-factor CFA (with the ML estimator) was used using IBM SPSS Amos and its standardized RMR plugin, with the cut-off points recommended by Boateng et al. (2018) and Collier (2020): $\chi^2/df < 5$, RMSEA $\geq .05$, TLI $\geq .90$, SRMR $\leq .08$, and CFI $\geq .95$ to consider an adequate fit of each of the models. Next, the factor loads were analyzed, considering the cut-off point $\lambda \geq .60$ as appropriate (Awang, 2014).

For *objective 6*, to analyze the convergent and discriminant validity of the scales within their own groups, consecutive CFAs of the IDPI-11 scales were analyzed by group using IBM SPSS Amos and its «master validity» and «standardized RMR» plugins. For this, the cut-off points mentioned by Boateng et al. (2018) and Collier (2020) were used for the indices of good fit. Then, the average variance extracted (AVE) and Pearson's r correlations were analyzed, and the appropriate values of convergent validity for a scale resulted in an AVE $\geq .50$ (Bello, 2016). The optimal values of discriminant validity implied that the square root of a scale AVE is greater than the correlations between said scale and the other scales of the subgroup (Ab Hamid et al., 2017; Bello, 2016).

For *objective 7*, to analyze the validity of differentiation by groups of the scales, we first evaluated whether the raw scores of the IDPI-11 scales, according to age groups (group 1 = 18 to 34 years old, group 2 = 35 years old and older) and type of crime (see Table 1), had a normal distribution using IBM SPSS Statistics. To assess the indices of Kolmogorov-Smirnov (KS) for groups of 30 or more cases and the indices of Shapiro-Wilk (SW) for groups with fewer than 30 observations, we used a p -value $> .05$ for both indices as confirmation of the non-normality of these distributions (Frey, 2022). Later, with the RStudio «onewaytests» package, the homoscedasticity of the variances was evaluated using the Fligner-Killeen index (χ_o^2) with the cut-off point of p -value (χ_o^2) $> .05$ to confirm said homogeneity (Frey, 2022). After confirming non-normal distributions

and similar variances in the scales, the Mann-Whitney (U) and Kruskal-Wallis (H) indices were used with IBM SPSS Statistics to assess whether there were significant differences between these groups, with the p -value $\leq .05$ of both indexes showing a clear difference (Frey, 2022). In addition, two algorithms were used to calculate the effect sizes of said differences through Microsoft Excel, with the Glass rank biserial correlation coefficient (r_g) for U –with King et al. (2018) algorithm– and with epsilon squared (ϵ^2) for H –with Tomczak and Tomczak (2014) algorithm–. The cut-off points of $|r_g| \geq .40$ and $\epsilon^2 \geq .04$ were considered moderate effects or more (King et al., 2018; Stikker, 2018).

Then, to analyze the criterion validity of Personality group scales (*objective 8*) using IBM SPSS Statistics, Nagelkerke's pseudo- R^2 coefficient of determination (R^2_N) and odds ratio (OR) statistics were first analyzed as fit indicators and predictors in logarithmic regression models to explain the presence of crimes against public health, heritage, liberty, and life, body, and health. For this, the parameters of $R^2_N \geq .50$ (Frey, 2022) and p -value (OR) $< .05$ (Sperandei, 2014) were evaluated and thus considered significant. Then, the R squared coefficient of determination (R^2) and beta (β) statistics were analyzed for predictive analyses of the linear regressions of the Personality group scales to explain the increase in the prison security level. The parameters of $R^2 \geq .50$ (Frey, 2022) and p -value (β) $< .05$ (Ali & Younas, 2021) were taken into account to consider a good fit of the model and a significant predictive capacity of the scale, respectively. And, finally, to analyze the reliability of IDPI-11 scales (*objective 9*), McDonald's omega (ω) was computed for items of each scale using the new features of IBM SPSS Statistics, and the parameter to assess an adequate degree of internal consistency was $\omega \geq .70$ (Olivas-Ugarte & Cipriani-Delgado, 2022).

Normalization. For *objective 10*, to develop norms for the raw scores of the IDPI-11 scales (traditional and uniform), the mean (M), standard deviation (SD) skewness (Sk), and position

(percentile score [Pc]) were calculated using IBM SPSS Statistics. The methodology of Ben-Porath and Tellegen (2020a) was followed to transform the raw scores of the IDPI-11 scales into uniform T scores (UT scores), similar to the scores established in the MMPI-3 scales with right-skewed PD distributions ($Sk \geq .1$). Thus, the final scores were generated for the raw scores of each IDPI-11 scale. Traditional T scores (TT scores) were designated for the Healthy Habits (F1), Self-Esteem (F2), Significant Activities (F3), and Openness to Treatment (F4) scales, because their distributions do not have a bias to the right. Additionally, UT scores were elaborated for the scales in case they had this bias. The steps for the construction of UT scores of the aforementioned scales involve a «matching» between the raw scores and integer composite traditional T-scores (ICTT scores) as described below:

- (a) First, the means and standard deviations of the raw scores on each scale were calculated.
- (b) TT scores of each raw score obtained for each scale were calculated.
- (c) Next, percentiles .5, 1, 2, ..., 99, 99.5 (101 percentiles) of the TT score of each scale were found.
- (d) Subsequently, the average of the TT scores of all scales for each percentile was calculated, the results of which were incorporated into a new column called the composite traditional T score (CTT score).
- (e) After obtaining the CTT scores for each percentile, all TT scores for each scale were replaced with their respective raw scores.
- (f) Next, the raw score of each ICTT score –ICTT scores were used as a template– were calculated through linear interpolation.
- (g) Then, to determine the UT scores of the raw scores obtained, UT scores were calculated from polynomial interpolations of degree 3: raw scores of the ICTT scores less than 60 and UT scores of the raw scores of the normative sample,

corresponding to ICTT scores equal to or greater than 60, were determined by linear interpolation, which served to verify that the UT scores were the same as the ICTT scores as the test intended, and indeed they were.

- (h) Finally, to establish raw scores that were outside the range of the normative sample, corresponding to the UT scores, linear extrapolations were performed by establishing a lower limit (UT score = 20) and an upper limit (UT score = 100). Thus, any extreme raw score outside the linear extrapolations would take the UT score corresponding to the upper or lower limit, as the case may be. For more details, see Appendix F.

Results

Development of scales. In item designation for the floating scales (*objective 1*), using content grouping with statistical refinement method, 55 pairs of items with high tetrachoric correlations ($r_{tc} > .77$) were obtained for Inconsistency (R2 scale). For example, for pairs of items with opposite content (49 pairs in total) corresponding to the scales of the Trait Facets group, most of the correlations were greater than .80; whereas for pairs of items with similar content (six pairs in total), the tetrachoric correlation was in the range of .78 to .86. Consequently, these 55 pairs in total were included with strong correlation coefficients to compose the R2 scale (Akoglu, 2018).

For the composition of scales of the Facets of Traits group, the sequential system of construct-oriented scale development was applied through a series of EFAs up to a five-factor solution with previous verification of its adequacy: $KMO = .81$, $\chi^2 = 2314.9$, p -value = .00 (Ferrando et al., 2022; Indrayan & Holt, 2017). Thus, following Padilla's criteria (2019), only three items from the a.1 scale, two items from the a.3 scale, two items from the a.4 scale, and two items from the a.6 scale obtained adequate factor loadings ($\lambda = .35$ to $.73$) to compose scale A. In addition, six items belonging to scale b.1 and two items belonging to scale b.2 obtained relevant

levels of factor loadings ($\lambda = .40$ to $.71$) to compose scale B. Likewise, two items from the c.1 scale and five items from the c.2 scale obtained appropriate factor loadings ($\lambda = .38$ to $.63$) to compose scale C. Similarly, three items from the d.1 scale, two items from the d.2 scale, and five items from the d.4 scale obtained adequate indices of factor loads ($\lambda = .32$ to $.71$) to compose scale D. Finally, three items from the d.1 scale, two items from the e.2 scale, and two items from the e.3 scale obtained relevant indices of factor loads ($\lambda = .37$ to $.69$) to compose scale E.

Then, for the composition of Functioning (PF) and Borderline Pattern (PL), the *content grouping with statistical refinement method* was used through one-factor confirmatory analysis (CFA) for each scale – with previous verification of its adequate goodness-of-fit indices: $\chi^2/df = 4.3$, RMSEA = $.06$, TLI = $.95$, SRMR = $.07$, and CFI = $.97$ for scale PF and $\chi^2/df = 3.8$, RMSEA = $.08$, TLI = $.92$, SRMR = $.06$, and CFI = $.99$ for scale PL (Boateng et al., 2018; Collier, 2020). Once the fit of the model was confirmed, appropriate levels of factor loads were evidenced ($\lambda = .71$ to $.86$) to designate items from the scales of the Facets of Trait group –from a.2 (one item), a.3 (4), a.5 (2), b.1 (4), b.2 (2), c.2 (2), e.1 (2), and e.3 (4)– to compose the PF scale (Awang, 2014). Similarly, considering the cut-off proposed by Awang (2014), appropriate levels of factor loadings were evidenced ($\lambda = .76$ to $.92$) to designate items from other scales –from a.2 (one item), a.2 (1), a.3 (3), a.4 (3), a.5 (1), a.7 (2), c.2 (1), d.1 (1), d.4 (3), and S8 (1)– to compose the PL scale.

In the formulation of differentiated item scores for solid scales according to their prototypicality (*objective 2*), the content grouping with statistical refinement method was applied. This implied that the items of the other solid scales were related to a moderate-to-high level with the total score of the objective scale. After the analyses, correlations were obtained in the range of $|rho| = .40$ to $.66$. These items were designated as direct non-prototype items for having positive correlations with the total score of the target scale (scored as F = '0' and T = '1') and as

inverse non-prototype items for having negative correlations (scored as F = '1' and T = '0'). See Table E.1 of Appendix E for details on these correlations. Consequently, solid scales with an average of 3.2 (from 1 to 7) direct non-prototype items and an average of 1.7 (from 1 to 5) inverse non-prototype items could be obtained, excluding from the account the solid scales that did not have any direct or inverse non-prototype items. The final composition by items of all IDPI-11 scales is shown in Appendix H.

