

Evaluación de

Cuestionario de

Personalidad Forense

 **CPF** 

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Cuestionario de Personalidad Forense
Autor	Pedro Manuel Medina García y Francesc Sintas Capdevilla
Autor de la adaptación española	--
Variable(s)	Divergencia Normativa Defensividad, Malestar, Evitación social, Eficacia, Sinceridad y Dominancia
Áreas de aplicación	- Ámbito penitenciario - Ámbito forense - Selección de personal
Soporte	Papel y lápiz, corrección automatizada via internet

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Adecuada-Buena	3.5
Fundamentación teórica	Adecuada, pero con algunas carencias- Adecuada	2.5
Adaptación	N/A (no aplicable)	--
Análisis de ítems	Adecuado, pero con algunas carencias- Adecuado	2.5
Validez: contenido	No se aporta	-
Validez: relación con otras variables	Adecuada-Buena	3.5
Validez: estructura interna	Adecuada	3
Validez: análisis del DIF	No se aporta	-
Fiabilidad: equivalencia	N/A (no aplicable)	-
Fiabilidad: consistencia interna	Adecuada	3
Fiabilidad: estabilidad	Adecuada-Buena	3.5
Fiabilidad: TRI	Adecuada-Buena	3.5
Fiabilidad inter-jueces	-	-
Baremos e interpretación de puntuaciones	Adecuada-Buena	3.8

Comentarios generales

El CPF se ha diseñado para cumplir una función social muy relevante en ámbito penitenciario: la de pronosticar la reincidencia de los reclusos. Se trata pues de un contexto aplicado muy relevante, donde son muy importantes las consecuencias derivadas de esta evaluación, tanto en el caso de falsos positivos como en el de falsos negativos. Además, a lo largo del manual los autores demuestran ser buenos conocedores del ámbito penitenciario y del contexto forense.

La aplicación de la prueba da lugar a una puntuación global (Divergencia Normativa; DN) y a seis puntuaciones específicas que reflejan variables de personalidad (Defensividad, Malestar, Evitación social, Eficacia, Sinceridad y Dominancia).

El test presenta una serie de puntos fuertes. Por ejemplo:

1) Es un test corto y con un lenguaje adaptado a las poblaciones específicas (p.ej. penitenciarias).

2) Presenta suficientes evidencias de validez y fiabilidad para justificar el uso de la puntuación global de Divergencia Normativa (DN) en el manejo de reclusos y la ayuda en la toma de decisiones (p.ej. permisos). Es de especial valor el estudio basado en la curva COR, para predecir la reincidencia, lo que ha requerido un seguimiento prolongado en el tiempo (de hasta 4 o 5 años, a partir del momento de aplicación del test).

3) Es también de valor el estudio experimental para el desarrollo de la escala para detectar la distorsión de respuestas, un riesgo siempre presente en la aplicación de este tipo de escalas.

4) Algunos procedimientos psicométricos están bien ejecutados y se han desarrollado con muestras grandes, lo cual representa un valor añadido.

Sin embargo, a lo largo de la obra también se reflejan algunas carencias y limitaciones:

1) Un punto débil de la escala, que afecta a la calidad de las evidencias de validez recogidas, es la ausencia de un modelo teórico de personalidad, especialmente en lo que atañe a las seis subescalas. Las escalas de personalidad que se contemplan han sido derivadas de una serie de análisis factoriales. No obstante, el origen de una prueba debería estar siempre en una buena definición de los constructos, basada en un modelo teórico -en este caso de personalidad- sólido. En el caso de que los modelos existentes se valoren inadecuados, debería crearse un modelo propio, definiendo los constructos y estableciendo relaciones entre los mismos y con otras variables. La obra carece de un modelo de personalidad que sustente a nivel teórico las seis dimensiones propuestas.

2) Las evidencias de fiabilidad y validez con respecto a las seis subescalas de personalidad son débiles. En primer lugar, se omite del manual información fundamental relativa a la consistencia interna de estas escalas. No obstante, a partir de la solución factorial se puede deducir que la consistencia interna será inadecuada para algunas de ellas. De igual manera, las evidencias de validez convergente son inadecuadas en algunos casos (p.ej., Eficiencia). Por último, la capacidad de algunas subescalas para discriminar poblaciones es reducida (p.ej., Sinceridad). En definitiva, convendría un análisis de la validez convergente, teniendo en cuenta una red nomológica más completa que incluya modelos de personalidad modernos y en boga, con una amplia evidencia empírica, como los Cinco Grandes o el HEXACO. De hecho, es previsible que algunas de las dimensiones que se miden tengan relación estrecha con constructos que ya están bien medidos por otros test (Cordialidad, Emocionalidad negativa, Extraversión, Responsabilidad, Honestidad y Narcisismo) y, por lo tanto, debería mostrarse en qué medida esas nuevas escalas tienen validez incremental sobre otras ya existentes.

