Consejo General de la Psicología
ESPAÑA

Miembro de la Federación Europea de Asociaciones de Psicólogos

Conde Peñalver, 45. 5º Izq.
28006 Madrid – España
Tel.: +34 91 444 90 20
Fax: +34 91 309 56 15
secop@cop.es
www.cop.es

Evaluación de

Inventario Clínico Multiaxial de Millon- IV

MCMI-IV

Prueba valorada en la ronda de evaluaciones del año 2019

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Inventario Clínico Multiaxial de Millon - IV (MCMI - IV)
Autor	Theodore Millon, Seth Grossman y Carrie Millon
Autor de la adaptación española	Ana Hernández, Cristina Aguilar y Frédérique Vallar. Departamento de I+D Pearson Clinical & Talent Assessment en colaboración con María Pilar Sánchez y Violeta Cardenal
Variable(s)	Personalidad, psicopatología y dominios funcionales en el ámbito clínico
Áreas de aplicación	Psicología clínica, forense, neuropsicologia, investigación
Soporte	Lápiz y papel, online

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Buena - Excelente	4,3
Fundamentación teórica	Buena	4
Adaptación	Buena - Excelente	4,5
Análisis de ítems	Adecuado, con algunas carencias	2
Validez: contenido	Adecuada	3
Validez: relación con otras variables	Adecuada - Buena	3,4
Validez: estructura interna	-	-
Validez: análisis del DIF	-	-
Fiabilidad: equivalencia	-	-
Fiabilidad: consistencia interna	Buena	4
Fiabilidad: estabilidad	Adecuada	3
Fiabilidad: TRI	-	-
Fiabilidad inter-jueces	-	-
Baremos e interpretación de puntuaciones	Buena	4

Nota. El signo – se interpreta como no se aporta información o bien no procede.

Comentarios generales

El test MCMI-IV está orientado a la evaluación de la personalidad y la psicopatología en pacientes adultos/as que buscan o reciben atención o tratamiento psicológico o psiquiátrico. Según el manual del test, se puede utilizar también en los ámbitos forense, neuropsiológico, de orientación y terapia de pareja, y de investigación.

Es un cuestionario de autoinforme con 195 afirmaciones que se responden como verdaderas o falsas y proporcionan medidas en 70 escalas con interpretación clínica y otras cinco indicadoras de la validez de las respuestas. Entre las escalas con interpretación clínica, 12 miden Patrones Clínicos de Personalidad (e.g., Esquizoide, Dependiente, Compulsivo), 3 miden Patologías Graves de la Personalidad (e.g., Límite), 7 miden Síndromes Clínicos (e.g., Ansiedad Generalizada, Consumo de Sustancias)y 3 miden Síndromes Clínicos Graves (e.g., Depresión Mayor). Cada una de las 15 escalas relacionadas con la personalidad lleva asociadas otras tres escalas que aportarían información más detallada y que en conjunto forman las 45 facetas de Grossman que se han introducido por primera vez en la versión IV del test MCMI.

Entre los puntos fuertes del test destacan una tradición de más de 40 años y su permananente actualización. Por ejemplo, en el siglo XXI se han editado en nuestro entorno las versiones II, III y IV además de una versión específica para adolescentes (MACI) y otra para pacientes con diagnóstico médico (MBMD). Su base en la teoría de la personalidad de Millon está suficientemente explicada en el manual donde se hace referencia también a las similitudes y diferencias con los trastornos de la personalidad tal como se valoran en el DSM-5. En cambio, en el manual está menos explicada la base teórica para la interconexión de las facetas de Grossman con las escalas de personalidad. El proceso de construcción de los ítems se ha desarrollado a lo largo de varios años y con la participación de diversos profesionales aunque en el manual de la presente edición se atribuyen los nuevos ítems al juicio clínico del autor principal y la revisión de materiales publicados. Finalmente, no queda clara la redundancia o diferenciación en los contenidos de cada escala puesto que con 195 ítems se construyen 70 escalas formadas por entre 8 y 21 ítems cada una, lo que se consigue mediante la reutilización de los mismos ítems para distintas finalidades. Los ítems que se consideran prototípicos de cada escala se decidieron en una fase inicial con la ayuda de un análisis factorial confirmatorio. El proceso de adaptación lingüística al idioma castellano cubre el proceso que sería de esperar incluyendo la participación de un equipo de personas con diversos puntos de vista, la realización de un estudio piloto con la población diana, y un estudio cuantitativo, todo ello antes de pasar a redactar la versión definitiva con la que se construirían los baremos.