Evaluation. In the analysis of the content validity of scales (*objective 3*), CVI-S was used as the CVI-I average of all items that integrate each scale, based on the assessment of the IDPI-11 items in step 2 conducted by the panel of experts (item revision by a panel of five experts). The results showed adequate levels of content validity for all scales, since CVI-S was found in the range of $.88$ to 1 (Almanasreh et al., 2019).

In the item-level analysis from CTT (*objective 4*) through one-factor CFAs for each scale, in general, with good fit indices, χ^2/df in the range of 2 to 4, CFI $> .96$, TLI $> .95$, RMSEA $< .07$, and SRMR $> .06$ (Boateng et al., 2018; Collier, 2020), IDPI-11 items obtained good levels of factor loads with their original scales ($\lambda_1 > .59$) (Awang, 2014). However, their mean levels of factor loadings with their target scales varied ($\lambda_2 = .30$ to $.78$).

In the item-level analysis from IRT, the unidimensionality of the IDPI-11 scales was confirmed by considering the adequate goodness-of-fit indices of the one-factor confirmatory analyses shown in the previous analysis (Nguyen et al., 2014). In addition, considering the criteria of Nguyen et al. (2014), the monotonicity, local independence and invariance of the items were adequate for each of the IDPI-11 scales with $crit = .23$ to $.65$, $a = .1$ to 3.6 , and $\chi^2 = 3.6$ to 5.8 (p -value = $.06$ to $.08$), respectively. Similarly, in the IRT model fit test, the most parsimonious model (2PL) was adequate with $\chi^2 = 4.6$ to 8.2 (p -value = $.05$ to $.07$) in the -2LL

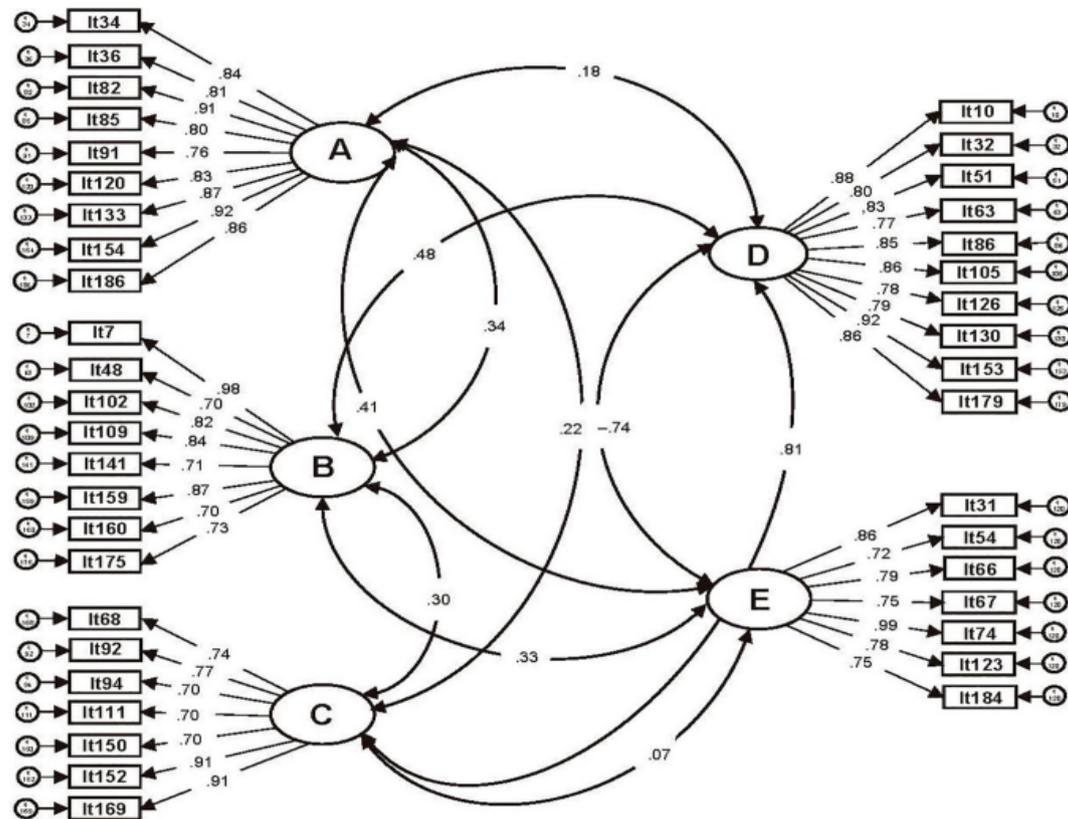
test; in addition, the adjustment at the item level showed $S-\chi^2 = 6.3$ to 8.7 (p -value = .06 to .09).

Once the assumptions of the IRT were verified and the best fit of the 2PL model was evidenced, the results of the IRT parameters showed that the IDPI-11 items differed well in the high and low levels of the construct according to the probability of supporting the item in the direction of its original scale ($a > .1$). Likewise, the difficulty parameters showed, in general, adequate levels according to the construct they measured, although there were some exceptions: for example, item 19 ($b = .3$) and item 26 ($b = -.2$), which focus on depressiveness and distrust, and are frequently supported in the study population. These parameters showed that the items of the clinical construct were difficult to support for the examinees

($b > .5$) and that the non-clinical constructs obtained a varied range of difficulty ($b > -.6$). These results are generally consistent with the levels of probability of support and the degree of change in said support in relation to the measured trait, according to some authors (Nguyen et al., 2014; Yang & Kao, 2014). See Table E.2 in Appendix E for more details.

Then, in the structural analysis of the Trait subgroup scales (*objective 5*), through the a five-factor CFA, adequate goodness-of-fit indices were obtained for the Five Factors model: $\chi^2/df = 4$, CFI = .99, TLI = .91, RMSEA = .65, and SRMR = .52 (Boateng et al., 2018; Collier, 2020), as well as appropriate levels of factor loads ($\lambda \geq .70$) (Awang, 2014). Figure 2 presents the CFA standardized factor loads for the items of these scales.

Figure 2
Standardized coefficients of the confirmatory factor analysis of the traits subgroup scales



Note: A = Serenity vs. Negative Affectivity; B = Humanity vs. Detachment; C = Integrity vs. Dissociality; D = Moderation vs. Disinhibition; E = Psychological Flexibility vs. Anankastia.

In the analysis of convergent and discriminant validity of the scales within their own groups (*objective 6*), carried out through five CFAs, the adjustment indices were previously verified, finding adequate values: $\chi^2/df = 2$ to 3 , CFI $> .96$, TLI $> .93$, RMSEA $> .07$, and SRMR $< .04$ (Boateng et al., 2018; Collier, 2020). It was also found that the AVE of a CFA series of the scales of its corresponding group was greater than $.50$, evidencing convergent validity (Bello, 2016). It was also noted that \sqrt{AVE} was higher than Pearson's r correlations, with the other scales of its group confirming discriminant validity of these scales (Ab Hamid et al., 2017; Bello, 2016). Once non-normality (KS/SW $\geq .09$) and homoscedasticity ($\chi_o^2 \geq .26$) in most of the scales by groups of crime type and age range were confirmed, a series of comparison tests were analyzed in each IDPI-11 scale for these groups (*objective 7*). Most of the scales showed significant differences between crime groups ($H \geq 32.97$, p -value $\leq .001$), although

with low effect sizes ($\epsilon^2 < .39$). Similarly, most of the scales showed significant differences between age groups ($U < 131$ 331.50, p -value $< .05$), as well as low effect sizes ($rg < .40$) (Frey, 2022; King et al., 2018; Stikker, 2018).

In the analysis of criterion validity of Personality group scales (*objective 8*), it was found that the IDPI-11 personality group scales were capable of significantly explaining –according to the thresholds of some authors (Frey, 2022; Sperandei, 2014)– the probability of belonging to the group of inmates with crimes against public health and property, in addition to crimes against liberty, and against life, body, and health (R^2_N between $.52$ and $.68$ with predictors OR between 1.02 and 1.98 , p -value $\geq .21$). Likewise, these scales can predict that the prison security level imposed will increase: $R^2 = .53$ to $.61$ with β predictors between 1.16 and 1.97 , p -value $\geq .008$ (Ali & Younas, 2021; Frey, 2022), as shown in Table 3.

Table 3
Predictive logistic and linear models of the Personality group scales

Scale	Crime A. Public Health	Crime A. Property	Crime A. Liberty	Crime A. Life, Body, and Health	Prison Security Level
	OR (p)	β (p)			
PF	<i>1.15 (< .001)</i> $R^2_N = .56$	<i>1.26 (< .001)</i> $R^2_N = .68$	<i>1.02 (.007)</i> $R^2_N = .52$	<i>1.38 (< .001)</i> $R^2_N = .57$	<i>1.46 (< .001)</i> $R^2 = .61$
A	.04 (.543)	.04 (.543)	<i>1.36 (< .001)</i>	<i>1.41 (< .001)</i>	.04 (.543)
B	.08 (.346)	.02 (.654)	<i>1.64 (.008)</i>	.04 (.543)	.04 (.564)
C	<i>1.24 (< .001)</i>	<i>1.09 (.005)</i>	.21 (.084)	.23 (.075)	<i>1.97 (< .001)</i>
D	<i>1.36 (< .001)</i>	<i>1.87 (< .001)</i>	<i>1.06 (.021)</i>	<i>1.95 (< .001)</i>	<i>1.84 (< .001)</i>
E	<i>-1.47 (< .001)</i> $R^2_N = .45$.04 (.287) $R^2_N = .52$	<i>1.10 (< .001)</i> $R^2_N = .54$	<i>1.98 (.004)</i> $R^2_N = .62$	<i>1.16 (< .001)</i> $R^2 = .53$
PL	.41 (.931) $R^2_N = .41$.58 (.126) $R^2_N = .23$	<i>1.54 (< .001)</i> $R^2_N = .61$	<i>1.66 (< .001)</i> $R^2_N = .59$	<i>1.25 (.008)</i> $R^2 = .53$

Note: $n = 1095$; OR = odds ratio; β = beta (standardized regression coefficient); p = significance value; R^2 = coefficient of determination of the linear regression; R^2_N = Nagelkerke's pseudo- R^2 coefficient of determination of the logistic regression; PF = Functioning; A = Serenity vs. Negative Affectivity; B = Humanity vs. Detachment; C = Integrity vs. Dissociality; D = Moderation vs. Disinhibition; E = Psychological Flexibility vs. Anankastia; PL = Borderline Pattern. Only four types of crimes were used for the regression models because there were more than 50 cases. The coefficients of determination of the models that explain a variance of the dependent variable greater than $.05$ are shown in **bold**. The significant regression coefficients (OR or β) of these models are in *italics*.