3) Con respecto a dimensión general, Divergencia Normativa, los autores la definen como una tendencia a desviarse de la norma, que deriva de un balance entre la sensibilidad al castigo y a la recompensa. Aquí se asume (con cierta cautela) el marco teórico de Gray. No obstante, el manual aporta muy poca información sobre cómo fueron desarrollados los ítems a este respecto. Tampoco se considera la abundante literatura previa. Esto es relevante pues existen medidas de sensibilidad al refuerzo y al castigo, adaptadas en nuestro país (por ejemplo, el SPSRQ; Torrubia et al., 2001), que podrían utilizarse para analizar la validez convergente de su escala. También contrasta el marco teórico con el hecho de que ninguna de las escalas de personalidad del CPF esté a priori relacionada con ambos constructos (sensibilidad al castigo y sensibilidad al refuerzo). Por último, se hace una interpretación de la puntuación global, pero no se argumenta desde el punto de vista psicométrico la presencia de un factor general.

4) Con respecto a la representatividad de las muestras, los autores sugieren que la prueba

es aplicable en ámbitos laborales, más allá de la población penitenciaria, pero los baremos proporcionados no son representativos de esas poblaciones y, por tanto, no debería recomendarse su uso en esas situaciones. Tampoco se proporciona ninguna evidencia de validez de las interpretaciones en esos ámbitos.

5) Por último, convendría mejorar el manual, para proporcionar la información de una manera más exacta, organizada, compacta y completa. Por ejemplo:

a) El manual adolece de alguna confusión terminológica en cuanto a las evidencias de validez. Por ejemplo, los autores se refieren a evidencias de validez de contenido, cuando no se recogen evidencias que permitan valorar la relevancia de los ítems o la representación de los constructos que se pretende evaluar.

b) Con relación a evidencias de validez basadas en la estructura factorial la información es dispersa e incompleta, y es necesario acudir a los anexos de la prueba para encontrar información relevante para su consideración, sin que en el manual se haga un llamamiento explícito a dichos anexos.

c) El análisis de curvas ROC no se incluye en el apartado de evidencias de validez, sino en el de interpretación de las puntuaciones.

d) Se aplica el modelo de Respuesta Graduada en la muestra total ($N = 2855$), lo que es una fortaleza, pero no se hace mención ninguna al cumplimiento o evaluación de los supuestos de un modelo TRI, o a su ajuste, y se proporciona un gráfico en el que la función del error típico de estimación es difícil de visualizar.

Futuras ediciones del manual deberían: a) solventar estas confusiones terminológicas; b) incorporar evidencias de validez relevantes que ahora mismo son inexistentes en su versión actual (p.ej., sobre funcionamiento diferencial de los ítems); c) actualizar las referencias (excepto un TFG de 2019, todas son anteriores a 2017); d) proporcionar casos ilustrativos, que pueden ayudar a los profesionales, para la interpretación y para establecer las necesarias cautelas (informando, por ejemplo, del intervalo de confianza de las puntuaciones estimadas).

En resumen, la escala tiene algunas fortalezas derivadas de un trabajo continuado y muy valorable, pudiendo ser una buena fuente de información complementaria para el manejo de reclusos y, especialmente, en la predicción de la reincidencia a partir de la puntuación global. No obstante, se requeriría un modelo teórico más fuerte y la evidencia empírica recogida no es especialmente favorable al uso de las subescalas. Además, por el momento, no parece recomendable su uso en el ámbito organizacional.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Cuestionario de Personalidad Forense (CPF)

1.2. Nombre del test en su versión original:

--

1.3. Autor/es del test original:

Pedro Manuel Medina García y Francesc Sintas Capdevilla

1.4. Autor/es de la adaptación española:

--

1.5. Editor del test en su versión original:

Editorial Giunti EOS Psychometrics

1.6. Editor de la adaptación española:

--

1.7. Fecha de publicación del test original:

2021

1.8. Fecha de publicación del test en su adaptación española:

--

1.9. Fecha de la última revisión del test:

--

1.10. Área general de la/s variable/s que pretende medir el test:

Personalidad

1.11. Breve descripción de la/s variable/s que pretende medir el test:

Este es un test que ha sido desarrollado para evaluar la personalidad en contextos forenses, bajo la premisa de que los rasgos de personalidad permiten predecir el comportamiento penitenciario y la reincidencia. En definitiva, se pretende medir la tendencia al comportamiento delictivo de una persona, y por ende, su probabilidad de reincidencia.

La prueba se puede aplicar en el ámbito penitenciario (p.ej., predecir reincidencia delictiva) y forense (p. ej., pronóstico de cumplimiento de medidas impuestas por un juez o tribunal). Los autores sugieren que la prueba se puede aplicar también en el ámbito laboral, en puestos de trabajo en los que resulte esencial para el buen desempeño cumplir con los protocolos, las normas y seguir el reglamento correspondiente (p. ej., contables de grandes empresas, cuerpos de seguridad) o, en el lado opuesto, en los que sea importante el logro de objetivos mediante incentivos materiales o simbólicos, sin importar mucho los procedimientos para ello (p. ej., directivos de empresa de nuevos mercados).