Los materiales son de elevada calidad, bien impresos, claros y fáciles de usar, las instrucciones para ello son explícitas y el proceso de registro de respuestas es sencillo y rápido, en el bienentendido que para la interpretación de puntuaciones se necesita ser especialista en la teoría de Millon.

La parte cuantitativa del desarrollo y de la adaptación del test están menos respresentadas en el manual y en cambio, serían muy útiles para complementar las pruebas en favor de la credibilidad de las distintas escalas. Concretamente, en el desarrollo de la versión estadounidense se cita un análisis de ítems inicial y también análisis factoriales confirmatorios que, según se dice en el manual, darían soporte a las agrupaciones de los ítems en las escalas clínicas y en las facetas de Grossman, pero en el manual no se presentan datos ni referencias sobre ello. En cambio, estos datos serían extremadamente útiles para clarificar la reutilización de ítems antes mencionada. La misma consideración es de aplicación a la adaptación española. En este sentido, parecería muy aconsejable un estudio de la equivalencia entre la versión original y la adaptación española, puesto que algunos ítems se asociaron a las escalas por su relevancia estadística en la muestra estadounidense, e incluso alguna escala al completo, como la de inconsistencia, se basó en resultados estadísticos de la muestra estadounidense.

Se valoran positivamente las muchas precauciones que se toman en la corrección y puntuación del test para determinar si el proceso de respuesta ha sido válido. Se tienen en cuenta el número de ítems en blanco o con doble marca, y también se evalúan preguntas con respuesta previsible, inconsistencia entre respuestas, y estilos de respuesta que incluyen franqueza o reserva al responder, deseabilidad social y tendencia a la devaluación. Estos indicadores se usan bien

para invalidar el cuestionario o bien para ajustar los resultados en función del estilo de respuesta. La corrección automatizada del test es ágil y proporciona un informe completo para cubrir la finalidad de orientar la interpretación de resultados.

Las puntuaciones directas se obtienen ponderando doblemente los ítems prototípicos de cada escala y de forma simple el resto de ítems. Se deduce que una puntuación más elevada para cada ámbito es indicativa de una mayor problemática, aunque sería oportuno clarificar el papel que juegan los ítems directos e inversos en la obtención de estas puntuaciones.

La interpretación de puntuaciones es totalmente normativa basada en la distribución de respuestas de la muestra de tipificación. Tanto en el manual como en el informe automatizado se presentan las equivalencias entre puntuaciones directas y dos tipos de puntuaciones normativas: las tasas base, propias de los test de Millon, y el baremo en percentiles, una puntuación mucho más divulgada. Las puntuaciones en tasa base se construyeron para clasificar a las personas en tres grupos (atributo no presente, atributo presente o nivel tipo de personalidad y atributo prominente o nivel trastorno de personalidad) de manera que dicha clasificación fuera concordante con la prevalencia de cada trastorno en la sociedad. Se valora postivamente el hecho de que estos datos se calcularon específicamente para la adaptación española y, aunque las distribuciones de prevalencia se basan en los datos de la muestra de tipificación, se considera un punto fuerte el hecho de que dichas distribuciones encajaron con datos epidemiólogicos oficiales en nuestro entorno. También se valora positivamente el esfuerzo realizado para asegurar en el grupo normativo la representación de diversas zonas geográficas españolas, así como la transparencia en cuanto a su composición respecto a sexo, edad, nivel educativo y tratamiento ambulatorio o con ingreso. En cambio, necesitaría más discusión la decisión que se ha tomado en relación con las diferencias en tasas base respecto al sexo y el estado del paciente (ingresado/ambulatorio). En el manual se ha optado por presentar un solo baremo e indicar que para la evaluación de cada paciente pueden tomarse en cuenta las diferencias de media por sexo y estado. Ello parece algo laborioso de realizar tras una corrección automatizada como es la de este test.