Table 4*Mean, standard deviation, range, and McDonald's omega of the IDPI-11 scales*

Scale	<i>M</i>	<i>SD</i>	<i>Range</i>	McDonald's ω
Response Styles (Group R)				
<i>Validity Indices</i>				
R1: Invalidity ^a	—	—	—	—
R2: Inconsistency	12.9	3.3	6-24	.12
<i>Response Distortion Indices</i>				
R3: Dissimulation	5.5	4.1	0-10	.88
R4: Simulation	4.4	3.9	0-10	.90
Mental Health Factors (Group F)				
<i>Protective Factors</i>				
F1: Healthy Habits	5.1	2.6	0-8	.81
F2: Self-Esteem	6.8	2.9	1-11	.80
F3: Meaningful Activities	4.8	3.6	0-10	.79
F4: Openness to Treatment	5.4	3.6	0-10	.79
<i>Risk Factors</i>				
F5: Childhood Abuse	3.8	3	0-11	.82
F6: Suicidal Tendency	5.8	3.6	0-11	.76
F7: Health Concern	3	2.5	0-9	.78
F8: Lack of Social Support	4.9	2.7	0-10	.94
Psychopathology (Group S)				
<i>Frequent Syndromes</i>				
S1: Major Depression	3.6	3.6	0-10	.95
S2: Post-Traumatic Stress	4.5	4.1	0-11	.83
S3: Alcohol Consumption	4.3	4	0-13	.82
S4: Generalized Anxiety	4.5	3.4	0-12	.91
S5: Social Phobia	3.8	3.8	0-12	.84
<i>Uncommon Syndromes</i>				
S6: Agoraphobia	3.1	3.2	0-11	.75
S7: Schizophrenia Spectrum	4.7	3.2	0-14	.95
S8: Dysthymia	4.5	3.6	0-13	.73
S9: Panic	3	3.3	0-11	.79
S10: Obsessive-Compulsive Disorder	3.2	2.8	0-11	.78
Personality (Group P)				
<i>Severity</i>				
PF: Functioning	7.3	3.8	1-16	.88
<i>Traits</i>				
A: Serenity vs. Negative Affectivity	2.7	2	0-7	.83
B: Humanity vs. Detachment	2.4	2.2	0-8	.77
C: Integrity vs. Dissociality	2.6	2.3	0-7	.79
D: Moderation vs. Disinhibition	3.4	2.4	0-10	.92
E: Psychological Flexibility vs. Anankastia	3.1	2	0-7	.84
<i>Additional Qualifier</i>				
PL: Borderline Pattern	5.5	2.9	0-12	.95

Scale	<i>M</i>	<i>SD</i>	<i>Range</i>	McDonald's ω
Trait Facets (Group P.2)				
<i>Serenity vs. Negative Affectivity</i>				
a.1: Calm vs. Anxiety	3.5	3.2	0-11	.75
a.2: Fortitude vs. Vulnerability	5.1	4.5	0-16	.93
a.3: Emotional Stability vs. Emotional Lability	4	4.6	0-14	.74
a.4: Patience vs. Anger	5.2	4.6	0-15	.80
a.5: Humor vs. Depressiveness	4.9	4	0-14	.83
a.6: Initiative vs. Shame	4.9	3.3	1-12	.77
a.7: Faith in Others vs. Distrust	3.2	4.1	0-10	.79
<i>Humanity vs. Detachment</i>				
b.1: Sociability vs. Social Detachment	3.8	4.1	0-12	.91
b.2: Love vs. Emotional Detachment	2.8	3.4	0-10	.84
b.3: Assertiveness vs. Lack of Assertiveness	3.1	3.4	0-10	.78
<i>Integrity vs. Dissociality</i>				
c.1: Altruism vs. Egocentrism	4.9	4.4	0-14	.74
c.2: Sensitivity vs. Lack of Empathy	4.6	3.7	0-11	.92
c.3: Kindness vs. Aggressiveness	4.6	3.6	0-12	.96
<i>Moderation vs. Disinhibition</i>				
d.1: Prudence vs. Temerity	4	3.1	0-11	.81
d.2: Commitment vs. Irresponsibility	3.5	2.7	0-10	.79
d.3: Planning vs. Disarray	5	4.6	0-13	.87
d.4: Emotional Plenitude vs. Emotion Pursuit	5.3	4	0-13	.81
<i>Psychological Flexibility vs. Anankastia</i>				
e.1: Tolerance to Frustration vs. Perfectionism	7	3.4	1-13	.78
e.2: Availability for Leisure vs. Work Addiction	6.1	3.9	0-14	.84
e.3: Docility vs. Inflexibility	4.6	4.1	0-12	.91

Note: *M* = mean; *SD* = standard deviation. Group names are shown in **bold** and subgroups names, in *italics*.

^a Invalidity (scale R1) was excluded from the analyses.

Finally, McDonald's omega (ω) was used to analyze the internal consistency of the IDPI-11 scales (*objective 9*), and the results showed adequate internal consistency indices for the scales of the Response Styles group: Dissimulation ($\omega = .88$) and Simulation ($\omega = .90$). However, the only scale that obtained a low level in this group was Inconsistency ($\omega = .12$) (Olivas-Ugarte & Cipriani-Delgado, 2022). Similarly, considering the criteria of Olivas-Ugarte and Cipriani-Delgado (2022), the internal consistency of the other scales of the Mental Health Factors ($\omega = .76$ to $.94$), Psychopathology ($\omega = .73$ to $.95$), Personality ($\omega = .77$ to $.95$), and Trait Facets ($\omega = .74$ to $.96$) obtained appropriate levels of internal consistency. See Table 4.

Normalization. After the process detailed in Appendix F, the development of norms for the raw scores of the IDPI-11 scales (*objective 10*), the UT scores generated for the scales of the Risk Factors subgroup, as well as for the scales of the Psychopathology, Personality and Trait Facets groups, managed to obtain common percentile scores that facilitate the differential diagnosis (Ben-Porath & Tellegen, 2020a). So, uniform PTs of 30, 35, 40, 45, 50, 55, 60, 65, 70, and 75 were equivalent to percentile score (Pc) of < 1, 12, 14, 41, 61, 73, 80, 87, 98, and > 99, respectively.

Discussion

The eleventh revision of the ICD has recently entered into force and is ready to be implemented by member states, continuing the tradition of having a global scope and being applicable in diverse cultures and settings, with a focus on primary care in underdeveloped countries unlike the DSM (Halliwell, 2021). One of the sections of this diagnostic system addresses personality disorders and related traits, a model that promises to correct the errors of its predecessor, which was criticized by the scientific community for being less restrictive and allowing greater specificity. Some measures have been developed based on this model to obtain adequate psychometric levels with a few exceptions. The limitations of these results mainly emerge from the methodology, the samples used, and the manner in which their manuals are specified. In fact, its application in prison settings and underdeveloped countries has not yet been rigorously evaluated. Therefore, in this research a comprehensive measure of personality based on the ICD-11 PD model was developed, considering an extensive theoretical, practical, and methodological foundation that favors the best evaluation of the construct in the Peruvian correctional system.

The item-level results supported the psychometric properties of each IDPI-11 scale. It was expected that the factor loadings of the items with their original scales would be higher than those of their target scales as found in the development of the MCMI-IV (Millon et al., 2015). Likewise, for the most part, both the IRT parameters of discrimination and those of difficulty reflected a good ability to differentiate and detail the levels of constructs measured through the probability of their support. In addition, it was found that in this study, as in Medina (2021), depressiveness and distrust, in addition to other characteristics and symptoms, are frequent in the criminal population. In this sense, it is key to make decisions to identify appropriate personalized and group treatments to reduce or control the exacerbation of these dysfunctional indicators in convicts.

The findings showed that the developed instrument had an adequate level of construct validity and an external criterion for the scales of the IDPI-11 Traits subgroup. These results are similar to those found in the construction of the Personality Inventory for the DSM-5 - Brief Form Plus (PID5-BF+), a multinational study (Bach et al., 2020) in which the structural validity demonstrated a CFI above .95 and an RMSEA below .06. Additionally, the instrument traits were correlated with personality constructs evaluated through interviews ($r > .17$). They were also similar to the optimal results of the research carried out by Oltmanns and Widiger (2018) in the USA for the construction of the Personality Inventory for the ICD-11 (PiCD): CFI = .77 to .83, TLI = .76 to .82, RMSEA = .10 to .11, and SRMR = .10 to .11. However, they are different from those found in the Spanish validation of the instrument because their data did not fit the CFA model (Gutiérrez et al., 2021).

The external criteria validity results found in this research are similar to those obtained in a study conducted in Italy (Ferretti et al., 2021), in which high irresponsibility (disinhibition) and high restricted affectivity (detachment) predicted child sexual abuse ($\beta = -.60$ and $\beta = -1.17$, respectively), and irresponsibility (disinhibition) predicted crimes against the person ($\beta = .017$). Similarly, it is partially similar to a study carried out in Huancayo, Peru, with child sexual abuse offenders (Medina, 2021), in which high prevalence of an obsessive-compulsive clinical pattern (anankastia) was found (41.5%). However, contrary to the results of this study, the dependent clinical pattern (negative affectivity) was more prevalent in this population, and was also shown as the second most prevalent (24.5%). This divergence may be due to the descriptive design used and certain shortcomings that prevented the instrument to be adapted to the Peruvian context. Similarly, it has been shown that psychopathy in its two variants (primary and secondary) is related to recidivism (Alonso et al., 2021), which is an important component in the designation of the prison security level in the Peruvian system.

