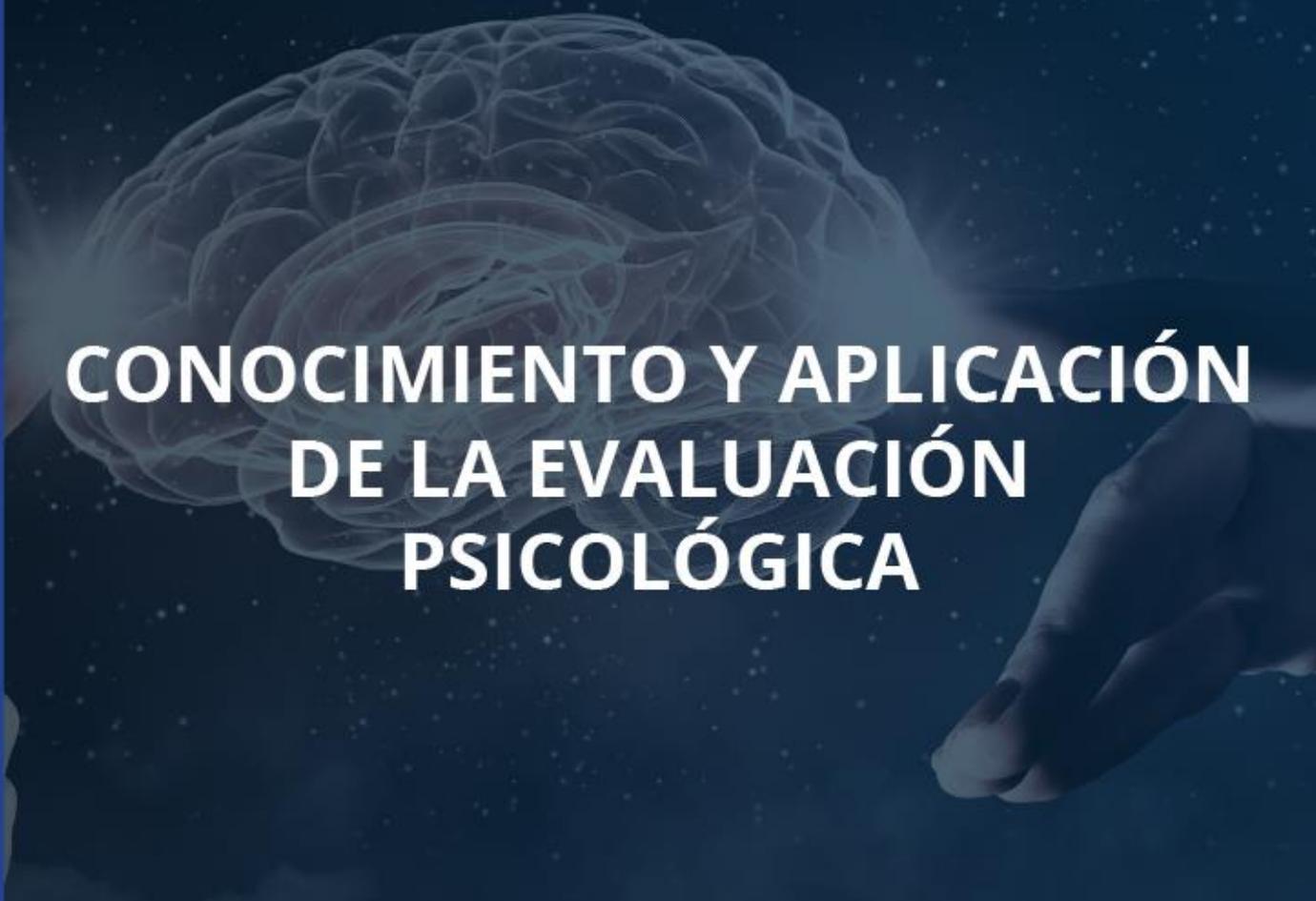




Capítulo 7/94

Tratado de psiquiatría clínica  
Massachusetts General Hospital  
2017 - 2018

A hand is shown holding a glowing, translucent brain against a dark blue background with a starry pattern. The brain is the central focus, with light rays emanating from it.

**CONOCIMIENTO Y APLICACIÓN  
DE LA EVALUACIÓN  
PSICOLÓGICA**

*Trabajamos para su tranquilidad...*

# Conocimiento y aplicación de la evaluación psicológica

*Mark A. Blais, PsyD*

*Samuel Justin Sinclair, PhD*

*Sheila M. O'Keefe, EdD*

## Puntos clave

### Contexto

- La evaluación psicológica es un servicio de consulta profesional cuyo objetivo es aportar a los clínicos una imagen más compleja, sobre una base empírica, de sus pacientes en lo relativo a varios dominios clínicos relevantes (p. ej., funcionamiento intelectual y neurocognitivo, psicopatología y diagnóstico clínico, estilo de personalidad).
- La evolución de la evaluación psicológica en el último siglo ha producido una mejora de la fiabilidad y la validez clínica, que a su vez ha mejorado la asistencia al paciente y los resultados clínicos.

### Historia

- La psicometría es un campo de estudio muy amplio dedicado al desarrollo y la aplicación de nuevos instrumentos diseñados para medir distintos constructos psicológicos.
- Históricamente, ha habido múltiples enfoques de la evaluación psicológica, con modelos racionales y de base empírica para el desarrollo de instrumentos psicológicos.
- La evaluación psicológica es un proceso complejo que incorpora un enfoque multimétodo a la recogida e integración de los datos, como medio de maximizar su utilidad clínica.

### Retos para la investigación y la clínica

- Los instrumentos psicológicos deben someterse primero a evaluaciones estrictas de fiabilidad y validez antes de su aplicación, que además debe ser reproducida en distintos entornos y poblaciones.

### Apuntes prácticos

- Una de las ventajas de la evaluación psicológica es su capacidad de cuantificar y aclarar presentaciones psiquiátricas confusas.
- Cada vez es más frecuente que los pacientes reciban información sobre los hallazgos de la prueba directamente del psicólogo evaluador.
- Se ha demostrado que la evaluación psicológica con información directa al paciente facilita el proceso de tratamiento.

## Perspectiva general

La evaluación psicológica es un servicio de consulta con gran potencial para aumentar y mejorar el conocimiento de los clínicos sobre sus pacientes y facilitar el proceso de tratamiento. A pesar de esto, las interconsultas de evaluación psicológica son infrutilizadas en el entorno asistencial actual de salud mental. Esto resulta desafortunado a la vista de los sólidos indicios de que las pruebas psicológicas consiguen por lo general coeficientes de fiabilidad y validez similares a los de muchas pruebas diagnósticas médicas estándar.<sup>1</sup> Este capítulo contiene una revisión detallada de los elementos de una evaluación psicológica y expone las posibles ventajas de una interconsulta para evaluación. Con este fin, revisaremos los métodos usados para construir instrumentos psicológicos válidos, los grupos principales de pruebas psicológicas (incluidos ejemplos detallados de cada uno), y la aplicación y utilidad de estos instrumentos en la valoración clínica. También se presentan los aspectos relacionados con el orden de las pruebas psicológicas y la integración de las conclusiones de un informe de evaluación en el proceso de tratamiento.