La prueba de validez más importante que se aporta en relación con variables externas viene a apoyar la interpretación prevista en términos de puntos de corte. Se trata del estudio de la sensibilidad y la especificidad de la clasificación de las pacientes a través de su autoinforme respecto al criterio clínico aportado por su terapeuta. Hay que interpretar los resultados con precaución puesto que no son datos obtenidos de forma completamente independiente, pero los resultados muestran valores de sensibilidad y especificidad que, con no ser excelentes, sí son acordes con el consejo que se da en el manual de utilizar los resultados del MCMI-IV de forma complementaria con otras informaciones disponibles sobre la persona.

Otras pruebas vienen a apoyar la validez convergente y discriminante de las escalas. El patrón de correlaciones obtenido con la versión anterior MCMI-III puede considerarse convincente para las personas que ya confiaban en este test. A su vez, el patrón de correlaciones obtenido con el test de cribado BSI-18 puede resultar convincente para un público más amplio. En ambos casos, en el manual se comentan fundamentalmente las correlaciones que apoyan la validez convergente del MCMI-IV. Puesto que el MCMI-III, el MCMI-IV y el BSI-18 miden varios constructos distintos, sería muy útil si se indicara de forma exhaustiva qué coeficientes de correlación serían indicadores de validez convergente, cuáles serían indicadores de validez discriminante y el grado de apoyo que aportan los resultados en cada caso.

Finalmente, se valoran positivamente las evidencias de fiabilidad que se presentan en el manual. En cuanto a fiabilidad de consistencia interna, los resultados son buenos y basados en una muestra grande. También son buenos los resultados de estabilidad temporal aunque basados en una muestra más reducida. De acuerdo con el pretendido uso individualizado del test en ámbitos como el forense, sería deseable la publicación del error estándar de medida, un dato muy informativo y más interpretable por un mayor número de colectivos sociales.

La información que se aporta en el manual del test MCMI-IV sustenta en conjunto la calidad psicométrica del test para su uso en la evaluación de la personalidad clínica basada en la teoría de la personalidad de Millon y las aportaciones de Grossman. Sería necesario proporcionar información que apoye en datos concretos el resto de usos previstos en el manual en los ámbitos forense, neuropsiológico, de orientación y terapia de pareja, y de investigación.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Inventario Clínico Multiaxial de Millon - IV (MCMI - IV)

1.2. Nombre del test en su versión original:

Millon Clinical Multiaxial Inventory - IV

1.3. Autor/es del test original:

T. Millon, S. Grossman y C. Millon

1.4. Autor/es de la adaptación española:

A. Hernández, C. Aguilar y F. Vallar. Departamento de I+D Pearson Clinical & Talent Assessment, en colaboración con M.P. Sánchez y V. Cardenal

1.5. Editor del test en su versión original:

NCS Pearson Inc

1.6. Editor de la adaptación española:

Pearson Educación

1.7. Fecha de publicación del test original:

2015 (corrección a la errata advertida en el manual)

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

2018

1.9. Fecha de la última revisión del test:

2018

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

Personalidad, escalas clínicas

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

El MCMI - IV está orientado a la evaluación de la personalidad y la psicopatología de pacientes adultos que buscan o reciben atención o tratamiento psicológico o psiquiátrico.