In this study, the IDPI-11 personality scales were correlated with prison criteria, as they are of greater practical relevance (Day & Cook, 2019). Psychopathy and PD are closely linked, since several five-factor traits overlap with the psychopathic personality variants. Certainly, psychopathic traits are the main predictor of recidivism. Nevertheless, comorbidity of ASPD and BPD, in addition to substance use, predicts the commission of violent crimes (McMurrin & Howard, 2019). It was also found that offenders with high levels of substance use and attention-deficit/hyperactivity disorder (ADHD) are related to theft, and those who commit matricide are generally people with psychosis (Davison & Janca, 2012).

Therefore, it is important to understand that any link between personality disorders and violence must be viewed within the context of a link between general mental disorders and violence. The more psychopath characteristics people have, the better prognosis they will have to feel distressed (Stricker & Pietrowsky, 2022) when committing a crime, attacking people, or attacking themselves, as shown by the significant predictive capacity of the Functioning and Borderline Pattern scales in most prison criteria. In this situation, it is advisable to assess the protective attributes and positive tendencies (traits) of personality as proposed by the IDPI-11 scales. This proposal is based on the evidence of successful studies of salutogenic perspectives that support the importance and necessity of evaluating positive personality in a penitentiary center (see, e.g., Miner, 2021; Pasowicz & Piotrowski, 2021). This will make it possible to identify the protective factors and strengths of the inmates. This, through the «High» levels of the scales of the Protective Factors subgroup (T score ≥ 65) and the «Positive» levels of the Functioning scale, the Traits subgroup scales, and the Facets of Traits group scales (T score = 20 to 34). Such evaluation can help determine the goals, plan and supervise the treatment, to help them «give back» to others and provide them with opportunities to participate in daily community life (Levak et al., 2011).

The reliability findings of each of our scales are similar to those found in the development of the MCMI-IV (Cronbach's $\alpha \geq .63$; see Millon et al., 2015) and MMPI-3 (Cronbach's $\alpha \geq .69$ in the Reestructured Clinical [RC] Scales; see Ben-Porath & Tellegen, 2020b). Furthermore, the low reliability of Inconsistency (R2 scale) is similar to a study carried out in Ohio, USA, on the Combined Response Inconsistency (CRIN scale) of the MMPI-3 (Cronbach's $\alpha = .27$), due to the willingness of examinees to do the evaluation (Whitman & Ben-Porath, 2021).

Finally, standardized scores were designed for the IDPI-11 scales, depending on the bias of their distribution and/or the nature of their conception. Moreover, it is necessary to mention that the UT score of Functioning (PF scale) needs to be modified using the Intensity/Comorbidity (IC) adjustment in order to increase the specificity of this scale. This is because the comorbidity of several maladaptive personality traits/types, followed by the severity itself that can exist in a single maladaptive personality trait/type, usually causes deterioration of personality functioning (Tyrer et al., 2019). For more information on the IC adjustment see Appendix G.

How does IDPI-11 fill the gap?

The rationale of this research is derived from the potentialities previously expressed in the ICD-11 PD model for evidence-based psychological assessment (EBPA). Theoretically, focusing on the *actor self* allows a better understanding of the personality dynamics of the prison population, conceptualizing current behavior (clinical diagnosis), determining the reason for it (etiological diagnosis), and understanding the early interaction and the current result between nature and nurture in this type of population. Likewise, focusing on the *agent self*, it is possible to understand how psychosocial factors and the manifestations of mental disorders in the prison population influence each other or whether they influence the maintenance of the personality disorder, thus enriching the theoretical basis of differential diagnosis.

From a practical point of view, by emphasizing the *author self*, the developed instrument contributes to the detection, diagnosis, and prognosis of PD and other psychopathological conditions; by emphasizing the *agent self*, group treatment programs can be planned and evaluate the dynamic (current) risk and protection factors of the group; and by emphasizing the *author self*, individual treatment programs can be established, taking into account the dynamic and static (historical) risk and protection factors, and the way in which the inmate himself understands all of this to fit in society by finding the meaning of life and causing rehabilitation to last over time (see Appendix I for evidence-based treatment for the IDPI-11 scale profiles). These benefits are important since, to date, most evaluations for classification and treatment at the HP have been guided more by psychologists criteria than by a measuring instrument. With the profile obtained from the IDPI-11 measurement, which can be even shown by codes and not with the full names of the scales, and following an interpretation based on a measurement theory, it is possible to avoid mentioning the «label» (stigma) of the clinical conditions and instead reveal a comprehensive look towards the possible reason for the current state of the inmate (Grossman, 2019). This subtle way of translating and reporting the results promotes the development of empathy and a therapeutic alliance with the inmate, thereby obtaining better results in rehabilitation programs.

Limitations and future directions

This study is not exempt of limitations. One of the limitations was in the sample ($n = 60$) used for meeting the objective of study 1: item selection analysis from the IRT. At that time, there was no adequate statistical guide; therefore, the assumptions and minimum size necessary for the analysis were not evaluated. However, this was not evidenced and did not affect study 2, since adequate psychometric levels were evidenced. This may be due to the fact that the analysis with the IRT of the items in step 6 was accompanied by the analysis of the difficulty and

discrimination indices of the CTT. In addition, the guidelines for establishing sample size and test length assumptions to obtain precision in IRT modeling have been determined based on simulated studies and not with real test data. In this sense, there are –although few– accuracy studies of dichotomous items in small samples of up to 50 participants with test lengths of up to 10 items (Sahin & Anil, 2017).

Another limitation was that the selected sample, although large and randomized, was only collected in a single center, and was mostly made up of inmates from the Peruvian highlands. For this reason, more validation and adaptation studies are still needed in other correctional contexts and they must also be carried out with populations of women and adolescents to further generalize its adequate psychometric properties. A possible bias in the evaluation with the IDPI-11 may be caused by professionals underqualified for this type tasks. An advanced level of expertise; extensive knowledge; and practice in personality, psychopathology, epidemiology, clinical health psychology, and correctional psychology are needed for adequate interpretation of the profile of the examinee. If this is achieved, professionals will have a valuable tool for their evaluation and treatment practices.

An additional limitation was the failure to identify cut-off points through semi-structured interviews for the IDPI-11 scales due to the restrictions generated by the COVID-19 pandemic. Consequently, the thresholds for interpretation were designated intervals based on the statistical distribution of the scores on each scale, which was supported by a dimensional rather than a categorical one. Professional users must consider an error threshold for interpretation, whether this is greater or less than that obtained by the examinee on the Psychopathology and Personality scales (Grossman, 2017): the IDPI-11, like all psychological measurement instruments in general, does not perfectly predict the specific level where a person is located; and, to make better decisions, all the available information must be evaluated. Diagnostic accuracy studies for this measure are

needed to be applied in clinical settings; however, in the absence of similar instruments, IDPI-11 can be used with caution in populations with similar characteristics, as suggested by Millon et al. (2015).

Despite its limitations, the developed instrument, whose scales are generally valid and reliable, has many potential applications in the practice of the psychologist in the penitentiary context as follows:

- (a) For *mental health purposes*, because IDPI-11 has been developed primarily to provide a comprehensive assessment of the inmate's personality and associated mental health factors for mental health purposes, such as detection (as primary health care), diagnosis, prognosis, case conceptualization, treatment monitoring, and treatment follow-up (as secondary health care), as previously explained in the introductory section (Day & Cook, 2019). It can also be used for evaluating the predisposition of any mental illness in healthy inmates, for stratifying the risk of mental illness in inmates with risk factors (primary health care), for staging the severity of general mental disorders that inmates already have (as secondary health care), for monitoring inmates' chronic mental disorders, and for surveilling the remission or recurrence of the prisoners' chronic mental disorders (tertiary health care) (see Deeks & Bossuyt, 2021).
- (b) For *legal purposes* as in competency to stand a trial, criminal responsibility, dangerousness, pre-sentence, risk, and recidivism assessments.

For competency to stand a trial assessments, Simulation (scale R4) can help verify the intention of over-reporting psychiatric symptoms after ruling out a significant increase in the score of Schizophrenia Spectrum (scale S7) (see Ben-Porath et al., 2022; Butcher et al., 2015). For criminal responsibility assessments, it is also important to identify if there is a significant elevation in the R4 and S7 scales to better understand said responsibility, and it is also essential to verify if the items of infrequent symptoms

of memory problems that make up the R4 scale (items 23 and 128) have been supported (Butcher et al., 2015; Sellbom et al., 2022). For dangerousness assessments, the elevations of the Patience vs. Anger (a.4 scale), Prudence vs. Temerity (d.1 scale), and Commitment vs. Irresponsibility (d.2 scale) should be determined to carry out its pertinent transfer to civil or forensic units.

For pre-sentence assessments, the «present» and «prominent» levels of any of the scales of the Psychopathology group and the mild, moderate or severe level of Functioning (PF scale) can provide guidelines for decision-making on the main mental health needs, as well as the «positive» levels of the Prudence vs. Temerity (d.1 scale), Commitment vs. Irresponsibility (d.2 scale), and Planning vs. Disarray (d.3 scale), to inform the judge about the mitigating factors that may be considered for the sentence.

For risk assessments, it is appropriate to assess the «present» and «prominent» levels of the Patience vs. Anger (a.4 scale) and Kindness vs. Aggressiveness (c.3 scale). And, finally, for recidivism assessments, the «positive» levels of the sociability scale (b.1 scale) –common in sociopaths (see «The mask of sanity»; Patrick, 2018)– for hetero aggression, «medium» or «high» level of Suicidal Tendency (F6 scale) for self-aggression, and the «present» and «prominent» levels of the scales can be assessed (S3, a.4, b.2, c.1, c.2, c.3, d.1, d.2, d.3, and d.4). It can also be considered the type of crime committed, such as crimes against property and public order, since they are commonly associated with recidivism and therefore require a higher level of prison security level.

Another strength is that the IDPI-11 scales conform to current health standards. As a result, correctional psychologists, prison inmates, and researchers who use the IDPI-11 will be able to obtain an evaluation tool, pertinent care, and an appropriate methodological resource, respectively. In the global context, a comprehensive instrument has

not yet been built and, at the same time, adjusted to the ICD-11 PD model. In the Latin American context, to date, no comprehensive instrument has been developed to measure personality. In this sense, it is expected that the authorities of the Peruvian prison system can incorporate this instrument into their guidelines for the evaluation and treatment of convicts. Finally, it will be of great theoretical and practical benefit that future studies evaluate the IDPI-11 scales from network psychometry, an emerging and promising methodology for exploring the direct and integral dynamics of observed variables (nodes) without the direct influence of latent constructs (see e.g., Christensen et al., 2020; Isvoranu et al., 2022; See et al., 2020).