## **Psicometría: ciencia del desarrollo de pruebas**

La psicometría es un campo de estudio muy amplio dedicado al desarrollo y la aplicación de nuevos instrumentos diseñados para medir distintos constructos psicológicos (p. ej., depresión, impulsividad, estilo de personalidad). Clásicamente, se han usado tres estrategias generales de desarrollo de pruebas psicométricas para dirigir su construcción: los métodos racional, empírico y de validación de constructos.

La construcción racional de pruebas psicométricas se basa en una teoría de la personalidad o la psicopatología (p. ej., teoría cognitiva de la depresión) que guía la construcción de una prueba psicológica. El proceso de desarrollo de puntos y escalas se realiza para operativizar los rasgos fundamentales de una teoría. El inventario multiaxial clínico de Millon (MCMI, Millon Clinical Multiaxial Inventory) es un ejemplo de prueba desarrollada en su origen mediante un proceso racional de construcción de pruebas psicométricas principalmente.

La construcción empírica de pruebas, sin embargo, comienza con un gran número de ítems (llamado conjunto de ítems) y a continuación emplea varios métodos estadísticos para determinar cuáles distinguen grupos clínicos de individuos (proceso denominado afinamiento empírico). Los ítems que distinguen eficazmente a un grupo de otro se organizan para formar una escala sin tener en cuenta su contenido temático o «validez facial». El inventario de personalidad multifásico de Minnesota (MMPI, Minnesota Multiphasic Personality Inventory) es un ejemplo de prueba desarrollada con este método.

El método de validación de constructos combina aspectos de los métodos racional y empírico. En este marco conceptual se escribe un gran conjunto de ítems que reflejen un constructo teórico (p. ej., impulsividad); a continuación, se evalúan empíricamente esos ítems para determinar si realmente distinguen a los individuos que cabe esperar que se diferencien respecto a ese constructo (personas impulsivas frente a no impulsivas). Los ítems que distinguen eficazmente los grupos clínicos conocidos y cumplen otros criterios psicométricos (p. ej., tienen la consistencia interna apropiada) se mantienen para la escala. Además, si ítems teóricamente importantes no diferencian entre los grupos conocidos, este hallazgo puede dar lugar a una revisión de la teoría. La metodología de validación de constructos se considera la estrategia más sofisticada para el desarrollo de pruebas. El inventario de evaluación de la personalidad (PAI, Personality Assessment Inventory) es un ejemplo de prueba construida con el enfoque de validación de constructos.

## **Premisas psicométricas básicas: fiabilidad y validez**

Las pruebas psicológicas deben cumplir los estándares psicométricos mínimos de fiabilidad y validez para usarse de forma útil en la investigación y en la clínica. La fiabilidad consiste en la capacidad de repetición, estabilidad o constancia de la puntuación en la prueba de un individuo, y por lo general se representa mediante algún tipo de coeficiente de correlación (que oscila entre 0 y 1). Los instrumentos de investigación pueden tener índices de fiabilidad de tan solo el 0,7, mientras que los clínicos deben alcanzar el intervalo comprendido entre 0,8 y 0,9. Esto es así porque los instrumentos de investigación se interpretan conjuntamente como medidas de grupo, mientras que los clínicos se interpretan para una sola persona y, por tanto, requieren una precisión mayor. Existen varios parámetros estadísticos de fiabilidad para valorar una prueba: consistencia interna (grado en que los ítems de una prueba se comportan de la misma manera), fiabilidad prueba-nueva prueba (constancia de la puntuación en una prueba a lo largo de cierto tiempo, que típicamente oscila entre unos días y 1 año) y fiabilidad interobservador (según las escalas de valoración juzgadas por un observador). El estadístico  $\kappa$  se considera la mejor estimación de la fiabilidad interobservador, porque refleja el grado de consenso entre los observadores tras tener en cuenta la puntuación al azar. La fiabilidad (cuantía del error presente en una puntuación de la prueba) puede verse afectada por variabilidad en el individuo (la persona cambia con el tiempo), en el evaluador (error y sesgo del evaluador), o en la propia prueba (si se administra con instrucciones distintas).

La validez es un concepto más difícil de entender y demostrar que la fiabilidad. La validez de una prueba refleja el grado en que la prueba mide realmente el constructo que fue diseñada para medir (también conocida como validez de constructo). A menudo se demuestra comparando la prueba en cuestión con una medida (o medidas) ya establecida. Al igual que la fiabilidad, las medidas de validez suelen representarse como coeficientes de correlación (que oscilan entre 0 y 1). Típicamente, los coeficientes de validez se elevan al cuadrado (señalado como  $R^2$ ) para reflejar la cuantía de la varianza compartida entre dos o más escalas. Se necesitan múltiples tipos de datos antes de que una prueba pueda considerarse válida. La validez de contenido valora en qué grado un instrumento cubre todo el intervalo del constructo objetivo (p. ej., una prueba de depresión que no incluya ítems sobre alteraciones del sueño y del apetito tendría una validez de contenido limitada). La validez predictiva es la eficacia de una prueba para predecir ocurrencias futuras del constructo, mientras que la validez concurrente muestra cómo de bien se correlaciona con otras medidas existentes del mismo constructo. La validez convergente y la validez divergente hacen referencia a la capacidad de las escalas con distintos métodos (entrevista o autoinforme) de medir el mismo constructo (validez convergente), mientras que también tienen correlaciones bajas o negativas con escalas que miden rasgos no relacionados (validez divergente). En conjunto, las correlaciones convergentes y divergentes indican la especificidad con la cual la escala mide el constructo previsto. Es importante tener en cuenta que, a pesar de la cuantía de datos afirmativos para una prueba determinada, las pruebas psicológicas no se consideran válidas en sí mismas. Más bien, son las puntuaciones de pruebas que resultan válidas en situaciones específicas para tomar decisiones concretas.

## **Definición de prueba psicológica**

Existen múltiples técnicas para facilitar el establecimiento de un diagnóstico psiquiátrico y ayudar al plan de tratamiento, pero no cumplen necesariamente los requisitos de las pruebas psicológicas. Una prueba psicológica se define como una herramienta de medida compuesta por una serie de estímulos estándar (es decir, preguntas o estímulos visuales) administrados de forma estandarizada. Las respuestas a los estímulos se registran y se puntúan según un método estandarizado (lo que garantiza que una respuesta determinada siempre se puntúe del mismo modo), y se interpretan los resultados de la prueba del paciente respecto a una muestra normativa representativa.