Las escalas de esta prueba se basan en la teoría evolutiva de la personalidad de Millon. Consiste en un cuestionario de autoinforme con 195 ítems verdadero-falso, que miden 12 patrones clínicos de la personalidad (Esquizoide, Evitativo, Melancólico, Dependiente, Histriónico, Tempestuoso, Narcisista, Antisocial, Sádico, Compulsivo, Negativista, y Masoquista); 3 patologías graves de la personalidad (Esquizotípico, Límite, y Paranoide); 7 síndromes clínicos (Ansiedad Generalizada, Síntomas Somáticos, Espectro Bipolar, Depresión Persistente, Consumo de Alcohol, Consumo de Sustancias, y Estrés Postraumático); y 3 síndromes clínicos graves (Espectro Esquizofrénico, Depresión Mayor y Delirante). Por poner algunos ejemplos, el patrón clínico de la personalidad denominado como Esquizoide se caracteriza en este test por falta de intensidad en el deseo y el dolor, indiferencia a las relaciones sociales, apatía, y necesidades afectivas mínimas. La patología grave de la personalidad denominada Esquizotípica se caracteriza por el deseo de aislamiento social, estar absorto/a en si mismo/a, excentricidad, desconfianza o aplanamiento emocional. Finalmente, el síndrome clínico de Espectro Esquizofrénico se caracteriza por mostrar conductas incongruentes, confusión, pensamiento fragmentado, sentimientos aplanados y/o aislamiento. Dentro de cada escala, se incluyen una serie de ítems prototípicos que caracterizan los aspectos de la personalidad que son distintivos de ese trastorno y complementan la interpretación de los resultados.

En esta versión se han adaptado las facetas de Grossman a la población española con el fin de aportar información más detallada para cada patrón clínico de la personalidad y cada patología grave de la personalidad. Para cada una de las escalas de personalidad se ofrecen tres facetas, sumando un total de 45 facetas derivadas de dominios funcionales y estructurales. Por ejemplo, el patrón clínico de la personalidad Esquizoide se detalla por las facetas Personalmente desvinculado, Contenido escaso y Temperamentalmente apático, mientras que la patología grave de la personalidad Esquizotípica se detalla por las facetas Cognitivamente circunstancial, Autoimagen disociada y Contenido caótico.

Así pues, el test proporciona información sobre 70 escalas o facetas con utilidad clínica. Adicionalmente, el MCMI-IV incluye cinco índices de validez de las respuestas: Indicadores de respuestas aleatorias (2 escalas), un Índice de Sinceridad, un Índice de Deseabilidad Social, y un Índice de Devaluación.

1.12. Áreas de aplicación:

Psicología clínica, Neuropsicología, Psicología forense, Investigación

1.13. Formato de los ítems:

Respuesta dicotómica verdadero/falso

1.14. Número de ítems:

195 ítems clasificados en 25 escalas clínicas, cinco de validez y 45 facetas de Grossman (un mismo ítem puede pertenecer a más de una escala y/o faceta). En las escalas clínicas los ítems se distribuyen así: Esquizoide (15), Evitativo (18), Melancólico (19), Dependiente (14), Histriónico (17), Tempestuoso (17), Narcisista (16), Antisocial (14), Sádico (14), Compulsivo (13), Negativista (18), Masoquista (18), Esquizotípico (18), Límite (27), Paranoide (24), Ansiedad Generalizada (13), Síntomas somáticos (10), Espectro Bipolar (13), Depresión Persistente (21), Consumo de Alcohol (8), Consumo de Drogas (11), Estrés Postraumático (14), Espectro Esquizofrénico (21), Depresión Mayor (17), Delirante (14). En las escalas de validez: Invalidez (3), Inconsistencia (50), Sinceridad (121), Deseabilidad Social (24), y Devaluación (30). Los mismos ítems se clasifican en las 45 facetas de Grossman con un rango de 7 a 12 ítems por faceta. Un ejemplo ilustrativo de la composición de las distintas escalas y facetas es el ítem 30. El ítem 30 es prototípico de la escala Histriónico y forma parte de la escala Esquizoide y de la escala