We conclude that the IDPI-11⁸ is a valid and reliable measure for determining HP inmates' personality from the model proposed in the ICD-11 nosology, since it provides a vast theoretical foundation from the integration of the models with greater scientific support and is a EBPA tool for the purposes of the primary, secondary, and tertiary mental health care, and for the guidance of legal decisions by the authorities of the Peruvian correctional setting.

Conflicts of interest

The authors have no conflicts of interest to declare.

Ethical responsibility

This study did not involve human or animal experimentation, and informed consent was obtained from the inmates and witnesses. To ensure the confidentiality of the data, consent was administered separately with the protocols, eliminating all forms of identification. Likewise, the study does not detail any

individual data on prisoners' responses or any data that allows their identification. This study was reviewed and approved by the Research Ethics Committee of the Faculty of Humanities (CEI-DD-HH) of Continental University.

Authors' contributions

LMHO, DNRC, and RMCA conceived the article; DNRC and PAAV collected the data; LMHO and PAAV performed the statistical analysis; LMHO reviewed the contents; and RMCA advised on the entire study and manuscript preparation process.

Acknowledgment

The authors thank Dr. Luis Centeno Ramírez for his recommendations regarding the statistical analysis of the data. We also thank Mike J. Crawford, M. D., Laura Weekers, Ph. D., and Joshua R. Oltmanns, Ph. D., who shared their instruments of personality psychopathology which served us as a guide for the construction of the IDPI-11.

References

- Ab Hamid, M. R., Sami, W., & Mohamad Sidek, M. H. (2017). Discriminant Validity Assessment: Use of Fornell & Larcker Criterion Versus HTMT Criterion. *Journal of Physics: Conference Series*, 890, 012163. <https://doi.org/10.1088/1742-6596/890/1/012163>
- Abdalla-Filho, E. (2022). Special Chapter: Personality Disorders in ICD-11. In E. Abdalla (Ed.), *Personality Disorders in the 10th and 11th Editions of the International Classification of Diseases* (pp. 226-265). Cambridge Scholars Publishing.
- Akoglu, H. (2018). User's Guide to Correlation Coefficients. *Turkish Journal of Emergency Medicine*, 18(3), 91-93. <https://doi.org/10.1016/J.TJEM.2018.08.001>

⁸ An image of the IDPI-11 materials is shown in Appendix J; in addition, the IDPI-Desktop™ scoring software can be accessed from the link: <https://doi.org/10.17605/OSF.IO/425N3>

- Ali, P., & Younas, A. (2021). Understanding and Interpreting Regression Analysis. *Evidence-Based Nursing*, 24(4), 116-118. <https://doi.org/10.1136/EBNURS-2021-103425>
- Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of Methods Used for Estimating Content Validity. *Research in Social and Administrative Pharmacy*, 15(2), 214-221. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Alonso, T., Peña, M. E., & Andreu, J. M. (2021). Psicopatía, agresión y violencia: un análisis de la interrelación en una muestra de delincuentes [Psychopathy, aggression, and violence: An interrelationship analysis in a criminal sample]. *Anuario de Psicología Jurídica*, 32(1), 61-69. <https://doi.org/10.5093/apj2021a25>
- American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR™)*. American Psychiatric Publishing.
- Arias, W. L., Canales, F. R., & De la Torre, N. L. (2016). Características psicopatológicas en los reclusos del Penal de Socabaya de Arequipa (Perú) [Psychopathological characteristics in prisoners from Socabaya Penitentiary, Arequipa (Peru)]. *Anuario de Psicología Jurídica*, 26(1), 80-87. <https://doi.org/10.1016/j.apj.2015.08.001>
- Arosquipa, Y. E., & Gutiérrez, R. J. A. (2016). *Trastornos y rasgos de personalidad según el delito cometido por internos varones de mínima seguridad del Establecimiento Penitenciario de Arequipa* [Disorders and Personality Traits According to the Crime Committed by Minimum Security Male Inmates of the Arequipa Penitentiary Establishment] [Bachelor's Thesis, National University of San Agustín]. UNSA Institutional Repository. <http://repositorio.unsa.edu.pe/handle/UNSA/3621>
- Awang, Z. (2014). *A Handbook on SEM*. MPWS Publisher.
- Ayinde, O. O., & Gureje, O. (2021). Cross-Cultural Applicability of ICD-11 and DSM-5 Personality Disorder. *Current Opinion in Psychiatry*, 34(1), 70-75. <https://doi.org/10.1097/YCO.0000000000000659>
- Bach, B. (2019, July 4). *Personality Disorders: Clinical rationale and application (ICD-11)*. [Video]. <https://www.youtube.com/watch?v=NjxrNXyPF3w>
- Bach, B., & Anderson, J. L. (2020). Patient-Reported ICD-11 Personality Disorder Severity and DSM-5 Level of Personality Functioning. *Journal of Personality Disorders*, 34(2), 231-249. https://doi.org/10.1521/pedi_2018_32_393
- Bach, B., & El Abiddine, F. Z. (2020). Empirical Structure of DSM-5 and ICD-11 Personality Disorder Traits in Arabic-Speaking Algerian Culture. *International Journal of Mental Health*, 49(2), 186-200. <https://doi.org/10.1080/00207411.2020.1732624>
- Bach, B., & Mulder, R. T. (2022). Empirical Foundation of the ICD-11 Classification of Personality Disorders. In S. K. Huprich (Ed.), *Personality Disorders and Pathology: Integrating Clinical Assessment and Practice in the DSM-5 and ICD-11 Era*. American Psychological Association.
- Bach, B., & Simonsen, S. (2021). How does Level of Personality Functioning Inform Clinical Management and Treatment? Implications for ICD-11 Classification of Personality Disorder Severity. *Current Opinion in Psychiatry*, 34(1), 54-63. <https://doi.org/10.1097/YCO.0000000000000658>
- Bach, B., Brown, T. A., Mulder, R. T., Newton-Howes, G., Simonsen, E., & Sellbom, M. (2021). Development and Initial Evaluation of the ICD-11 Personality Disorder Severity Scale: PDS-ICD-11. *Personality and Mental Health*, 15(3), 223-236. <https://doi.org/10.1002/PMH.1510>
- Bach, B., Kerber, A., Aluja, A., Bastiaens, T., Keeley, J. W., Claes, L., Fossati, A., Gutierrez, F., Oliveira, S. E. S., Pires, R., Riegel, K. D., Rolland, J. -P., Roskam, I., Sellbom, M., Somma, A., Spanemberg, L., Strus, W., Thimm, J. C., Wright, A. G. C., & Zimmermann, J. (2020). International Assessment of DSM-5 and ICD-11 Personality Disorder Traits: Toward a Common Nosology in DSM-5.1. *Psychopathology*, 53(3-4), 179-188. <https://doi.org/10.1159/000507589>
- Bach, B., Kramer, U., Doering, S., di Giacomo, E., Hutsebaut, J., Kaera, A., Panfilis, C. De, Schmahl, C., Swales, M., Taubner, S., & Renneberg, B. (2022). The ICD-11 Classification of Personality Disorders: a European Perspective on challenges and Opportunities. *Borderline Personality Disorder and Emotion Dysregulation*, 9(12), 1-11. <https://doi.org/10.1186/S40479-022-00182-0>

- Bach, B., Sellbom, M., Kongerslev, M., Simonsen, E., Krueger, R. F., & Mulder, R. T. (2017). Deriving ICD-11 Personality Disorder Domains from DSM-5 Traits: Initial Attempt to Harmonize Two Diagnostic Systems. *Acta Psychiatrica Scandinavica*, *136*(1), 108-117. <https://doi.org/10.1111/acps.12748>
- Bach, B., Somma, A., & Keeley, J. W. (2021). Editorial: Entering the Brave New World of ICD-11 Personality Disorder Diagnosis. *Frontiers in Psychiatry*. <https://doi.org/10.3389/fpsy.2021.793133>
- Baker, J., Edwards, I., & Beazley, P. (2021). Juror decision-making regarding a defendant diagnosed with borderline personality disorder. *Psychiatry, Psychology and Law*. <https://doi.org/10.1080/13218719.2021.1938273>
- Bartlett, P. (2011). DSM-5 and ICD-11 on Personality Disorder: A Lawyer's Perspective. *Personality and Mental Health*, *5*(2), 144-151. <https://doi.org/10.1002/pmh.165>
- Bayne, K. (2018). *Re: What is the Permissible Margin of Error when Computing a Social Research Sample Size?* Researchgate. https://www.researchgate.net/post/What_is_the_permmissible_margin_of_error_when_computing_a_social_research_sample_size/5bf367263d48b73b556f9d23/citation/download
- Bello, L. D. (2016, December 31). *Vídeo 16 Validez Convergente y Discriminante* [Video 16 Convergent and Discriminating Validity] [Video]. <https://www.youtube.com/watch?v=2ShNzkcQpmo>
- 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*(4), 332-346. <https://doi.org/10.1080/00223891.2011.583808>
- Bender, D. S., Skodol, A. E., First, M. B., & Oldham, J. (2018). Module I: structured clinical interview for the level of personality functioning scale. In M. B. First, A. E. Skodol, D. S. Bender, & J. M. Oldham (Eds.), *User's Guide for the Structured Clinical Interview for the DSM-5® Alternative Model for Personality Disorders (SCID-5-AMPD)*. American Psychiatric Association.
- Ben-Porath, Y. S., & Tellegen, A. (2020a). Chapter 3: The MMPI 3 Normative Sample and Derivation of Norms. In *Minnesota Multiphasic Personality Inventory-3 (MMPI-3): Manual for Administration, Scoring, and Interpretation* (pp. 17-22). University of Minnesota Press.
- Ben-Porath, Y. S., & Tellegen, A. (2020b). *Minnesota Multiphasic Personality Inventory-3 (MMPI-3): Manual for administration, scoring, and interpretation*. University of Minnesota Press. <https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/mmpi-3/mmpi-3-adminman-ch1.pdf>
- Ben-Porath, Y. S., & Tellegen, A. (2022). *Interpreting the MMPI-3*. University of Minnesota Press.
- Ben-Porath, Y. S., Heilbrun, K., & Rizzo, M. (2022). Using the MMPI-3 in Legal Settings. *Journal of Personality Assessment*, *104*(2), 162-178. <https://doi.org/10.1080/00223891.2021.2006672>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, *6*(149). <https://doi.org/10.3389/fpubh.2018.00149>
- Butcher, J. N., Hass, G. A., Greene, R. L., & Nelson, L. D. (2015). The MMPI-2 in Correctional Settings. In *Using the MMPI-2 in Forensic Assessment* (pp. 205-219). American Psychological Association. <https://doi.org/10.1037/14571-012>
- Cain, N. M., & Mulay, A. L. (2022). The DSM-5, the Alternative Model for Personality Disorders, and Their Empirical Support. In S. K. Huprich (Ed.), *Personality Disorders and Pathology: Integrating Clinical Assessment and Practice in the DSM-5 and ICD-11 Era*. American Psychological Association.
- Carlson, K. A. (1981). A Modern Personality Test for Offenders: The Carlson Psychological Survey. *Criminal Justice and Behavior*, *8*(2), 185-200. <https://doi.org/10.1177/009385488100800205>
- Carroll, A., Walvisch, J., & Marsh, T. (2022). Personality Disorders and Forensic Assessments: The Benefits of ICD-11. *Medicine, Science and the Law*, *1*. <https://doi.org/10.1177/00258024221094188>
- Christensen, A. P., Golino, H. F., & Silvia, P. J. (2020). A Psychometric Network Perspective on the Validity and Validation of Personality Trait Questionnaires. *European Journal of Personality*, *34*(6), 1095-1108. <https://doi.org/10.1002/per.2265>