## Grupos principales de pruebas psicológicas

### Pruebas de inteligencia

A Alfred Binet (1857-1911) se le atribuye la primera medida auténtica de la inteligencia. El French School Board encargó a Binet y a Theodore Simon que desarrollaran una prueba para identificar alumnos que podrían beneficiarse de programas de educación especial. Las escalas de Binet de 1905 y 1908 constituyen la base de las pruebas de inteligencia actuales. De hecho, la creación de la prueba de Binet en 1905 marcó el comienzo de las pruebas psicológicas modernas. Su enfoque era práctico y eficaz, ya que desarrolló un grupo de pruebas con la amplitud y la profundidad suficientes para separar a los niños con malos resultados y capacidad intelectual normal de aquellos cuyos malos resultados se debían a una capacidad intelectual menor. Además de tareas de matemáticas y lectura, Binet también introdujo otras áreas (como identificación de objetos, juicio y conocimiento social). Cerca de una década después, en la Stanford University, Lewis Terman tradujo la prueba de Binet al inglés, añadió más puntos y realizó algunas revisiones de la puntuación. La prueba de Terman sigue usándose hoy en día, con el nombre de escalas de inteligencia de Stanford-Binet.<sup>2</sup>

David Wechsler, para ayudar a evaluar a los reclutas durante la Primera Guerra Mundial, combinó lo que básicamente eran las tareas verbales de la Stanford-Binet con sus propias pruebas y creó la prueba de Wechsler-Bellevue (1939). A diferencia de la Stanford-Binet, la prueba de Wechsler-Bellevue generaba un cociente de inteligencia (CI) global, así como medidas de capacidades intelectuales verbales y no verbales. El uso de tres puntuaciones para describir el CI se popularizó entre los clínicos y las escalas de Wechsler fueron adoptadas de forma generalizada. Hasta el día de hoy siguen siendo la medida de la capacidad intelectual predominante en EE. UU.

La inteligencia es un constructo difícil de definir. Wechsler escribió que «la inteligencia, como constructo hipotético, es la suma o capacidad global del individuo de actuar con un objetivo, pensar racionalmente y enfrentarse eficazmente al entorno».<sup>3</sup> Esta definición ayuda a aclarar lo que las pruebas de CI modernas intentan medir (es decir, el funcionamiento adaptativo) y por qué las pruebas de inteligencia o CI pueden ser complementos importantes en la evaluación clínica y la planificación del tratamiento. Si una puntuación de CI refleja aspectos del funcionamiento eficaz, entonces las pruebas de CI miden aspectos de la capacidad adaptativa. La serie Wechsler de instrumentos para valorar el funcionamiento intelectual cubre la mayor parte del intervalo de edades humanas, y comienza con la escala de inteligencia de Wechsler de preescolar y primaria (Wechsler Preschool and Primary Scale of Intelligence) (4-6 años), seguida de la escala de inteligencia de Wechsler para niños IV (Wechsler Intelligence Scale for Children-IV) (6-16 años) y la escala de inteligencia de

Wechsler para adultos IV (Wechsler Adult Intelligence Scale-IV) (16-90 años).<sup>4,5</sup> Más recientemente, también se ha desarrollado la escala de inteligencia de Wechsler abreviada II (WASI-II, Wechsler Abbreviated Scale of Intelligence-II; Wechsler, 2011) con el fin de proporcionar medidas más breves (pero fiables) de inteligencia global, además de capacidad verbal y no verbal.<sup>6</sup>

Con el tiempo, la serie Wechsler ha pasado de proporcionar tres medidas globales de funcionamiento intelectual (CI total, CI verbal y CI manipulativo) a un modelo de funcionamiento cognitivo con más matices (comprensión verbal, razonamiento perceptivo, memoria operativa y velocidad de procesamiento). Al igual que sus predecesoras, las escalas de Wechsler actuales se puntúan para obtener una media de 100 y una desviación estándar (DE) de 15 en la población general, lo que permite comparar al paciente con una referencia normativa. Además, esta estrategia de puntuación también posibilita que los clínicos detecten discrepancias significativas entre el funcionamiento verbal y el no verbal: en la mayoría de los casos, una diferencia de unos 15 puntos (o 1 DE) puede considerarse estadística y clínicamente significativa. La [tabla 7-1](#) recoge un resumen de las categorías del CI.

Tabla 7-1

Intervalos de cociente intelectual (CI) con las puntuaciones de CI correspondientes y su distribución en percentiles

CI total	Grupos de inteligencia (CI)	Percentil en la distribución normal
≥ 130	Muy superior	2,2
120-129	Superior	6,7
110-119	Media-alta	16,1
90-109	Media	50
80-89	Media-baja	16,1
70-79	Límite	6,7
≤ 69	Extremadamente baja	2,2

Las cifras de CI no representan la inteligencia innata e inmodificable de un paciente, sino que es más exacto considerar el CI como representativo de la posición ordinal de un paciente, o rango en percentiles, en la prueba respecto a la muestra normativa en un momento determinado. En otras palabras, una puntuación en el percentil 50 está por encima del 50% de las personas en el intervalo de edad del paciente. Clínicamente, las cifras de CI pueden contemplarse como representantes del nivel actual de función adaptativa del paciente. Además, como las cifras de CI contienen cierto grado de error de medida y puntuación, deben señalarse con intervalos de confianza que indiquen el intervalo de puntuaciones en el que es probable que se encuentre el verdadero CI del paciente.

Las pruebas de CI de Wechsler están compuestas por 10-15 subpruebas diseñadas para medir dominios más concretos del funcionamiento cognitivo: comprensión verbal (verbal comprehension; VCI: semejanzas, vocabulario, información, comprensión), razonamiento perceptivo (perceptual reasoning; PRI: diseño de bloques, matrices, rompecabezas visuales, pesos de figuras, completar figuras), memoria operativa (working memory; WMI: prueba de dígitos, aritmética, secuencias de letras y números) y velocidad de procesamiento (processing

speed; PSI: búsqueda de símbolos, claves y cancelación). Las subpruebas se puntúan para llegar a una media de 10 y una desviación estándar de 3, lo que permite también aquí establecer distintas interpretaciones sobre el nivel de funcionamiento de una persona según la variabilidad de las puntuaciones. También es importante señalar que todas las puntuaciones de la escala de Wechsler están ajustadas según la edad.