Tempestuoso; además, también forma parte de la faceta Contenido escaso que, por su parte, aportaría información más detallada sobre la escala Esquizoide, pero no sobre la escala Tempestuoso. Otro tipo de ejemplo es el ítem 15, que forma parte de la faceta de Grossman Interpersonalmente aversivo y por ello aportaría información más detallada sobre la escala Evitativo, pero en cambio, no es un ítem que forme parte de la escala Evitativo. Otro ejemplo más, ya en el ámbito de los índices de validez, es la escala de Sinceridad, formada por casi dos tercios de los ítems del cuestionario.

1.15. Soporte:

Papel y lápiz, informatizado

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:

Personas adultas a partir de 18 años que buscan o reciben atención o tratamiento psicológico o psiquiátrico con una comprensión lectora igual o superior a la de educación primaria, sin diagnóstico de enfermedad mental o física que interfiera con la evaluación.

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

Sólo cambia el soporte

1.19. Procedimiento de corrección:

Automatizado por ordenador

1.20. Puntuaciones:

La obtención de las puntuaciones se realiza únicamente mediante una plataforma online. Este sistema calcula las puntuaciones directas a partir de los ítems que componen cada escala, ponderando doblemente los ítems prototípicos. Se deduce que una puntuación más elevada en cada ámbito es indicativa de mayor problemática, aunque sería oportuno detallar esta información. Por ejemplo, aunque se informa tangencialmente de la presencia de ítems directos e inversos, no se explica qué papel juegan en la obtención de las puntuaciones directas.

Los resultados obtenidos en las escalas de validez se interpretan a partir de la puntuación directa. Por ejemplo, el test se considera inválido si en la escala de Sinceridad, formada por 121 ítems, se ha obtenido una puntuación de 0 a 6 o bien de 115 a 121.

La puntuación de las escalas clínicas puede ser ajustada en función de los resultados obtenidos en las escalas de Sinceridad, Ansiedad y Depresión mayor. Su interpretación se realiza atendiendo a baremos y puntos de corte que se comentarán más adelante en el apartado de este informe destinado a ello. En el manual también se recomienda complementar la interpretación de los resultados atendiendo a las respuestas dadas a los ítems prototípicos de cada escala.

1.21. Escalas utilizadas:

Percentiles, Tasas base

1.22. Posibilidad de obtener informes automatizados:

Sí. El informe incluye las siguientes secciones: Portada con datos personales, Aspectos de interpretación, Resumen de las puntuaciones y perfil, Facetas de Grossman con la puntuación más alta, Puntuaciones en las facetas de Grossman, Respuestas significativas y Respuestas a los ítems.

El informe es completo, y pretende orientar a la interpretación de los resultados, diagnósticos y problemas susceptibles de una evaluación posterior, en su caso. Incluye comentarios sobre la validez del perfil, el estilo de respuesta, y el ajuste de las escalas si corresponde.

1.23. Tiempo estimado para la aplicación del test:

30 minutos

1.24. Documentación aportada por el editor:

Manual, cuaderno y hoja de respuestas, acceso a correción online

1.25. Precio de un juego completo de la prueba:

225 € (14/06/2019)

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

25 cuadernillos, 64 € (14/06/2019)

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

25 hojas de respuesta, 64 € (14/06/2019)

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

Incluido en el precio de las hojas de respuesta (14/06/2019)

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Buena - Excelente	4,5
2.2. Calidad de la documentación aportada	Buena	4
2.3. Fundamentación teórica	Buena	4
2.4. Adaptación del test	Buena - Excelente	4,5
2.5. Desarrollo de los ítems del test	Bueno	4
2.6. Calidad de las instrucciones para el participante	Buena - Excelente	4,5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Buena	4
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Excelente	5
2.10. Datos sobre el análisis de los ítems	Adecuados, con algunas carencias	2