- Cirlugea, O., Benuto, L. T., & Leany, B. D. (2013). Assessing Risk, Recidivism, and Dangerousness. In L. T. Benuto (Ed.), *Guide to Psychological Assessment with Hispanics* (pp. 377-398). Springer.
- Clark, L. A., Corona-Espinosa, A., Khoo, S., Kotelnikova, Y., Levin-Aspenson, H. F., Serapio-García, G., & Watson, D. (2021). Preliminary Scales for ICD-11 Personality Disorder: Self and Interpersonal Dysfunction Plus Five Personality Disorder Trait Domains. *Frontiers in Psychology, 12*, 2827. <https://doi.org/10.3389/fpsyg.2021.668724>
- Collier, J. E. (2020). Measurement Model Analysis. In *Applied Structural Equation Modeling Using AMOS*. Taylor & Francis.
- David, D., Lynn, S. J., & Montgomery, G. H. (Eds.). (2018). *Evidence-Based Psychotherapy: The State of the Science and Practice*. Wiley-Blackwell.
- Davison, S., & Janca, A. (2012). Personality Disorder and Criminal Behaviour: What is the Nature of the Relationship? *Current Opinion in Psychiatry, 25*(1), 39-45. <https://doi.org/10.1097/ycp.0b013e32834d18f0>
- Day, A., & Cook, J. (2019). Psychological Assessment in the Correctional Setting. In D. L. L. Polaschek, A. Day, & C. R. Hollin (Eds.), *The Wiley International Handbook of Correctional Psychology* (pp. 488-497). Wiley-Blackwell.
- De la Iglesia, G., & Castro, A. (2018). The Positive Personality Model (PPM): Exploring a New Conceptual Framework for Personality Assessment. *Frontiers in Psychology, 9*, 2027. <https://doi.org/10.3389/fpsyg.2018.02027>
- Deeks, J. J., & Bossuyt, P. M. M. (2021). Chapter 3: Evaluating diagnostic tests. Draft version. In J. J. Deeks, P. M. M. Bossuyt, M. M. G. Leeftang, & Y. Takwoingi (Eds.), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 2* (2th ed.). Cochrane.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Covering the Population and Selecting Who to Survey. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th ed.). Wiley.
- Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., & Muñoz, J. (2022). Decálogo para el análisis factorial de los ítems de un test [Decalogue for the Factor Analysis of the Test Items]. *Psicothema, 34*(1), 7-17. <https://doi.org/10.7334/PSICOTHEMA2021.456>
- Ferretti, F., Carabellese, F., Catanesi, R., Coluccia, A., Ferracuti, S., Schimmenti, A., Caretti, V., Lorenzi, L., Gualtieri, G., Carabellese, F., & Pozza, A. (2021). DSM-5 Personality Trait Facets Amongst Child Molesters: An Exploratory Comparison with Other Types of Offenders. *BMC Psychology, 9*(117), 1-13. <https://doi.org/10.1186/S40359-021-00619-1>
- First, M. B., Skodol, A. E., Bender, D. S., & Oldham, J. M. (2018). Module III: structured clinical interview for the personality disorders (including personality disorder-trait specified). In *User's Guide for the Structured Clinical Interview for the DSM-5® Alternative Model for Personality Disorders (SCID-5-AMPD)*. American Psychiatric Association.
- Frey, B. B. (Ed.). (2022). *The SAGE Encyclopedia of Research Design* (2th ed.). SAGE Publications, Inc.
- Gamache, D., Savard, C., Leclerc, P., Payant, M., Berthelot, N., Côté, A., Faucher, J., Lampron, M., Lemieux, R., Mayrand, K., Nolin, M. C., & Tremblay, M. (2021). A Proposed Classification of ICD-11 Severity Degrees of Personality Pathology Using the Self and Interpersonal Functioning Scale. *Frontiers in Psychiatry, 12*, 292. <https://doi.org/10.3389/fpsyg.2021.628057>
- Glen, S. (2016). *Tetrachoric Correlation: Definition, Examples, Formula*. StatisticsHowTo.Com: Elementary Statistics for the Rest of Us! <https://www.statisticshowto.com/tetrachoric-correlation/>
- Gomez, P. G. (2018). *Patrones clínicos de personalidad de internos con diagnóstico de adicciones de sustancias psicoactivas en el Establecimiento Penitenciario Huancayo - 2017* [Clinical Personality Patterns of Inmates Diagnosed with Addictions to Psychoactive Substances in the Huancayo Penitentiary Establishment - 2017] [Bachelor's thesis, Continental University]. Continental Institutional repository. <https://hdl.handle.net/20.500.12394/4988>
- Goth, K., Birkhölzer, M., & Schmeck, K. (2019). Assessment of Personality Functioning in Adolescents With the LoPF-Q 12-18 Self-Report Questionnaire. *Journal of Personality Assessment, 100*(6), 680-690. <https://doi.org/10.1080/00223891.2018.1489258>
- Grossman, S. D. (2017). *Therapeutic Applications of the Millon Clinical Multiaxial Inventory-IV (MCMI-IV)*. [Video]. <https://www.pearsonassessments.com/profes>

sional-assessments/blog-webinars/webinars/2017/11/therapeutic-applications-of-the-millon-clinical-multi-axial-inventory-iv-mcmi-iv-110917.html

- Grossman, S. D. (2019). The Millon Clinical Multiaxial Inventory-IV (MCMI-IV). In *The Cambridge Handbook of Clinical Assessment and Diagnosis* (pp. 249-262). Cambridge University Press. <https://doi.org/10.1017/9781108235433.018>
- Gutiérrez, F., Aluja, A., Ruiz, J., García, L. F., Gárriz, M., Gutiérrez-Zotes, A., Gallardo-Pujol, D., Navarro-Haro, M. V., Alabèrnia-Segura, M., Mestre-Pintó, J. I., Torrens, M., Peri, J. M., Sureda, B., Soler, J., Pascual, J. C., Vall, G., Calvo, N., Ferrer, M., Oltmanns, J. R., & Widiger, T. A. (2021). Personality Disorders in the ICD-11: Spanish Validation of the PiCD and the SASPD in a Mixed Community and Clinical Sample. *Assessment, 28*(3), 759-772. <https://doi.org/10.1177/1073191120936357>
- Halliwell, M. (2021). *Enhancing Mental Health Communications in Guyana*. MNS Disorders in Guyana's Jails: 1825 to the Present Day. <https://mnsguyana.le.ac.uk/2021/04/26/enhancing-mental-health-communications-in-guyana/>
- Hengartner, M. P., Zimmermann, J., & Wright, A. G. C. (2018). Personality Pathology. In *The SAGE Handbook of Personality and Individual Differences: Volume III: Applications of Personality and Individual Differences* (pp. 3-35). SAGE Publications Ltd. <https://doi.org/10.4135/9781526451248.n1>
- Huprich, S. K., Nelson, S. M., Meehan, K. B., Siefert, C. J., Haggerty, G., Sexton, J., Barry Dauphin, V., Macaluso, M., Jackson, J., Zackula, R., & Baade, L. (2018). Introduction of the DSM-5 levels of personality functioning questionnaire. *Personality Disorders: Theory, Research, and Treatment, 9*(6), 553-563. <https://doi.org/10.1037/PER0000264>
- 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), 192-197. <https://doi.org/10.1037/PER0000159>
- Hutsebaut, J., Weekers, L. C., Tuin, N., Apeldoorn, J. S. P., & Bulten, E. (2021). Assessment of ICD-11 Personality Disorder Severity in Forensic Patients Using the Semi-Structured Interview for Personality Functioning DSM-5 (STiP-5.1): Preliminary Findings. *Frontiers in Psychiatry, 12*, 505. <https://doi.org/10.3389/fpsy.2021.617702>
- Indrayan, A., & Holt, M. P. (2017). *Concise Encyclopedia of Biostatistics for Medical Professionals*. Taylor & Francis Group.
- Isvoranu, A.-M., Epskamp, S., Waldorp, L., & Borsboom, D. (Eds.). (2022). *Network Psychometrics with R: A Guide for Behavioral and Social Scientists*. Routledge.
- Kerber, A., Schultze, M., Müller, S., Rühling, R. M., Wright, A. G. C., Spitzer, C., Krueger, R. F., Knaevelsrud, C., & Zimmermann, J. (2022). Development of a Short and ICD-11 Compatible Measure for DSM-5 Maladaptive Personality Traits Using Ant Colony Optimization Algorithms. *Assessment, 29*(3), 467-487. <https://doi.org/10.1177/1073191120971848>
- Kim, Y. -R., Tyrer, P., & Hwang, S. -T. (2021). Personality Assessment Questionnaire for ICD-11 Personality Trait Domains: Development and Testing. *Personality and Mental Health, 15*(1), 58-71. <https://doi.org/10.1002/PMH.1493>
- King, B. M., Rosopa, P. J., & Minium, E. W. (2018). *Statistical Reasoning in the Behavioral Sciences* (7th ed.). John Wiley & Sons, Inc.
- Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42*(9), 1879-1890. <https://doi.org/10.1017/S0033291711002674>
- Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2013). *The Personality Inventory for DSM-5 Brief Form (PID-5-BF)*. American Psychiatric Association. <https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/assessment-measures#Personality>
- Kuhn, T. S. (2012). *The structure of scientific revolutions: 50th anniversary edition* (4th ed.). The University of Chicago Press.
- Levak, R. W., Siegel, L., Nichols, D. S., & Stolberg, R. (2011). *Therapeutic feedback with the MMPI-2: a positive psychology approach*. Taylor and Francis Group.