Una de las estrategias iniciales para interpretar el rendimiento de un paciente en la WAIS es revisar la constancia de las puntuaciones. Por ejemplo, un CI de 105 está en el intervalo medio y por sí mismo no suscitaría ninguna «señal de alarma». Sin embargo, esta cifra puede darse en situaciones en las que todas las puntuaciones de índices compuestos están en el intervalo promedio (lo que refleja una variabilidad mínima) o en casos cuyas puntuaciones verbales son bastante altas y las no verbales, bastante bajas (indicativo de una variabilidad mayor en el funcionamiento). Las implicaciones clínicas en estos dos supuestos son bastante diferentes y llevarían a interpretaciones muy distintas; por este motivo, es esencial el examen de las discrepancias a la hora de interpretar el perfil. Sin embargo, la existencia de una discrepancia no siempre es sinónimo de anomalía. De hecho, no es infrecuente la aparición de discrepancias pequeñas o medianas incluso en la población general. Típicamente, se necesitan discrepancias de 12 a 15 puntos para poder considerarlas significativas, y deben reseñarse en el informe.

En resumen, aunque todas las medidas de inteligencia están muy interrelacionadas, lo mejor es considerar la inteligencia como un fenómeno polifacético. De acuerdo con la intención original de Binet, las pruebas de CI deberían usarse para valorar puntos fuertes y débiles de una persona respecto a una muestra normativa. Con demasiada frecuencia los profesionales de la salud mental se centran demasiado en el CI total y caen en el error proverbial de que los árboles no les dejan ver el bosque. Para contrarrestar este error es esencial conocer las subpruebas y los índices de la WAIS-IV, y así comprender la complejidad de una cifra de CI.

## **Pruebas de personalidad y psicopatología objetivas (autoinforme)**

La valoración moderna y objetiva de la personalidad (más correctamente denominada autoinforme) tiene sus raíces en la Primera Guerra Mundial, cuando las fuerzas armadas recurrieron a la psicología como ayuda para evaluar y clasificar a los nuevos reclutas. Se pidió a Robert Woodworth que desarrollara una prueba de autoinforme con el fin de evaluar la estabilidad emocional de los nuevos reclutas del ejército. Desafortunadamente, esta prueba, denominada hoja de datos personales (Personal Data Sheet), se terminó más tarde de lo previsto y apenas afectó al transcurso de la guerra. Sin embargo, la metodología usada por Woodworth influyó después en el desarrollo del instrumento de personalidad más usado, el MMPI.

Hathaway y McKinley (1943) publicaron la versión original del MMPI en la University of Minnesota.<sup>7</sup> (Aunque la versión original del MMPI se produjo en 1943, el manual oficial del MMPI no fue publicado hasta 1967.<sup>7</sup>) El objetivo de la prueba era ser capaz de diferenciar pacientes psiquiátricos de individuos normales, así como clasificar correctamente a los pacientes en su grupo diagnóstico. Se generó un gran conjunto de ítems, y se entrevistó a cientos de pacientes psiquiátricos, pidiéndoles que dieran su aprobación a cada uno de estos ítems. Se hizo lo mismo con una muestra de gran tamaño de personas que no estaban recibiendo tratamiento psiquiátrico. Los resultados de este proyecto mostraron que, aunque el conjunto de ítems resultaba excepcionalmente bueno para diferenciar a los grupos

normales de los clínicos, distinguir un grupo psiquiátrico de otro era más difícil. Un factor de confusión importante era que los pacientes con trastornos distintos tendían a suscribir los mismos ítems; esto llevaba a escalas con un alto grado de solapamiento entre sus ítems (es decir, los ítems aparecían en más de una escala). Este método de desarrollo de pruebas, conocido como afinamiento empírico (descrito anteriormente) fue innovador para su tiempo, porque la mayoría de las pruebas de personalidad precedentes se basaban exclusivamente en ítems que los creadores de la prueba hipotetizaban que medirían el constructo en cuestión (desarrollo racional de pruebas). La segunda innovación introducida con el MMPI fue el desarrollo de escalas de validez cuya intención era identificar el estilo de respuesta de los que realizaban la prueba. En respuesta a las críticas de que algunos ítems tenían un lenguaje pasado de moda y de que el grupo normativo original se consideraba una «muestra de conveniencia», el MMPI fue revisado en 1989. El MMPI-2 es el resultado de este proceso de revisión, y es la versión de la prueba usada hoy en día.<sup>8</sup>

## **Inventario multifásico de la personalidad de Minnesota 2**

El MMPI-2 es una prueba de función psicológica, tipo autoinforme, compuesta por 567 ítems de verdadero o falso.<sup>8</sup> Como ya se ha mencionado, el MMPI fue diseñado para separar a las personas en «normales» y «anómalas», y subdividir el grupo anómalo en clases específicas.<sup>9</sup> El MMPI-2 contiene 10 escalas clínicas que valoran las grandes categorías de la psicopatología y 6 escalas de validez diseñadas para evaluar actitudes en la realización de la prueba. Las puntuaciones brutas del MMPI se transforman en puntuaciones T estandarizadas cuya media es 50 y cuya DE es 10. Una puntuación T de 65 o más indica psicopatología clínicamente relevante en el MMPI-2. Una característica interesante del MMPI-2 es que se han desarrollado más de 300 escalas «nuevas» o experimentales para la prueba a lo largo de los años. Esto es posible por el método de afinamiento empírico descrito anteriormente. Los grupos de ítems que han demostrado diferenciar con fiabilidad dos o más muestras o poblaciones pueden añadirse al MMPI-2 como escala clínica o complementaria. La adición de estas escalas ayuda a refinar e individualizar la interpretación clínica de los resultados del MMPI-2.

Las escalas de validez del MMPI-2 son disimulación (lie; L), infrecuencia (F), corrección (K), inventario de respuesta variable (VRIN, Variable Response Inventory), inventario de respuesta verdadera (TRIN, True Response Inventory) y F inversa (F back; F[b]). La escala L fue diseñada para identificar respondedores que intentan minimizar alteraciones patológicas hasta el punto de que incluso niegan faltas menores que la mayoría de las personas admitiría. Habitualmente se considera un intento rudimentario de parecer más sano de lo que uno podría estar realmente (es decir, fingir que está bien). La escala F contiene puntos de trastornos inusuales o graves que la mayoría de las personas casi nunca señalarían como ciertos. Por este motivo, se cree que una escala F elevada es un «grito de ayuda» o bien un intento más voluntario de aparecer psicológicamente mal (es decir, fingir que se está mal). Al igual que la escala L, la K mide presuntamente la defensa psicológica, pero los datos han indicado que las personas con mayor nivel educativo tienden a puntuar más alto en la escala K que en la L.<sup>10</sup> Una puntuación elevada en la escala K (más defensa) significa que las escalas clínicas probablemente serán más bajas de lo que deberían, de modo que el MMPI-2 tiene una fórmula «corregida según la K» que proporciona una estimación de lo que podrían ser las escalas clínicas si la K se encontrara en los límites normales. Las correcciones K se