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	Adecuada	3
2.11.1.2. Consultas a expertos	Adecuada	3

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 de sexo y de estado (ingreso/ambulatorio)	
2.11.2.1.2. Tamaño de las muestras	Un estudio con una muestra moderada (200 < N ≤ 500) o varios estudios con muestras pequeñas (N < 200)	2
2.11.2.1.3. Procedimiento de selección de las muestras	Incidental. Obtenida a través de psiquiatras y profesionales de la psicología con amplia experiencia en el uso del MCM-III, siguiendo para la selección de pacientes las indicaciones de Pearson Educación con criterios de inclusión detallados.	
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	Excelente (r ≥ 0.70)	5
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	Adecuada, pero con algunas carencias	2
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	-	-
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	Criterio diagnóstico de profesionales aplicado a la muestra de tipificación procedente de la población de pacientes de salud mental.	
2.11.2.2.2. Calidad de los criterios empleados	Buena	4
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente	
2.11.2.2.4. Tamaño de las muestras	Un estudio con una muestra grande (N ≥ 200)	3
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental. Obtenida a través de psiquiatras y psicólogos/as con amplia experiencia en el uso del MCM-III, siguiendo para la selección de pacientes las indicaciones de Pearson Educación con criterios de inclusión detallados. Se procuró que participaran pacientes de casi todas las comunidades autónomas de España, y que procedieran de diversos ámbitos de atención psicológica y con distintas situaciones clínicas.	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Buena (0.45 ≤ 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	-	-
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:

Las evidencias de validez basadas en el contenido se valoran en parte como muy convincentes y en otra parte como muy discutibles. Entre los aspectos convincentes, se destaca que en el manual se proporciona suficiente información sobre el modelo teórico de Millon y también sobre las condiciones de administración del test y cómo interpretar los resultados. También la adaptación española del test se considera buena, destacando la participación de un equipo de profesionales y la realización de un estudio piloto. Son puntos más débiles la información proporcionada sobre la elaboración de los nuevos ítems para la versión IV así como la falta de información sobre el apoyo teórico para la determinación de las facetas de Grossman. Finalmente, el hecho de que en el manual se sitúe la posición del test respecto a los ejes diagnósticos del DSM-5 ha sido valorado como un punto fuerte en una revisión por considerarse una forma de clarificar la interpretación de los resultados en el ámbito clínico. En cambio, ha sido destacado como un punto débil en otra de las revisiones por considerar criticable el DSM-5 en este sentido.

En esta fase de desarrollo inicial del test se valora positivamente el análisis de ítems basado en análisis factorial para depurar los ítems incluidos en cada escala, aunque se echa de menos la publicación, ni que fuera sucinta, de los principales resultados tanto del análisis que se realizó para determinar los ítems en las escalas clínicas como del que se realizó para definir las facetas de Grossman. También se sugiere considerar la oportunidad de realizar un estudio de funcionamiento diferencial de los ítems para apoyar la equivalencia entre el test americano y la adaptación española. Un ejemplo claro de la necesidad de discutir al menos esta opción es el índice de inconsistencia (escala W) basado en los 25 pares de ítems con las correlaciones más altas en la muestra de tipificación americana que se han conservado sin discusión en la adaptación española. En el manual, la estructura interna del test sólo queda apoyada empíricamente a partir del análisis visual de la matriz de correlaciones entre las 25 escalas clínicas.

Las evidencias principales de validez del MCMI–IV proceden de su correlación con la anterior versión del mismo, el MCMI-III, con el test BSI-18, y también del juicio clínico de las profesionales que han colaborado en la tipificación de este inventario, utilizado este último para analizar la sensibilidad, la especificidad y los valores predictivos de las escalas.