Se trata de una prueba razonablemente corta, de solo 55 ítems con un formato de frecuencia de cinco categorías ("nunca o casi nunca", "algunas veces", "a menudo", "muchas veces", "siempre o casi siempre").

Según los autores, el test mide una dimensión general, la "DIVERGENCIA NORMATIVA" (DN), que constituye una tendencia a la desviación de la norma, explicada por un balance entre la sensibilidad al refuerzo y la sensibilidad al castigo. La puntuación DN deriva de una selección de ítems para predecir la reincidencia, operativizada como el ingreso en prisión posterior a la aplicación de la escala (con un seguimiento de cuatro a cinco años, según el caso).

La prueba también incluye una serie de subescalas específicas para medir dimensiones de personalidad. Estas escalas se han derivado empíricamente, a partir del análisis factorial de los ítems que conforman la escala principal, DN, y son:

- DEFENSIVIDAD. Indica la tendencia de los individuos a reconocer abiertamente que algo les molesta y a actuar en consecuencia para disminuir su incomodidad. Las personas que puntúan alto en este rasgo se comportan independientemente de cómo puedan pensar los demás.

- MALESTAR. Indica la tendencia de los sujetos a experimentar malos momentos. Se relaciona con una baja tolerancia a la frustración, intranquilidad emocional y autopercepción emocional inestable. Las personas que puntúan alto en este rasgo sienten que no tienen el control de su vida y atribuyen su sufrimiento a la mala suerte.

- EVITACIÓN SOCIAL. Indica la preferencia de los sujetos por estar la mayor parte del tiempo solos y por pasar desapercibidos. A las personas que puntúan alto en este rasgo les disgusta relacionarse con la gente, tanto en contextos de ocio como laborales.

- EFICACIA. Indica tendencia a la autodisciplina. Las personas que puntúan alto están orientados a la tarea y a los resultados. A las personas que puntúan alto en este rasgo les gusta mantenerse ocupados. Las personas que puntúan alto en este rasgo son sociables, receptivos a la crítica y a los retos personales.

- SINCERIDAD. Indica el grado en que los sujetos son conscientes de sus deseos y de aquello que rechazan, sin ocultar su forma de ver las cosas por quedar bien ante los demás. Las personas que puntúan alto en esta dimensión pueden ser percibidas como agresivas y poco tolerantes.

- DOMINANCIA. Indica la tendencia de los sujetos a estar orientados al éxito, al logro y a sus propios objetivos. Las personas que puntúan alto en esta dimensión son competitivas y narcisistas.

Además, incluye dos ítems de control de respuesta y dos escalas de control:

- MANIPULA. Es una escala intersticial, formada por 37 de los 55 ítems, que evalúa el grado de manipulación de las respuestas de los sujetos que contestan el test. Esta escala se desarrolla a partir de un estudio experimental, que incluye una condición de distorsión intencional.

- INFRECUENCIA. Pretende detectar si un evaluado emite patrones de respuesta muy poco usuales o extraños. Se contabiliza el número de veces que se ha seleccionado una opción de respuesta poco frecuente (elegida por menos del 6.2% de la muestra).

1.12. Áreas de aplicación:

Psicología forense, Psicología del trabajo y las organizaciones

1.13. Formato de los ítems:

Respuesta graduada (formato de frecuencia con cinco categorías ordenadas: "nunca o casi nunca", "algunas veces", "a menudo", "muchas veces", "siempre o casi siempre").

1.14. Número de ítems:

El test en total tiene 55 ítems y tiene 2 ítems de control (que no se emplean para el cálculo de ninguna puntuación). Es decir, hay 53 ítems que se combinan para obtener todas estas puntuaciones:

- Divergencia Normativa: 53 ítems.

Subescalas de personalidad:

- Defensividad: 11 ítems.

- Malestar: 12 ítems.

- Evitación social: 7 ítems.

- Eficacia: 7 ítems.

- Sinceridad: 5 ítems.

- Dominancia: 11 ítems.

Subescalas de validez:

- Manipula: 37 ítems.

- Infrecuencia: 14 ítems.

1.15. Soporte:

Papel y lápiz

1.16. Cualificación requerida para el uso del test de acuerdo con la documentación aportada:

Nivel C.

1.17. Descripción de las poblaciones a las que el test es aplicable:

El rango de edad propuesto por los autores es de 18 a 90 años.

En principio, la población de interés es la población penitenciaria. No es aplicable si la persona evaluada tiene problemas de lectura o no tiene un buen dominio del castellano. Se desaconseja el uso en personas con problemas de visión, ya que con la lectura del cuestionario aumenta el riesgo de inducir la respuesta e invalidar los resultados.

Los autores proponen la adecuación de esta prueba para determinados procesos de selección de personal, donde el cumplimiento/violación de las normas resulte fundamental.

1.18. Existencia o no de diferentes formas del test y sus características:

Es un cuestionario de lápiz y papel con una sola forma. No incluye administración informatizada.