generaron racionalmente y no han sido verificadas empíricamente. Las tres escalas de validez siguientes, VRIN, TRIN y F(b) se añadieron en la revisión. La VRIN consiste en parejas de ítems que cabe esperar que se respondan del mismo modo (p. ej., «me siento triste casi todo el tiempo» y «considero que estoy deprimido»). Un número excesivo de parejas de ítems con respuestas incompatibles suscita la sospecha de invalidez de la prueba. La TRIN está compuesta de parejas de ítems que no cabe esperar que se respondan del mismo modo (p. ej., «me siento triste casi todo el tiempo» y «por lo general soy feliz»). Se considera que un número excesivo de parejas de ítems señaladas como verdaderas indica un sesgo de respuesta verdadera, que también puede conducir al cuestionamiento de los resultados de la prueba. Por último, la escala F(b) es la adición más reciente y se trata básicamente de la escala F para la segunda mitad de la prueba. Los ítems que componen la escala F original están todos en la primera mitad, y se observó que algunos pacientes tendían a esforzarse menos en responder con precisión a medida que avanzaban por los 567 ítems del MMPI. Una escala F(b) elevada indica que el individuo tiende a señalar como ciertos menos ítems en la segunda mitad de la prueba, lo que pone en duda la exactitud de los resultados.

Las escalas clínicas del MMPI-2 son las siguientes: 1) Hs (hypochondriasis): hipocondría; 2) D: depresión; 3) Hy (hysteria): histeria de conversión; 4) Pd (psychopathic deviate): desviación psicopática; 5) Mf: masculinidad-feminidad; 6) Pa: paranoia; 7) Pt: psicastenia; 8) Sc (schizophrenia): esquizofrenia; 9) Ma: hipomanía, y 0) Si (social introversion): introversión social. Las escalas se nombraron según los grupos diagnósticos que se estaban intentando identificar en el MMPI original (lo cual explica por qué siguen apareciendo términos tales como «psicastenia» en la prueba aunque ya no se use este término clínico). Para evitar la confusión, la mayoría de los profesionales que usan actualmente la prueba se refieren a las escalas por su número. El [cuadro 7-1](#) contiene una descripción breve de las conductas asociadas con las escalas clínicas del MMPI-2.

Cuadro 7-1 Descripciones de las conductas asociadas con elevaciones en las escalas del inventario multifásico de la personalidad de Minnesota 2 (MMPI-2)

#### Validez

(L) Disimulación:

Intento rudimentario de negar problemas psicológicos

(F) Infrecuencia:

Suscripción excesiva de síntomas infrecuentes; aparentar estar mal

(K) Defensa psicológica:

Intentos sofisticados o sutiles de negar problemas psicológicos

#### Escalas clínicas

(1) Hs–hipocondría (hypochondriasis): aparición excesiva de quejas somáticas vagas

(2) D–depresión: tristeza general y estado de ánimo deprimido con culpa y aislamiento

(3) Hy–histeria (hysteria) de conversión: ausencia de introspección, negación de problemas psicológicos, atención centrada en quejas físicas

(4) Pd–desviación psicopática (psychopathic deviate): rebeldía, hostilidad y conflictos con figuras de autoridad

- (5) Mf–masculinidad-feminidad: hombres pasivos, interesados por la estética y sensibles; mujeres no interesadas en el rol femenino tradicional
- (6) Pa–paranoia: desconfianza, hostilidad y sensibilidad a la crítica
- (7) Pt–psicastenia: preocupado, tenso e indeciso
- (8) Sc–esquizofrenia (schizophrenia): alienado, distante, escasa concentración y lógica
- (9) Ma–hipomanía: hiperactivo, emocionalmente lábil, pensamientos rápidos
- (0) Si–introversión social (social introversion): introvertido, tímido, falta de confianza social

Habitualmente se usa un programa informático para puntuar el MMPI-2. Se introducen y puntúan las respuestas brutas; a continuación, el programa produce un perfil de las puntuaciones en todas las escalas del MMPI-2. Una línea continua aparece dibujada a la altura de  $T = 65$  para proporcionar un punto de referencia sencillo. Las puntuaciones por encima de la línea son clínicamente significativas y las que se encuentran por debajo de ella no son clínicamente significativas (aunque se debe mencionar que las puntuaciones de escalas por debajo del punto de corte pueden suponer igualmente una información clínica importante y no suelen ignorarse). El programa de puntuación también proporciona una interpretación preprogramada de los resultados, al igual que una lista de ítems que podrían requerir una atención clínica más inmediata (p. ej., «recientemente he pensado en el suicidio»); estos también se conocen como «ítems críticos». Aunque el informe preparado, o «informe precocinado», del programa de puntuación puede ser un resumen informativo de los resultados, no se recomienda su uso en exclusiva porque no está adaptado a un paciente específico.

El MMPI-2 se interpreta comenzando por las escalas de validez. Las escalas situadas en el punto de corte de la significación clínica o por encima de este ( $T \geq 65$ ) podrían indicar la presencia de un estilo de respuesta motivado. Sin embargo, la elevación de una sola escala de validez no significa necesariamente que no sea posible interpretar las escalas clínicas. De hecho, un metaanálisis reciente indica que, de las escalas de validez, la diferencia en las puntuaciones brutas entre las escalas F y K lograba el mayor tamaño del efecto para identificar perfiles simulados y reales de MMPI-2.<sup>11</sup> Un perfil de MMPI-2 determinado como inválido significa en esencia que hay indicios suficientes que apuntan a que los ítems no se respondieron de forma constante o directa; esto pone en duda la validez de los restantes ítems. En los perfiles válidos, la primera fase de la interpretación consiste en identificar el tipo de código del perfil, determinando las dos o tres escalas más elevadas. Hay muchos manuales de interpretación del MMPI-2 que ayudan a entender las características de personalidad de los sujetos que producen elevaciones de escalas específicas o tipos de código. Por ejemplo, un código 2-4-7 indica la presencia de depresión (escala 2), ansiedad (escala 7) e impulsividad (escala 4) y la probabilidad de un trastorno de la personalidad.<sup>7</sup> Dado que es el instrumento de personalidad más investigado, lo que le falta al MMPI-2 en fundamentos teóricos lo ha ganado en la información adquirida del inmenso número de estudios realizados con esta herramienta. Buena parte de esta investigación también ha ayudado a crear subescalas para muchas de las escalas clínicas principales.