Respecto a la correlación con el BSI-18, hay que señalar que éste es un instrumento de cribado que, según datos publicados, se puede considerar válido para la población española de pacientes de psiquiatría. El BSI-18 aporta cuatro indicadores de psicopatología: somatización, depresión, ansiedad, y un índice general de gravedad. En conjunto, el promedio de las correlaciones de las distintas escalas del MCMI-IV con los componentes del BSI-18 se puede considerar adecuado. Sin embargo, existe una gran dispersión entre las distintas escalas, estando en algunos casos por debajo de un valor de r = ,10 (como en las escalas de los patrones clínicos Narcisista y Antisocial), y en otros por debajo de un valor mínimo de r = ,20 (patrón clínico Compulsivo, o los síndromes clínicos del Espectro bipolar y de Consumo de drogas). Probablemente esta dispersión pueda atribuirse al hecho de que algunas de las escalas del BSI-18 miden constructos similares a algunas de las escalas del test MCMI-IV, y su correlación constituiría evidencia de validez convergente, mientras que otras escalas miden constructos conceptualmente distintos, y su correlación constituiría evidencia de validez discriminante. Sería una mejora relevante del manual que esto se comentara con toda claridad. También sería muy útil establecer hipótesis previas sobre las correlaciones esperadas entre las distintas escalas para facilitar la evaluación de los resultados que se publican.

El mismo comentario vale para el estudio de las relaciones entre el MCMI-IV y su versión anterior MCMI-III. En este caso, la prueba se considera informativa y potencialmente convincente para las personas profesionales que ya confiaban en la versión anterior.

Las evidencias más relevantes sobre la validez de este instrumento provienen del estudio de la sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo a partir del juicio clínico de los/as profesionales que colaboraron en la tipificación del instrumento. La mayoría de escalas tienen una sensibilidad inferior a ,60, y una especificidad superior a ,70. Los valores predictivos positivos y negativos se comportan adecuadamente conforme aumenta la tasa de prevalencia del trastorno. Así pues, los niveles de eficiencia diagnóstica de las escalas del MCMI-IV son acordes con el consejo que se da en el manual de utilizar sus resultados de forma complementaria con otras informaciones. Sin embargo, la gran variabilidad entre escalas hace recomendable utilizar con mayor prudencia las conclusiones que se puedan derivar de algunas de ellas. Como limitación a estas evidencias, hay que señalar que el criterio de los/as profesionales que colaboraron en categorizar la gravedad de los trastornos de los pacientes se solapa con el cálculo de las tasas de prevalencia en las que se basa el cálculo de las Tasas base utilizadas para establecer los puntos de corte para el diagnóstico con el MCMI-IV. Esto implica un grado de redundancia entre ambos juicios, que limita su consideración como criterios externos independientes.

La comparación de las puntuaciones del test obtenidas en función de grupos definidos por el sexo y el tratamiento ambulatorio o mediante ingreso proporcionó algunas diferencias que según se comenta en el manual eran de esperar y que podrían ser útiles como referencia en el momento de evaluar a cada persona.

Finalmente, tomando la validez como el grado en que se aportan evidencias que dan soporte a la interpretación de las puntuaciones para los usos previstos del test, se puede concluir que queda claro el apoyo que recibe su uso para la evaluación de la personalidad clínica. Sería necesario ampliar la discusión del valor de las evidencias aportadas para sustentar la interpretabilidad de las puntuaciones para los usos previstos en los ámbitos forense, de investigación y neuropsicológico.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad para cada escala o subescala	

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	Un estudio con una muestra grande (N < 200)	3
2.12.3.2. Coeficientes de consistencia interna presentados	alfa ordinal	
2.12.3.3. Promedio de coeficientes de consistencia	Excelente (r ≥ 0.85)	5