1.19. Procedimiento de corrección:

Automatizado por ordenador (plataforma de corrección on-line).

1.20. Puntuaciones:

El proceso de obtención de puntuaciones es poco claro (p.ej., resulta difícil determinar si se utiliza la suma de las puntuaciones factoriales de regresión o las puntuaciones directas). El examinador introduce las respuestas directas en un programa informático. Las puntuaciones son procesadas automáticamente mediante la aplicación de "un algoritmo de cálculo" para devolver las puntuaciones normalizadas y generar un informe con el perfil de personalidad y una clave para la interpretación (página 65 del manual). Se indica que la corrección requiere que se hayan respondido todas las preguntas.

No se proporciona en el manual ningún ejemplo de este informe y tampoco se informa de la presencia o ausencia (ni de la cantidad) de ítems inversos presentes en el instrumento.

1.21. Escalas utilizadas:

Percentiles.

Puntuaciones estandarizadas T (media 50 y desviación típica 10).

1.22. Posibilidad de obtener informes automatizados:

Los autores y/o el editor no proporcionan en el Manual ningún ejemplo de informe automatizado para poder valorar estos aspectos.

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 12-15 minutos.

En aplicación colectiva: 12-15 minutos.

1.24. Documentación aportada por el editor:

Manual. Cuestionario.

1.25. Precio de un juego completo de la prueba:

CPF Juego completo: Manual Técnico + 20 hojas de registro + usos de corrección: 132 euros. Fecha de consulta: 18/04/22. Se entiende que la hoja de registro es el propio cuestionario.

1.26. Precio y número de ejemplares del paquete de cuadernillos:

CPF. Hoja de registro + uso de corrección: 5,70 euros. Fecha de consulta: 18/04/22. Se entiende que la hoja de registro es el propio cuestionario.

1.27. Precio y número de ejemplares del paquete de hojas de respuesta:

CPF. Hoja de registro + uso de corrección: 5,70 euros. Fecha de consulta: 18/04/22. Se entiende que la hoja de registro es el propio cuestionario.

1.28. Precio de la administración y/o corrección, y/o elaboración de informes por parte del editor:

(está incluido en el precio del paquete de cuadernillos, ya que solamente puede corregirse a través de la plataforma online)

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Buena	4
2.2. Calidad de la documentación aportada	Adecuada	3
2.3. Fundamentación teórica	Adecuada, con algunas carencias- Adecuada	2.5
2.4. Adaptación del test	N/A	-
2.5. Desarrollo de los ítems del test	Adecuada	3
2.6. Calidad de las instrucciones para el participante	Buena	4
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Adecuada-Buena	3.5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Adecuada	3
2.10. Datos sobre el análisis de los ítems	Adecuada, con algunas carencias	2.5

2.11. Validez:

2.11.1. Evidencias de validez de contenido:

Contenido	Valoración	Puntuación
2.11.1.1. Calidad de la representación del contenido o dominio	No se aporta información en la documentación	-
2.11.1.2. Consultas a expertos	No se aporta información en la documentación	-

2.11.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

2.11.2.1. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

Contenido	Valoración	Puntuación
2.11.2.1.1. Diseños empleados	Correlaciones con otros tests, Diferencias entre grupos, Diseños experimentales o cuasi experimentales	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con muestras grandes	5
2.11.2.1.3. Procedimiento de selección de las muestras	<p>La muestra para estos estudios procede de población reclusa de centros penitenciarios del Estado Español. Los autores informan (comunicación personal) que, sólo menos del 1% de reclusos se negaron a participar.</p> <p>La población diana es población penitenciaria, que puede resultar de difícil acceso, a pesar de lo cual el tamaño de las muestras empleadas es de tamaño moderado o grande (N = 389, para valorar las relaciones entre las escalas del CPF y de Cuestionario de Personalidad Situacional, CPS; N = 260, para un estudio experimental sobre la escala Manipula; N = 1502, para valorar las relaciones entre la puntuación en la escala total, DN, y algunas variables tales como trayectoria penitenciaria, tipología delictiva o tiempo de privación en libertad).</p> <p>También se incluyen algunas comparaciones con una muestra de control de la población general, formada por una muestra grande (N = 1145) y con una muestra de fuerzas y cuerpos de seguridad (N = 208). Para la muestra normalizada o control (N = 1145) los autores emplearon redes sociales reales consiguiendo un número inicial de colaboradores muy variados: familiares, amigos, compañeros de trabajo, vecinos y contactos que cada uno de ellos podía conseguir (página 29 del manual). Para la muestra de fuerzas y cuerpos de seguridad no se</p>	

	proporciona información sobre el procedimiento de selección.	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	Adecuada-Buena	3.5
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Adecuada con algunas carencias	2
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	No se aporta	-
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	No se aporta	-
2.11.2.1.8. Resultados de las diferencias intergrupo	Buena	4

2.11.2.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y un criterio:

Contenido	Valoración	Puntuación
2.11.2.2.1. Criterios empleados	El criterio empleado ha sido la reincidencia en una muestra diana de 370 reclusos de los cuales 173 fueron reincidentes y 197 fueron no reincidentes. La reincidencia se operativiza como el ingreso en prisión posterior a la aplicación del cuestionario (con un seguimiento de al menos cuatro años).	
2.11.2.2.2. Calidad de los criterios empleados	Adecuada	3
2.11.2.2.3. Relación temporal entre test y criterio	Predictivo	
2.11.2.2.4. Tamaño de las muestras	Adecuada	3
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental. Las muestras de población reclusa procedían de centros penitenciarios del Estado Español. La muestra es incidental. El carácter de reincidente o no reincidente se obtiene después de un seguimiento de al menos cuatro años desde la excarcelación.	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Buena ($0.45 \leq r < 0.55$)	4

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Adecuada	3

2.11.3.2. Funcionamiento diferencial de los ítems	-	-
--	---	---

2.11.4. Acomodaciones en la administración del test:

Contenido	Valoración	Puntuación
2.11.4. El manual del test informa sobre las acomodaciones en la administración del test	No	

2.11.5. Comentarios generales sobre evidencias de validez:

Los autores presentan las siguientes evidencias de validez:

- Evidencias de validez sobre la estructura interna. En uno de los anexos se muestran los resultados de ajuste de los análisis factoriales exploratorios de una versión inicial de la prueba (68 ítems) en una muestra inicial de tamaño grande formada por varones (N = 1030) o incluyendo mujeres (N = 1183). En general las decisiones para el análisis son adecuadas (p.ej., implementación óptima del análisis paralelo para decidir el número de factores, RULS como método de estimación, método de rotación oblicua). No obstante, en ocasiones, la información es incompleta, desorganizada y difícil de seguir. Por ejemplo, los modelos exploratorios aplicados muestran un ajuste satisfactorio (p.ej., CFI > 0,95), pero no se muestra la estructura factorial obtenida. En el manual (apartado 3.2.1, figura 2 y tabla 3.5) sí se muestran las saturaciones estimadas para un modelo confirmatorio estimado en la mitad de la muestra (N =1435) y sobre el test definitivo de 53 ítems, pero no hay información sobre la estimación de este modelo, no se indica el programa con el que se ha estimado y no se muestran los índices de ajuste. Por otro lado, atendiendo a este último modelo, algunos pesos son algo bajos (menores que 0.3) y algunas correlaciones entre factores excesivamente altas (p.ej., de 0,76). En realidad, cabe la duda de si los pesos mostrados en la tabla 3.5 se han obtenido realmente para un modelo confirmatorio, en el sentido que se entiende tradicionalmente, o, por el contrario, corresponde a un modelo exploratorio con rotación target, ya que en el anexo se refieren a algunos análisis factoriales exploratorios como confirmatorios. Por último, no queda claro que las muestras con las que se hacen los distintos análisis sean independientes, como se dice, ya que se indica que la muestra para el análisis confirmatorio se forma con la mitad de todas las muestras.

- Evidencias de validez de relación con otras variables. Las evidencias se pueden resumir en los siguientes puntos:

a) Se presentan evidencias sobre las relaciones de sus seis escalas de personalidad con el Cuestionario de Personalidad situacional (CPS), en una muestra de 389 reclusos. Se encuentran algunas relaciones entre adecuadas y buenas para cuatro de las seis dimensiones (Defensividad, con Agresividad (+); Malestar con Estabilidad emocional(-), Ansiedad(+), Confianza en sí mismo(-), Deseabilidad social(-), Autoconcepto(-) y Control cognitivo(-); Evitación social con Sociabilidad(-) y Dominancia con Liderazgo(+)) e inadecuadas para Eficacia (que correlaciona 0,194 con otra escala de Eficacia). Para la escala de Sinceridad su significado difiere del entendido en el lenguaje coloquial (los autores la entienden como una especie de autoconsciencia de preferencias personales y deseos), y no hay hipótesis claras a priori sobre las variables con las que debería relacionarse. Una limitación es que se indica que no se aplicaron las pruebas en el mismo momento, aunque no se proporciona un resumen sobre los intervalos temporales. En todo caso, la amplitud del intervalo puede haber contribuido a atenuar las relaciones.

b) Se presentan las correlaciones o diferencias entre grupos en DN, con respecto a diversas variables (edad, nivel de estudios, sexo, nacionalidad, trayectoria penitenciaria, tipología delictiva, etc.). En general, la presencia y la dirección de las relaciones sigue el patrón y la magnitud esperada. En la muestra penitenciaria (N = 1502), se encuentran diferencias por tipo de delito (robo > otros delitos; eta cuadrado = 0,066), trayectoria penitencia (presencia > ausencia de antecedentes de regresión de grado, eta cuadrado = 0,035) y comparaciones no significativas ($p > 0,05$) para sexo, nacionalidad (nacional vs. extranjero), situación penal (preventivo vs. penado) y tipo de estancia en centros (centro único vs. varios centros). En cuanto a las relaciones, estas son positivas con tiempo de privación ($r = 0,105$), y negativas con edad ($r = -0,199$) y nivel de estudios ($r = -0,174$). Debe considerarse que la evidencia

mostrada es, por tanto, favorable, si bien en algunos casos los tamaños del efecto esperados son muy pequeños y no se proporciona evidencia convergente/discriminante con respecto a otras medidas del constructo.