Dos productos fundamentales de este cuerpo bibliográfico son las subescalas de Harris-Lingoes y las escalas de contenido del MMPI-2.<sup>12-14</sup> Las escalas de Harris-Lingoes se basaron en un análisis del contenido de las escalas clínicas más heterogéneas con el fin de extraer los factores que tenían ítems muy similares. Por ejemplo, la escala 2 (depresión) tiene cinco escalas: depresión subjetiva, retraso psicomotor, mal funcionamiento físico, torpeza mental y melancolía. Aunque se mantienen en el MMPI-2 con ciertos cambios y están incluidas en el informe automatizado del MMPI-2, los datos relativos a las escalas de Harris-Lingoes son

algo escasos y obsoletos.<sup>8</sup> No obstante, el examen de estas escalas junto con el de todas las escalas clínicas puede potenciar la riqueza clínica de una interpretación del MMPI-2. Las escalas de contenido se sometieron a un proceso empírico más estricto y se usaron todos los ítems originales del MMPI, en vez de incluir ítems solo de ciertas escalas clínicas, como hicieron Harris y Lingoes.<sup>12,13</sup> Las escalas de contenido se crearon identificando en primer lugar las áreas relevantes medidas por el conjunto de ítems original. Tras un proceso repetitivo que eliminaba escalas e ítems con características psicométricas de mala calidad, se crearon 15 escalas (con un solapamiento mínimo de los ítems) que medían áreas clínicamente relevantes y no habían sido recogidas por ninguna escala clínica. La investigación acumulada sobre las escalas de contenido indica una relación más sólida con los diagnósticos del DSM-IV que la presente para algunas de las escalas clínicas originales.

Recientemente se ha intentado reformular las escalas clínicas del MMPI-2 para abordar el impacto indeseable del solapamiento de ítems y fortalecer al mismo tiempo la asociación de las escalas con el concepto diagnóstico original.<sup>14</sup> Serán necesarias más investigaciones para determinar si estas nuevas escalas, llamadas escalas MMPI-RC, logran los objetivos deseados.

### **Inventario clínico multiaxial de Millon III**

El MCMI-III es un cuestionario autoadministrado con 175 ítems de verdadero/falso diseñado para identificar trastornos de síntomas (trastornos del eje I) y trastornos de la personalidad (TP).<sup>15</sup> El MCMI se creó inicialmente como medida de la teoría integradora de Millon de psicopatología y personalidad.<sup>16</sup> Las revisiones de la prueba han reflejado los cambios en la teoría de Millon, así como las modificaciones de la nomenclatura diagnóstica. El MCMI-III está compuesto por 3 índices modificadores (escalas de validez), 10 escalas de personalidad básica, 3 escalas de personalidad grave, 6 escalas de síndromes clínicos y 3 escalas de síndromes clínicos graves. Una de las características exclusivas del MCMI-III es que intenta valorar simultáneamente la psicopatología de los ejes I y II. Las escalas del eje II se parecen, aunque no son idénticas, a los trastornos del eje II del DSM-5. Por su extensión relativamente corta (175 ítems en vez de los 567 del MMPI-2), el MCMI-III puede tener ventajas en la valoración de pacientes agitados, sujetos cuya resistencia está alterada o personas con una motivación subóptima por otros motivos. Una innovación del MCMI mantenida en el MCMI-III es el uso de puntuaciones de tasa base (TB) en vez de las puntuaciones T clásicas para interpretar las elevaciones en las escalas. Las puntuaciones de TB correspondientes a cada escala están determinadas para reflejar la prevalencia del trastorno en la muestra de estandarización. Los valores críticos de TB son 75 y 85. Una puntuación de TB de 75 en las escalas de personalidad indica rasgos problemáticos, mientras que en las escalas de síntomas señala la presencia probable del trastorno como problema secundario. Las puntuaciones de TB de 85 o superiores en las escalas de personalidad indican la presencia de un trastorno de la personalidad. Una elevación similar en las escalas de síntomas apunta a que el trastorno es prominente o primario.

### **Inventario de evaluación de la personalidad**

El PAI es una de las últimas pruebas psicológicas objetivas.<sup>17</sup> El PAI se desarrolló usando un marco conceptual de validación del constructo que daba importancia por igual a la

selección de ítems guiada por la teoría y a la función empírica de las escalas. El PAI contiene 344 ítems y un formato de respuesta de cuatro opciones (falso, ligeramente verdadero, bastante verdadero o completamente verdadero), que genera 22 escalas sin solapamiento de ítems. El PAI posee ciertas ventajas psicométricas respecto a otros instrumentos autoadministrados. A diferencia del MMPI-2 y el MCMI, la elevación en una escala del PAI no causará la elevación de una segunda escala por contagio simplemente porque esas escalas compartan ítems. Esta característica permite una interpretación más directa de cada escala. Las 22 escalas del PAI son 4 escalas de validez (inconsistencia [INC], infrecuencia [INF], impresión negativa [IMN], impresión positiva [IMP]), 11 escalas clínicas (quejas somáticas [SOM], ansiedad [ANS], trastornos relacionados con la ansiedad [TRA], depresión [DEP], manía [MAN], paranoia [PAR], esquizofrenia [ESQ], rasgos límites [LIM], rasgos antisociales [ANT], problemas con el alcohol [ALC], problemas con las drogas [DRG]), 5 escalas relacionadas con el tratamiento (agresión [AGR], ideas suicidas [SUI], estrés [EST], falta de apoyo social [FAS] y rechazo al tratamiento [RTR]) y 2 escalas interpersonales (dominancia [DOM] y afabilidad [AFA]). El PAI posee unas características psicométricas destacadas y es una prueba excelente para valorar a grandes rasgos múltiples dominios del funcionamiento psicológico relevante.<sup>18-20</sup> También están bien establecidas la validez y la utilidad clínica de esta prueba.<sup>21,22</sup>

Como se ha explicado en el MMPI-2 y el MCMI, el PAI también contiene escalas de validez creadas para ayudar a detectar estilos de respuesta anómalos. La primera escala de validez, la INC, es un conjunto de parejas de ítems que cabe esperar que se respondan del mismo modo (similar a la VRIN del MMPI-2). Típicamente, las elevaciones de esta escala indican confusión, problemas de lectura o incluso deterioro cognitivo.<sup>23</sup> La INF consiste en ítems que deberían responderse de un modo determinado. La mitad de estos tendrían que ser verdaderos (p. ej., «la mayoría de las personas prefieren ser felices»), y la mitad, falsos (p. ej., «realmente me encanta pagar impuestos»). La elevación de la INF indica que el respondedor estaba interpretando excesivamente los puntos, fue descuidado o tiene dificultades de lectura. Si la INC o la INF están por encima de su punto de corte respectivo, se invalidan los resultados de la prueba y no se recomienda seguir interpretando las escalas clínicas. La escala de IMN está diseñada para identificar respondedores que intentan presentarse con un tono globalmente negativo. Del mismo modo que la escala F del MMPI-2, los puntos de la IMN casi nunca los suscriben la mayoría de las personas (p. ej., «no he tenido ni un solo día feliz»). Otro uso de esta escala, no obstante, es identificar pacientes que tal vez desean que sus terapeutas conozcan cuánto malestar psicológico sienten (p. ej., un grito de ayuda). No es infrecuente que la IMN esté elevada en muestras clínicas. Cuando se evalúan pacientes psiquiátricos ingresados, las puntuaciones de la IMN pueden pasar incluso de 80, o 3 DE por encima de las normas comunitarias. La IMP pretende identificar respondedores que están intentando presentarse con un tono excesivamente positivo. La investigación ha demostrado que la escala IMP es muy sensible a los intentos de presentarse de una forma excesivamente positiva («nunca me siento mal»); las elevaciones, incluso discretas (puntuaciones  $T \geq 57$ ) pueden poner en duda la precisión del perfil.