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Un estudio con una muestra pequeña (N < 100)	1
2.12.4.2. Coeficientes de estabilidad	Excelente (r ≥ 0.80)	5

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	-	-
2.12.5.2. Coeficientes proporcionados	-	
2.12.5.3. Tamaño de los coeficientes	-	-

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:

El estudio acerca de la fiabilidad de las puntuaciones de MCMI-IV se basa por un lado en la consistencia interna (coeficiente alfa ordinal) a partir de la muestra de tipificación de pacientes de salud mental españoles. Estos valores muestran un nivel excelente de consistencia interna para estudios grupales (> ,80) del conjunto de las escalas del instrumento. Tan sólo en una de las facetas de Grossman de una de las escalas (escala de patrón clínico Dependiente, faceta Interpersonalmente sumiso) se encuentra un valor no totalmente adecuado (,68).

Por otro lado, las autoras de esta adaptación han calculado la estabilidad de las escalas del instrumento mediante el procedimiento test-retest, con una muestra pequeña de la misma población, mostrando en la práctica totalidad de los casos valores de fiabilidad considerados adecuados en la fase de desarrollo de un cuestionario (r > ,70), excepto en tres facetas de Grossman de tres patrones clínicos diferentes (Narcisista, Antisocial y Sádico).

Sin embargo, se echa de menos la publicación del error estándar de medida que daría información más aplicable al pretendido uso individual del test y, además, sería más interpretable por un mayor número de colectivos potencialmente implicados, algo especialmente relevante si, tal y como se indica en el manual, se pretende avalar el test también para uso forense.

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	Un baremo para la población objetivo	3
2.13.1.2. Tamaño de las muestras	Grande (600 ≤ N < 1000)	4
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	Misma muestra que en apartado 2.11.2.2.5	
2.12.1.5. Actualización de baremos	Excelente (menos de 10 años)	5

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:

Para ayudar en el diagnóstico y la interpretación de los resultados, las puntuaciones directas se convierten mediante interpolación lineal en puntuaciones de Tasa base que son un tipo de puntuaciones derivadas específicas de este test. Para la interpetación de las Tasas base se han fijado cinco valores de anclaje(0, 60, 75, 85 y 115), asociando el valor de 60 a la media de las puntuaciones directas. Los puntos de corte de 75 (atributo presente o nivel tipo de personalidad) y 85 (atributo prominente o nivel trastorno de personalidad) se establecen para ajustarse la prevalencia de cada característica en la población de pacientes de salud mental españoles.

Dicha prevalencia ha sido obtenida a partir del diagnóstico proporcionado por las personas profesionales que participaron en la administración de los cuestionarios de la muestra de tipificación. En el futuro sería útil disponer de datos epidemiológicos oficiales para determinar los puntos de corte de las Tasas base. Su falta se palía en el manual indicando que se ha comprobado que las prevalencias así obtenidas coinciden con datos epidemiológicos oficiales publicados por el Observatori del Sistema de Salut de Catalunya.

Hay que destacar que también se puede consultar el baremo en percentiles, lo que se considera adecuado puesto que se trata de un tipo de baremo muy divulgado y de fácil interpretación.

Así pues, la interpretación de los resultados es únicamente de tipo normativo y los puntos de corte estarían apoyados por su sensibilidad y especificiad diagnóstica, características que se han comentado en el apartado de este informe relativo a la validez. Se utiliza un solo baremo sin separar por sexo ni por estado del paciente, pero se comentan las diferencias de medias encontradas en función de estas características y se aconseja tenerlas en cuenta como referencia.

En cuanto a la calidad de la muestra de tipificación, esta se considera relativamente grande al estar formada por cerca de 900 personas; actualizada, puesto que se obtuvo en el año 2017; y con evidencia de haber realizado un esfuerzo encomiable por representar las diversas zonas de la geografía española y por informar de su composición respecto al sexo, la edad, el nivel educativo, estado civil y situación respecto al tratamiento.