c) Los autores aportan las diferencias de medias en "Divergencia Normativa" y en las escalas de personalidad de varias poblaciones. La comparación más importante se hace con una muestra de 370 reclusos (173 reincidentes vs. 197 no reincidentes). Puede observarse que en los reincidentes la media en DN es mucho más elevada (eta cuadrado de 0,300). Los autores incluyen también información sobre la sensibilidad y especificidad en esa muestra, mediante un análisis de curvas COR, para determinar la utilidad de la escala principal (DN) para pronosticar la reincidencia. Los datos de este análisis son aceptables encontrando un AUC de 0,82, que puede considerarse bueno, e informando de los distintos valores de sensibilidad y especificidad para varias puntuaciones de corte de DN. Los autores no se comprometen con un punto de corte concreto, pero para un punto de corte de 50,51 se obtendría una sensibilidad de 0,762 y una tasa de falsos positivos del 17,7%.

Con respecto a las escalas de personalidad, se encuentran puntuaciones más elevadas para los reincidentes en Defensividad (eta cuadrado de 0,177), Malestar (eta cuadrado de 0,049) y Evitación social (eta cuadrado de 0,029), pero diferencias no significativas en Eficacia, Sinceridad y Dominancia. La muestra utilizada no parece totalmente independiente de la muestra a partir de la que se deriva la selección de los ítems (a la vista de lo mostrado en la tabla 4.5), por lo que puede haber habido cierta sobreestimación de la sensibilidad.

Por último, el anexo recoge un estudio experimental para validar la escala Manipula, diseñada para detectar la distorsión de las respuestas. La evidencia obtenida es favorable a la interpretación de las puntuaciones en esta escala, lo que supone una fortaleza del instrumento. No obstante, llama la atención que esta escala esté formada por 37 de los 53 ítems (cuando además la versión inicial se conformaba sólo de 22 de 68 ítems). Sería importante proporcionar las correlaciones entre las distintas puntuaciones y la escala Manipula.

Por otro lado, hay algunos aspectos que se pueden mejorar:

- En primer lugar, un punto débil de la escala que afecta a la calidad de las evidencias de validez que se recogen es la ausencia de un modelo teórico de personalidad. Las escalas de personalidad que se contemplan han sido derivadas empíricamente de una serie de análisis factoriales.

- En la misma línea, los autores hacen una interpretación de las puntuaciones en la dimensión general, divergencia normativa, como una tendencia a desviarse de la norma, producida por un conflicto de tipo atracción-repulsión, basado en la sensibilidad hacia la recompensa respecto al castigo. Si esto es así, convendría que hubieran utilizado medidas específicas de sensibilidad al refuerzo y al castigo para analizar la validez convergente de su escala. También es llamativo que ninguna de las escalas de personalidad que ellos identifican factorialmente en el CPF esté a priori relacionada con ambos constructos (sensibilidad al castigo y sensibilidad al refuerzo).

- En tercer lugar, la definición operativa de reincidencia vs no reincidencia es poco clara y cambia de un estudio a otro. Parece que, inicialmente, la reincidencia se define a la vez retrospectiva y prospectivamente (i.e., antecedentes de ingreso en prisión en el momento de hacer el test o que habiendo sido excarcelado después de hacer el test, haya reincidido en los dos años posteriores a la excarcelación). Se entiende que esta ha sido la definición de reincidencia para la selección de 70 ítems iniciales (para predecir la reincidencia mediante regresión logística). Posteriormente, la reincidencia se define prospectivamente, y con un periodo más amplio (que, después de la aplicación del test, haya reincidido en los cuatro años posteriores a la excarcelación, tenga antecedentes penales o no). Sin embargo, de acuerdo con la Tabla 4.5, parece excluirse al grupo con antecedentes penales del grupo de no reincidentes. Estas definiciones pueden ser importantes porque afectan al significado de la curva COR.

- En cuarto lugar, llama la atención que los ítems del cuestionario hayan sido escogidos para maximizar la predicción de la reincidencia, pero que, sin embargo, varias de las subescalas (Eficacia, Sinceridad y Dominancia) no muestren diferencias significativas entre los dos grupos (reincidentes vs. no reincidentes; ver Tabla 2.1 del Anexo). Convendría profundizar en este resultado.

- En quinto lugar, convendría incluir una evaluación del funcionamiento diferencial de los ítems, incluir evidencias de validez de contenido y el contraste de una estructura bi-factorial, con un factor general y seis específicos, que permita avalar el modelo propuesto por los autores.