El PAI también cuenta con un programa informático de puntuación; es similar en formato al programa del MMPI-2 e incluye un informe precocinado y una lista de ítems críticos. Al igual que en el MMPI-2, se produce un perfil que muestra la elevación de cada escala en una puntuación T (recuérdese que las puntuaciones T tienen una media de 50 y una DE de 10). Para la PAI, el punto de corte de la significación clínica está en  $T \geq 70$  y hay una línea continua dibujada en los puntos de corte de  $T = 50$  (puntuación media para la muestra

comunitaria) y  $T = 70$ . Otra ventaja del PAI es la inclusión de un grupo de referencia clínico, lo que permite establecer más comparaciones con grupos clínicos conocidos. Este se representa con una «línea del horizonte» marcada 2 DE por encima de la media para el grupo de comparación clínico.

La interpretación del PAI comienza con la valoración de las escalas de validez para observar si se encuentran por debajo de los puntos de corte. A continuación, se revisan las escalas clínicas en busca de elevaciones situadas por encima del punto de corte de la significación clínica. Con la excepción de ALC y DRG, las escalas contienen como mínimo tres subescalas, igualmente sin ítems solapados. El siguiente paso en la interpretación consiste en revisar las subescalas para determinar qué facetas del síndrome clínico están más respaldadas. Por ejemplo, las escalas de depresión y ansiedad del PAI tienen subescalas que se ocupan de las características cognitivas, afectivas y fisiológicas de estos trastornos. La revisión de las subescalas permite una evaluación más precisa de dónde cree el paciente que se encuentran sus dificultades principales. Las escalas relacionadas con el tratamiento (AGR, SUI, EST, FAS, RTR) miden dominios que pueden influir en la evolución y el tipo de tratamiento. Por ejemplo, los pacientes con una actitud fuerte y agresiva podrían tener menos capacidad de participar eficazmente en su propio tratamiento. La escala RTR es una estimación de la disposición del paciente a colaborar en el tratamiento, y se puntúa a la inversa, de modo que puntuaciones bajas en la escala RTR indican un deseo de ser ayudados y el interés por recibir ayuda. Las escalas interpersonales (DOM y AFA) se revisan porque aportan una estimación de cómo se ven los pacientes a sí mismos respecto a las interacciones en las relaciones. Estas escalas son bipolares, lo que significa que es posible interpretar las puntuaciones altas y las bajas. Una vez revisadas las escalas de PAI e identificadas las áreas problemáticas, los clínicos también tienen la posibilidad de consultar un índice de bondad del encaje generado por ordenador, que es una estimación estadística de la cercanía de un perfil determinado al perfil de pacientes de la muestra clínica que fueron diagnosticados minuciosamente de diagnósticos o problemas clínicos específicos (p. ej., intento de suicidio o fingir estar bien). Como regla general conservadora, los coeficientes de 0,9 o más indican que un perfil determinado tiene una relación sólida, o «buen encaje».

Por último, en los últimos años se han desarrollado varios índices complementarios para el PAI que, entre otras cosas, valoran factores como el potencial suicida y de violencia, y necesidad de aumentar el nivel de asistencia.<sup>24,25</sup> Estos índices también han demostrado que poseen propiedades psicométricas sólidas y validez de constructo, y ofrecen una información adicional potencialmente útil en la asistencia a pacientes.

## **Pruebas objetivas y DSM-5**

Los informes informatizados disponibles de las pruebas objetivas incluyen con frecuencia diagnósticos del DSM. Aunque se ha intentado validar las relaciones entre las pruebas de autoinforme y el DSM mediante la construcción de escalas o la investigación con numerosas poblaciones clínicas, estas pruebas no deben usarse en exclusiva para establecer un diagnóstico psiquiátrico. Esto no quiere decir que las pruebas de autoinforme revisadas en este capítulo no sean útiles para identificar áreas problemáticas, estilos interpersonales y consideraciones terapéuticas. En el mejor de los casos, los diagnósticos propuestos a partir de los resultados de las pruebas de autoinforme (y a este respecto también de las proyectivas) son indicaciones fundamentadas, y en el peor, trucos de mercadotecnia con escasa validez

clínica. Las pruebas psicológicas no están planteadas para sustituir a clínicos competentes capaces de integrar múltiples fuentes de datos en el proceso de evaluación. No obstante, estas pruebas sí proporcionan un gran conjunto de información estandarizada y pueden valorar la gravedad de los síntomas de un paciente de forma precisa.

## **Pruebas de personalidad basadas en la ejecución**

El desarrollo histórico de las pruebas proyectivas (actualmente denominadas pruebas de ejecución) está ligado al psicoanálisis y a la idea de motivación inconsciente. Carl Jung desarrolló la prueba de asociación de palabras (1910) para medir el concepto freudiano de «conflicto mental»; fue uno de los primeros intentos de estandarizar una técnica proyectiva. Según Jung, el tiempo que tardaba alguien en responder a una palabra revelaba lo cerca que estaba esa materia del complejo particular de la persona. Tiempos de respuesta mayores significaban que la persona se estaba defendiendo del afecto provocado por la palabra.

En 1921, Hermann Rorschach publicó su prueba de manchas de tinta. Curiosamente, la idea de usar manchas de tinta no fue exclusiva de Rorschach: Binet (1895) intentó sin éxito usar este método en sus primeras pruebas con niños en el sistema escolar francés para medir la imaginación visual.<sup>26</sup> En el caso de Rorschach, un juego llamado Klecksographie (Juego del coronel Blotto) era muy popular en Europa cuando él llegó al colegio. Este juego empleaba manchas de tinta para generar descripciones artísticas y coloridas destinadas al entretenimiento. Rorschach observó que los pacientes psicóticos producían respuestas muy diferentes a los demás en este juego. Tras una extensa exploración sistemática del potencial diagnóstico de hacer que los pacientes percibieran una serie de manchas de tinta estándar, publicó el libro *Psychodiagnostik*, con una tibia acogida. Rorschach falleció el año siguiente a la publicación del libro sin llegar a conocer el impacto que su prueba lograría finalmente en la valoración de la personalidad.