- Livesley, W. J. (2018). Integrated Modular Treatment. In W. J. Livesley & R. Larstone (Eds.), *Handbook of personality disorders: theory, research, and treatment* (pp. 645-673). The Guilford Press.
- Maples, J. L., Carter, N. T., Few, L. R., Crego, C., Gore, W. L., Samuel, D. B., Williamson, R. L., Lynam, D. R., Widiger, T. A., Markon, K. E., Krueger, R. F., & Miller, J. D. (2015). Testing whether the DSM-5 personality disorder trait model can be measured with a reduced set of items: An item response theory investigation of the personality inventory for DSM-5. *Psychological Assessment, 27*(4), 1195-1210. <https://doi.org/10.1037/pas0000120>
- Markon, K. E., Quilty, L. C., Bagby, R. M., & Krueger, R. F. (2013). The Development and Psychometric Properties of an Informant-Report Form of the Personality Inventory for DSM-5 (PID-5). *Assessment, 20*(3), 370-383. <https://doi.org/10.1177/1073191113486513>
- Martin-Key, N. A., Spadaro, B., Funnell, E., Barker, E. J., Schei, T. S., Tomasik, J., & Bahn, S. (2022). The Current State and Validity of Digital Assessment Tools for Psychiatry: Systematic Review. *JMIR Ment Health, 9*(3), e32824. <https://doi.org/10.2196/32824>
- McAdams, D. P., & Pals, J. L. (2006). A New Big Five: Fundamental Principles for an Integrative Science of Personality. *American Psychologist, 61*(3), 204-217. <https://doi.org/10.1037/0003-066X.61.3.204>
- McGrath, M., & Turvey, B. E. (2013). Ethical Issues for Treatment Staff in Forensic Settings. In B. E. Turvey & S. Crowder (Eds.), *Ethical Justice: Applied Issues for Criminal Justice Students and Professionals* (pp. 403-416). Elsevier.
- McMurran, M., & Howard, R. C. (2019). Personality Disorders and Offending. In L. L. Polaschek, A. Day, & C. R. Hollin (Eds.), *The Wiley International Handbook of Correctional Psychology* (pp. 265-281). John Wiley & Sons Ltd.
- Meaux, L., Cox, J., Edens, J. F., DeMatteo, D., Martinez, A., & Bownes, E. (2021). The Personality Assessment Inventory in U.S. Case Law: A Survey and Examination of Relevance to Legal Proceedings. *Journal of Personality Assessment, 104*(2), 179-191. <https://doi.org/10.1080/00223891.2021.1975723>
- Medina, D. E. (2021). *Patrones clínicos de personalidad en personas investigadas por delitos contra la libertad sexual, en la Sección de Psicología Forense PNP Huancayo, 2017-2019* [Clinical patterns of personality in people investigated for crimes against sexual freedom, in the PNP Huancayo Forensic Psychology Section, 2017-2019] [Bachelor's thesis, Continental University]. In Repositorio Institucional Continental. <https://hdl.handle.net/20.500.12394/10097>
- Mia, A., Motz, A., Logan, C., Bull, C., Nicklin, E., Shaw, J., Davies, J., Tew, J., Garvey, K., Scholey, K., Turner, K., Chauhan, M., Minoudis, P., O'Rourke, R., Paget, S., Simons, S., & Chuan, S. J. (2020). *Practitioner Guide: Working with people in the Criminal Justice System showing Personality Difficulties*. NHS England Publications. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1035881/6.5151_HMPPS_Working_with_Offenders_with_Personality_Disorder_accessible_version.pdf
- Millon, T., Grossman, S. D., & Millon, C. (2015). *Millon Clinical Multiaxial Inventory-IV manual* (4th ed.). Pearson Assessments.
- Miner, T. (2021, May 17). *Exploring the role of core positive selves with men convicted of child sexual offenses: A character strengths initiative*. Digital Repository UNM; Doctoral dissertation, University of New Mexico. https://digitalrepository.unm.edu/education_etds/106
- Moran, P. (2021). *Introducing the MMPI-3*. https://azpa.org/resources/Documents/Conventions/2021/10-9-21_800-930_MMPI-3_to_post.pdf
- Morey, L. C. (2007). *Personality assessment inventory (PAI): Professional manual* (2nd ed.). Psychological Assessment Resources.
- Morey, L. C. (2017). Development and initial evaluation of a self-report form of the DSM-5 level of personality functioning scale. *Psychological Assessment, 29*(10), 1302-1308. <https://doi.org/10.1037/PAS000045>
- Morey, L. C. (2019). On parsimony in personality disorder classification: Rejoinder to Sleep et al. (2019). *Psychological Assessment, 31*(10), 1206-1207. <https://doi.org/10.1037/pas0000774>

- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale Development: Ten Main Limitations and Recommendations to Improve Future Research Practices. *Psicologia: Reflexão e Crítica*, 30(3), 1-20. <https://doi.org/10.1186/S41155-016-0057-1>
- Mulder, R. T. (2021). ICD-11 Personality Disorders: Utility and Implications of the New Model. *Frontiers in Psychiatry*, 0, 709. <https://doi.org/10.3389/fpsy.2021.655548>
- Mulder, R. T., & Bach, B. (2022). Assessment and Treatment of Personality Disorders within the ICD-11 Framework. In S. K. Huprich (Ed.), *Personality Disorders and Pathology: Integrating Clinical Assessment and Practice in the DSM-5 and ICD-11 Era*. American Psychological Association.
- National Institute of Mental Health. (2019). *Estudio Epidemiológico de Salud Mental en Hospitales Regionales - 2015* [Epidemiological Study of Mental Health in Regional Hospitals - 2015]. <https://www.insm.gob.pe/investigacion/archivos/estudios/2020/VolXXXV2019Nro1EESMenHospitalesRegionales.pdf>
- Neal, T. M. S., Sellbom, M., & de Ruiter, C. (2022). Personality Assessment in Legal Contexts: Introduction to the Special Issue. *Journal of Personality Assessment*, 104(2), 127-136. <https://doi.org/10.1080/00223891.2022.2033248>
- Nguyen, T. H., Han, H. R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient - Patient-Centered Outcomes Research*, 7(1), 23-35. <https://doi.org/10.1007/s40271-013-0041-0>
- Olajide, K., Munjiza, J., Moran, P., O'Connell, L., Newton-Howes, G., Bassett, P., Akintomide, G., Ng, N., Tyrer, P., Mulder, R. T., & Crawford, M. J. (2018). Development and Psychometric Properties of the Standardized Assessment of Severity of Personality Disorder (SASPD). *Journal of Personality Disorders*, 32(1), 44-56. https://doi.org/10.1521/pedi_2017_31_285
- Olivas-Ugarte, L. O., & Cipriani-Delgado, C. S. (2022). Escala de timidez revisada (ETR-13): propiedades psicométricas en adolescentes peruanos [Revised Shyness Scale (ETR-13): Psychometric Properties in Peruvian Adolescents]. *Liberabit*, 28(1), e506. <https://doi.org/10.24265/liberabit.2022.v28n1.03>
- Oltmanns, J. R., & Widiger, T. A. (2018). A Self-Report Measure for the ICD-11 Dimensional Trait Model Proposal: The Personality Inventory for ICD-11. *Psychological Assessment*, 30(2), 154-169. <https://doi.org/10.1037/pas0000459>
- Oltmanns, J. R., & Widiger, T. A. (2019). Evaluating the Assessment of the ICD-11 Personality Disorder Diagnostic System. *Psychological Assessment*, 31(5), 674-684. <https://doi.org/10.1037/pas0000693>
- Oltmanns, J. R., & Widiger, T. A. (2020). The Five-Factor Personality Inventory for ICD-11: A Facet-Level Assessment of the ICD-11 Trait Model. *Psychological Assessment*, 32(1), 60-71. <https://doi.org/10.1037/pas0000763>
- Oltmanns, J. R., & Widiger, T. A. (2021). The Self- and Informant-Personality Inventories for ICD-11: Agreement, Structure, and Relations with Health, Social, and Satisfaction Variables in Older Adults. *Psychological Assessment*, 33(4), 300-310. <https://doi.org/10.1037/pas0000982>
- Padilla, J. A. (2019). *Análisis factorial exploratorio, SPSS, formato APA* [Exploratory factor analysis, SPSS, APA format]. [Video]. Estadística Con Joaquín Padilla. <https://www.youtube.com/watch?v=VxqkayFztTA>
- Pasowicz, M., & Piotrowski, P. (2021). Psychological resources in male prisoners: An application of the growth resources model. *Archives of Psychiatry and Psychotherapy*, 23(2), 24-37. <https://doi.org/10.12740/app/127670>
- Patrick, C. J. (2018). Psychopathy as Masked Pathology. In Author (Ed.), *Handbook of psychopathy* (2nd ed., pp. 3-21). Guilford Press.
- Peterson, C., & Seligman, M. E. P. (2004). *Character Strengths and Virtues: A Handbook and Classification*. Oxford University Press.
- Ramoz, N. (2022). Genetics of personality disorders. *Psychiatric Genomics*, 67-84. <https://doi.org/10.1016/b978-0-12-819602-1.00005-x>
- Reynolds, C., & Livingston, R. (2019). How to develop an empirically based psychological test. In G. Goldstein, D. N. Allen, & J. DeLuca (Eds.), *Handbook of Psychological Assessment*. Elsevier.