- Por último, el manual adolece de ciertas confusiones terminológicas en cuanto a las evidencias de validez. Por ejemplo:

a) La información incluida en el epígrafe de Evidencias de validez de contenido (epígrafe 3.2.2 del manual) no recoge lo que se entiende en la literatura psicométrica por "evidencias de validez de

contenido" (ver por ejemplo los manuales de Muñiz (1996) o Abad et al (2011)). No hay una recogida de evidencias que permita valorar la relevancia de los ítems o la representación de los constructos que se pretende evaluar. En este apartado se presentan evidencia de las relaciones de las escalas de personalidad del CPF con otro instrumento el CPS. Esta información encajaría mejor en el apartado "Evidencias basadas en las relaciones con otras variables".

b) En la validación de la escala de Manipulación se hace referencia a la obtención de una matriz multirasgo-multimétodo, pero únicamente se está comprobando la convergencia de las medidas para medir un mismo constructo.

c) En ocasiones, el manual resulta difícil de seguir, sin que queden claras las características de las muestras que se están analizando (p.ej., las muestras de $N = 1030$ y $N = 1183$) o su solapamiento de sujetos.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad (para cada escala o subescala) Coeficientes de fiabilidad para diferentes grupos de personas Cuantificación del error mediante TRI (Función de información u otros)	

2.12.2. Equivalencia formas paralelas:

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	-	-
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	-	-
2.12.2.3. Promedio de coeficientes de equivalencia	-	-

2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado o con alguna muestra grande y otras pequeñas	4
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa o KR-20 Omega ordinal GLB	
2.12.3.3. Promedio de coeficientes de consistencia	Adecuada, pero con algunas carencias ($0.60 \leq r < 0.70$)	2

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Varios estudios con muestras grandes	5
2.12.4.2. Coeficientes de estabilidad	Adecuada, pero con	2

	algunas carencias ($0.55 \leq r < 0.65$)	
--	--	--

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	Un estudio con una muestra grande	3
2.12.5.2. Coeficientes proporcionados	Función de información	
2.12.5.3. Tamaño de los coeficientes	Buena ($5 \leq \text{información} < 10$)	4

2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	-	
2.12.6.2. Promedio de los coeficientes	-	-

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

Los autores proporcionan las siguientes evidencias sobre fiabilidad:

1) Consistencia interna. El coeficiente alfa de Cronbach para la escala de DN es excelente (superior a 0,85) en las diferentes muestras consideradas (tabla 3.4 del manual), aunque no excesivamente elevado considerando la longitud del test (53 ítems). Este indicador se complementa con el coeficiente Omega ordinal, basado en el análisis factorial con variables categóricas. Las muestras utilizadas para calcular estos valores son de tamaño grande ($N > 1000$) o moderado en el caso de la población diana ($N = 370$). Se proporcionan también las correlaciones ítems-test corregidas para la escala DN, siendo todas aceptables ($> 0,22$).

El principal problema en este punto es que se omite del manual la información relativa a la fiabilidad de las seis escalas de personalidad, que es una información fundamental, teniendo en cuenta que es un instrumento para evaluar personalidad (los autores asumen que DN es la única dimensión subyacente al test, lo cual es contradictorio con el uso que se propone para las subescalas). Si se obtiene el coeficiente omega a partir de los pesos estandarizados en el análisis factorial (Tabla 3.5 del manual), se observa que para tres de las escalas (Evitación social, Eficacia y Sinceridad) se obtienen valores inadecuados (próximos o menores que 0,6). Para Dominancia y Defensividad se encuentra un valor adecuado (en torno a 0,7) y para Malestar un valor bueno (en torno a 0,8).

2) Fiabilidad test-retest. En primer lugar, en el estudio de la estabilidad temporal los autores consideran tres muestras distintas, a dos de ellas se les administraron diferentes versiones del CPF y los intervalos temporales entre el test y el retest también fueron diferentes en las tres muestras (Tabla 3.1 del manual), con intervalos de tiempo muy variables dentro de cada muestra. Dos de las muestras tienen un tamaño grande ($N > 200$). Se informa de la correlación test-retest para las escalas de DN y para las seis escalas de personalidad. Para dos de los grupos (mediana del intervalo de tiempo entre tres y cuatro meses), el coeficiente promedio es de 0,61 (con un rango de 0,387 a 0,730), mientras que para el grupo de presos en el que se aplicó una versión CPF de 214 ítems con un mayor intervalo temporal (la mediana era de aproximadamente seis meses), las correlaciones encontradas son próximas a cero. Los autores proporcionan una explicación, haciendo referencia a efectos iatrogénicos de la prisión.

3) Función de información del test. Los autores ajustan un modelo de teoría de respuesta al ítem (TRI), el modelo de Respuesta Graduada de Samejima para la muestra total ($N = 2855$). Se utilizó la función de información para seleccionar los ítems del test definitivo (reduciendo la longitud de 68 a 53 ítems), aunque no queda claro el modo exacto (p.ej., si se consideró la función de información para un nivel de rasgo en torno a un punto de corte o para un intervalo de rasgo). El manual no hace mención al cumplimiento o evaluación de los supuestos de un modelo TRI, o del ajuste del modelo. No obstante, se presenta un gráfico con la función de información y el error típico (página 40). En la figura se puede

observar que el pico máximo de información del test está en un nivel de habilidad cercano a 1 (un valor entre 9 y 10) y que la escala mide con buena precisión a lo largo de la mayor parte de los niveles de rasgo.

2.13. Baremación e interpretación de las puntuaciones:

2.13.1. Interpretación normativa de las puntuaciones:

Contenido	Valoración	Puntuación
2.13.1.1. Calidad de las normas	Se incluyen varios baremos dirigidos a diversos estratos poblacionales/Los baremos son buenos para el ámbito forense/penitenciario pero no para la población objetivo del ámbito laboral.	3.5
2.13.1.2. Tamaño de las muestras	Grande ($600 \leq N < 1000$)	4
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	<p>Incidental. Las muestras de población reclusa procedían de centros penitenciarios del Estado Español.</p> <p>Para la muestra normalizada o control los autores emplearon redes sociales reales consiguiendo un número inicial de colaboradores muy variados: familiares, amigos, compañeros de trabajo, vecinos y contactos que cada uno de ellos podía conseguir.</p>	
2.12.1.5. Actualización de baremos	No se aporta información en la documentación.	-

2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	-	-
2.13.2.2. Procedimiento empleado para fijar los puntos de corte	-	
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	-	
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	-	-

2.13.3. Comentarios generales sobre baremación e interpretación de las puntuaciones:

Se presentan baremos para cinco grupos diferentes: Muestra penitenciaria no reincidente (N=197), Muestra penitenciaria reincidente (N =173), Muestra penitenciaria (N=1502), Muestra no penitenciaria (control o normalizada) (N = 1145) y Muestra de cuerpos y fuerzas de seguridad (N=208). Por tanto, las muestras pueden considerarse de tamaño suficiente (en tres de los casos) o muy grande (en los otros dos)[Promedio N = 645]. Las muestras están compuestas mayoritariamente por varones (un 91% en la muestra penitenciaria; un 89% en la muestra de control).

Los baremos presentados para cada grupo solamente incluyen las puntuaciones transformadas (centiles y puntuaciones T), tanto para la escala principal de DN como para las seis escalas de personalidad y también para la escala Manipula. Para las escalas de personalidad no se proporcionan puntos de corte concretos, pero sí para la escala Manipula, en la que una puntuación inferior a 40 podría indicar un intento de distorsionar las respuestas y se asocia a una sensibilidad del 96% en muestra penitenciaria. También existe un punto de corte para la escala de Infrecuencia de respuesta. Estos puntos de corte están bien justificados. Por otro lado, con respecto a la escala DN el evaluador puede utilizar la tabla de percentiles de la curva COR (tabla 5.1.) para elegir un punto de corte con una determinada sensibilidad y especificidad (en función del momento y las circunstancias de la evaluación).

En general, el establecimiento de los baremos y las puntuaciones típicas es correcto y adecuado, y ejecutado de forma apropiada. Por otro lado, dadas las características de las muestras de baremación es claro que estas no se pueden considerar suficientemente representativas para la utilización de la escala en el ámbito laboral.

Algunos aspectos que se podrían mejorar:

- En el manual se informa de que se obtienen las puntuaciones típicas normalizadas, lo que puede dar a entender que se proporcionan las puntuaciones T normalizadas. No obstante, la observación de los baremos permite deducir que se han utilizado las puntuaciones estandarizadas, sin normalizar.
- No se proporcionan puntos de corte claros para las interpretaciones de las subescalas de personalidad. Aunque se trate de puntuaciones T y percentiles, convendría que los autores proporcionen una guía clara al respecto.
- Es importante informar de la precisión de las puntuaciones (i.e., intervalos de confianza derivados del error típico de medida). En el manual no se hace ninguna referencia a este aspecto, por lo que sería fundamental añadir esa información.
- Para algunas de las escalas (Eficacia, Sinceridad y Dominancia), las interpretaciones propuestas no están fundamentadas en la evidencia recogida, dadas las bajas correlaciones encontradas con otras variables. De igual manera, los valores de consistencia interna para Evitación social, Eficacia y Sinceridad son inadecuados.
- Convendría obtener muestras representativas del ámbito laboral, en el que se propone aplicar la prueba (p.ej., como mínimo con igual representación por sexos), así como recoger evidencias sobre relativas a la validez de las interpretaciones de las puntuaciones en este ámbito, ya que apenas se proporciona evidencia al respecto.
- No se aporta información sobre la actualización de baremos, lo que en este caso es relevante explicitar, pues es claro que la recogida de datos se ha prolongado en el tiempo y es necesario informar de los periodos de recogida de datos para poder valorar este punto, ahora y en el futuro.