- Sahin, A., & Anil, D. (2017). The Effects of Test Length and Sample Size on Item Parameters in Item Response Theory. *Educational Sciences: Theory and Practice*, 17(1), 321-335. <https://doi.org/10.12738/estp.2017.1.0270>
- See, A. Y., Klimstra, T. A., Cramer, A. O. J., & Denissen, J. J. A. (2020). The Network Structure of Personality Pathology in Adolescence With the 100-Item Personality Inventory for DSM-5 Short-Form (PID-5-SF). *Frontiers in Psychology*, 11, 823. <https://doi.org/10.3389/fpsyg.2020.00823>
- Sellbom, M., Solomon-Krakus, S., Bach, B., & Bagby, R. M. (2020). Validation of Personality Inventory for DSM-5 (PID-5) Algorithms to Assess ICD-11 Personality Trait Domains in a Psychiatric Sample. *Psychological Assessment*, 32(1), 40-49. <https://doi.org/10.1037/pas0000746>
- Sellbom, M., Wygant, D. B., Tarescavage, A. M., & Ben-Porath, Y. S. (2022). Using the MMPI-3 in Forensic Psychological Assessments. In C. Garofalo & J. J. Sijtsema (Eds.), *Clinical Forensic Psychology: Introductory Perspectives on Offending* (pp. 515-537). Springer International Publishing. https://doi.org/10.1007/978-3-030-80882-2_26
- Siefert, C. J., Sexton, J., Meehan, K., Nelson, S., Haggerty, G., Dauphin, B., & Huprich, S. (2019). Development of a Short Form for the DSM-5 Levels of Personality Functioning Questionnaire. *Journal of Personality Assessment*, 102(4), 516-526. <https://doi.org/10.1080/00223891.2019.1594842>
- Sindeev, A., & Guzmán-Negrón, E. (2019). Perfil epidemiológico, clínico-psiquiátrico y legal del sentenciado por delitos sexuales, Establecimiento Penitenciario Lurigancho, Lima, 2017 [Epidemiological, clinical-psychiatric and legal profile of those sentenced for sexual crimes, Lurigancho Penitentiary Establishment, Lima, 2017]. *Acta Medica Peruana*, 35(3), 168-173. <https://doi.org/10.35663/amp.2018.353.510>
- Skodol, A. E., First, M. B., Bender, D. S., & Oldham, J. M. (2018). Module II: structured clinical interview for personality traits. In M. B. First, A. E. Skodol, D. S. Bender, & J. M. Oldham (Eds.), *User's Guide for the Structured Clinical Interview for the DSM-5® Alternative Model for Personality Disorders (SCID-5-AMPD)*. American Psychiatric Association.
- Sleep, C., Lynam, D. R., & Miller, J. D. (2021). Personality Impairment in the DSM-5 and ICD-11: Current Standing and Limitations. *Current Opinion in Psychiatry*, 34(1), 39-43. <https://doi.org/10.1097/ycp.0000000000000657>
- Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 24(1), 12. <https://doi.org/10.11613/BM.2014.003>
- Stikker, P. (2018). *Part 3c: Effect size*. Analysing a Nominal and Ordinal Variable. <https://peterstatistics.com/CrashCourse/3-TwoVarUnpair/NomOrd/NomOrd3c.html>
- Stover, J. B., Liporace, M. F., & Solano, A. C. (2020). Personality functioning scale: A scale to assess DSM-5's criterion a personality disorders. *Interpersona*, 14(1), 40-53. <https://doi.org/10.5964/ijpr.v14i1.3925>
- Stricker, J., & Pietrowsky, R. (2022). Incremental validity of the ICD-11 personality disorder model for explaining psychological distress. *Personality Disorders: Theory, Research, and Treatment*, 13(2), 97-107. <https://doi.org/10.1037/PER0000489>
- Thylstrup, B., Simonsen, S., Nemery, C., Simonsen, E., Noll, J. F., Myatt, M. W., & Hesse, M. (2016). 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*, 16(1), 1-8. <https://doi.org/10.1186/s12888-016-1011-6>
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 21(1), 19-25.
- Toop, C., Olver, M. E., & Jung, S. (2019). Forensic Assessment With the PAI in Correctional Samples: Implications for RNR: *Criminal Justice and Behavior*, 46(6), 866-883. <https://doi.org/10.1177/0093854819834718>
- Tyrer, P. (2017). *Personality Assessment Schedule-ICD-11 Version (PAS-ICD-11)*. Home edition.
- Tyrer, P., & Mulder, R. T. (2022). *Personality Disorder: From Evidence to Understanding*. Cambridge University Press.
- Tyrer, P., Mulder, R. T., Kim, Y. R., & Crawford, M. J. (2019). The Development of the ICD-11 Classification of

- Personality Disorders: An Amalgam of Science, Pragmatism, and Politics. *Annual Review of Clinical Psychology*, 15, 481-502. <https://doi.org/10.1146/annurev-clinpsy-050718-095736>
- VandenBos, G. R. (Ed.). (2015). *APA Dictionary of Psychology* (2nd ed.). American Psychological Association (APA).
- Watts, J. (2019). Problems with the ICD-11 classification of personality disorder. *The Lancet Psychiatry*, 6(6), 461-463. [https://doi.org/10.1016/S2215-0366\(19\)30127-0](https://doi.org/10.1016/S2215-0366(19)30127-0)
- Weekers, L. C., Hutsebaut, J., & Kamphuis, J. H. (2019). The Level of Personality Functioning Scale-Brief Form 2.0: Update of a brief instrument for assessing level of personality functioning. *Personality and Mental Health*, 13(1), 3-14. <https://doi.org/10.1002/pmh.1434>
- Whitman, M. R., & Ben-Porath, Y. S. (2021). Distinctiveness of the MMPI-3 Self-Importance and Self-Doubt Scales. *Journal of Personality Assessment*, 103(5), 613-620. <https://doi.org/10.1080/00223891.2021.1883628>
- Widiger, T. A. (2016). An Integrative Model of Personality Strengths and Weaknesses. In A. M. Wood & J. Johnson (Eds.), *The Wiley Handbook of Positive Clinical Psychology* (pp. 261-277). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118468197.ch18>
- Widiger, T. A., & Oltmanns, J. R. (2021). Personality and Personality Disorder. In O. P. John & R. W. Robins (Eds.), *Handbook of Personality: Theory and Research* (4th ed., pp. 755-772). Guilford Press.
- Williams, C. L., Butcher, J. N., & Paulsen, J. A. (2019). Overview of multidimensional inventories of psychopathology with a focus on the MMPI-2. In G. Goldstein, D. N. Allen, & J. DeLuca (Eds.), *Handbook of Psychological Assessment*. Academic Press.
- World Health Organization. (2022a). *ICD-11, the 11th Revision of the International Classification of Diseases*. <https://icd.who.int/browse11/l-m/es#/http%3A%2F%2Fid.who.int%2Ficd%2Fentity%2F941859884>
- World Health Organization. (2022b). *Personality Disorders and Related Traits*. In *ICD-11 Clinical Descriptions and Diagnostic Requirements*. WHO Department of Mental Health and Substance Use. <https://gcp.network/groupings/personality-disorders-and-related-traits/>
- Yang, B., Mallett, S., Takwoingi, Y., Davenport, C. F., Hyde, C. J., Whiting, P. F., Deeks, J. J., & Leeflang, M. M. G. (2021). QUADAS-C: A Tool for Assessing Risk of Bias in Comparative Diagnostic Accuracy Studies. *Annals of Internal Medicine*, 174(11), 1592-1599. <https://doi.org/10.7326/m21-2234>
- Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry*, 26(3), 171-177. <https://doi.org/10.3969/j.issn.1002-0829.2014.03>
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, 4(2), 165-178. <https://doi.org/10.15171/jcs.2015.017>
- Zimmermann, R., Steppan, M., Zimmermann, J., Oeltjen, L., Birkhölzer, M., Schmeck, K., & Goth, K. (2022). A DSM-5 AMPD and ICD-11 compatible measure for an early identification of personality disorders in adolescence - LoPF-Q 12-18 latent structure and short form. *MedRxiv*. <https://doi.org/10.1101/2022.05.19.22275330>

Luis Miguel Hualparuca-Olivera

Universidad Continental, Perú.

B. S. in Psychology. He is currently pursuing an MSc in Clinical and Health Psychology. Recently, he has collaborated in the translation of the PDS-ICD-11 into Spanish, together with colleagues Bo Bach, Ph. D., Martin Sellbom, Ph. D., Fernando Gutiérrez, Ph. D., Anton Aluja, Ph. D., and Silvia Aguirre, B. S.

ORCID: <https://orcid.org/0000-0002-2026-3524>

Corresponding author: luisolivera22@gmail.com

Dayana Nicol Ramos Campos

Universidad Continental, Perú.

B. S. in Psychology. She is currently working as a clinical psychologist in Uncanny.

ORCID: <https://orcid.org/0000-0001-6012-8215>

76304542@continental.edu.pe

Paola Andrea Arauco Vivas

Universidad Continental, Perú.

B. S. in Psychology.

ORCID: <https://orcid.org/0000-0001-6689-851X>

77230148@continental.edu.pe

Rocío Milagros Coz Apumayta

Universidad Continental, Perú.

Ph. D. in Education, B. S. in Psychology, professor at the School of Psychology of Continental University (UC), and thesis advisor.

ORCID: <https://orcid.org/0000-0002-9916-5547>

rcoz@continental.edu.pe

This study is based on the thesis:

Hualparuca, L. M., Ramos, D. N., & Arauco P, A. (2021). *Construcción y validación del Inventario Dimensional Integrativo de la Personalidad-11 (IDIP-11) en internos del Establecimiento Penitenciario de Huancayo* [Construction and Validation of the Dimensional Integrative Personality Inventory-11 (IDIP-11) in inmates of the Huancayo Penitentiary Establishment]. [Bachelor's degree thesis]. Continental Institutional Repository. <https://hdl.handle.net/20.500.12394/10507>