El segundo hito en la evaluación proyectiva se produjo con la publicación de la prueba de apercepción temática (TAT, Thematic Apperception Test).<sup>27</sup> A diferencia del Rorschach, que usa manchas de tinta ambigua, la TAT consiste en una serie de dibujos de personas de varios sexos y edades que realizan alguna actividad. Se pide al evaluado que cuente una historia sobre el dibujo que tenga una introducción, un nudo y un desenlace, y que describa lo que están pensando y sintiendo los personajes del dibujo. Los estímulos de la TAT fueron seleccionados de dibujos, ilustraciones de revistas y cuadros. Murray, miembro del profesorado de Harvard, se había sometido a psicoanálisis con Carl Jung y la teoría de este le influyó enormemente. A diferencia de la aceptación vacilante del Rorschach en ese mismo período, la TAT fue muy aceptada y usada por investigadores de muchas orientaciones teóricas distintas. Esto se debe en parte a indicios emergentes a favor de su validez de constructo y su utilidad clínica. La TAT ha suscitado un gran interés, y con los años se han desarrollado muchas otras pruebas de apercepción con fines clínicos y de investigación.

Las pruebas del funcionamiento psicológico basadas en la ejecución se diferencian sustancialmente de las objetivas. Estas pruebas son menos estructuradas y requieren un mayor esfuerzo por parte del paciente para comprender y responder a los estímulos de la prueba, y del evaluador, que debe realizar la búsqueda apropiada entre las vagas respuestas al mismo tiempo que registra literalmente todas ellas. Incluso las instrucciones de las pruebas proyectivas tienden a ser menos específicas que las de las pruebas objetivas. Como resultado, el paciente tiene más libertad para poner de manifiesto sus características de personalidad y

sus procesos de organización psicológica únicos. Mientras que las pruebas de autoinforme aportan una visión de las motivaciones explícitas «conscientes» del paciente (lo que desea que el examinador conozca), las proyectivas dejan ver las motivaciones implícitas del paciente, así como su estilo típico de percibir, organizar y responder a estímulos ambiguos externos e internos. Cuando se combinan, los datos de las pruebas objetivas y de las basadas en la ejecución pueden proporcionar una descripción multidimensional exhaustiva del funcionamiento de un paciente.

## **Método de las manchas de tinta de Rorschach**

El método de las manchas de tinta tiene como fin medir la totalidad del funcionamiento de una personalidad. Consiste en manchas de tinta sobre 10 tarjetas o láminas (las mismas manchas de tinta creadas en la publicación original de Rorschach siguen usándose hoy en día) ante las cuales se pide al paciente que diga lo que podría ser la mancha. De las manchas, cinco son blancas y negras; dos son negras, rojas y blancas, y tres tienen varios colores. El Rorschach tiene una fase de respuesta y otra de encuesta. En la fase de respuesta, se presentan al paciente las 10 manchas de tinta de una en una y se le pregunta «¿qué podría ser esto?». Las respuestas se registran literalmente y el evaluador intenta conseguir dos respuestas a cada una de las dos primeras tarjetas para ayudar al paciente a establecer un conjunto de respuesta. Tiene que haber como mínimo 14 respuestas totales sin rechazos de tarjetas (tarjeta sin respuestas, p. ej., «no veo nada en esta») para que la administración se considere válida.

En la segunda fase, el evaluador revisa las respuestas del paciente y pregunta en qué zona de la tarjeta observó la respuesta (lo que se conoce como «localización» en el lenguaje del Rorschach) y qué hizo que pareciera eso (lo que se conoce como determinantes). Es importante no dar respuestas adicionales en este momento, y así se les indica a los pacientes si refieren que pueden ver algo distinto al contemplar por segunda vez las tarjetas. El sistema comprensivo (SC) de Exner para puntuar el Rorschach es el sistema de puntuación más usado en EE. UU.<sup>26,28</sup> A continuación resumimos el proceso para puntuar una respuesta del Rorschach.

Por ejemplo, un paciente responde que la tarjeta I es «un murciélago volando» en la fase de respuesta. Durante la encuesta, primero se lee la respuesta al paciente:

Paciente: Aquí usé toda la tarjeta. [Hace un movimiento circular con la mano.]

Evaluador: ¿Qué hizo que le pareciera un murciélago?

Paciente: El color, el negro hizo que me pareciera un murciélago.

Entonces, el evaluador tiene la información suficiente para traducir la respuesta en una puntuación de Rorschach. Según el nivel de funcionamiento del paciente, su disposición a participar y la relación entre evaluador y paciente, la fase de encuesta puede prolongarse bastante. De media, administrar un Rorschach lleva en torno a 1-1,5 h.

En ocasiones, puntuar el Rorschach es complejo y dominarlo requiere una práctica considerable. Las respuestas se puntúan según tres principios básicos: dónde estaba localizada la respuesta en la mancha (localización), características de la mancha que ayudaron a generar la respuesta en el paciente y cuál fue la respuesta. La localización de una mancha ayuda al evaluador a determinar si lo que el paciente ve es frecuente y si la forma que ve el paciente sigue el contorno real de la mancha. Los determinantes son las características que el paciente usa para justificar lo que vio y pueden ser color, tonalidades, texturas o características que implican movimientos activos (p. ej., «parece como si esta hoja

se estuviera cayendo») o pasivos (p. ej., «es un hombre arrodillado rezando»). Cada respuesta se revisa respecto a la presencia de puntuaciones relacionadas con los tres principios. Una revisión detallada de todas las características individuales puntuables queda fuera del objetivo de este capítulo; no obstante, según el ejemplo señalado anteriormente, la respuesta al Rorschach se codificaría como sigue: Wo C'Fo A P 1.0. Este código indica que la persona usó toda la mancha y que lo que vio (el murciélago) tiene una forma percibida frecuentemente (Wo), las características de blanco y negro de la mancha fueron prominentes para formar la percepción (C'F), la percepción encajaba en la forma de la mancha (o) y se trata de una respuesta popular (P). Todas las respuestas de una prueba se traducen a estos códigos Rorschach, siguiendo las directrices del SC. A continuación, estos códigos se organizan en cocientes, porcentajes e índices y se presentan en lo que se conoce como sumario estructural. Las variables del sumario estructural se interpretan comparándolas con una muestra normativa.

La combinación de variables individuales del Rorschach en índices aumenta su fiabilidad y validez. Los índices del Rorschach son el índice de pensamiento perceptivo (IPP), la constelación de suicidio (CON-S), el índice de depresión (IDEP), el índice de déficit de afrontamiento (IDA), el índice de hipervigilancia (IHV) y el índice de estilo obsesivo (OBS). Los que han recibido la mayor atención empírica son el IPP, la CON-S y el IDEP. Aunque la validez del IDEP ha obtenido revisiones contradictorias, el IPP y la CON-S parecen mantenerse bien como medidas válidas de trastorno de pensamiento y para identificar personas con un riesgo elevado de suicidio, respectivamente.<sup>29,30</sup>

Durante muchos años la puntuación del Rorschach fue criticada por ser demasiado subjetiva. Sin embargo, con la creación por parte de Exner del sistema de puntuación comprensivo (Comprehensive Scoring System), la fiabilidad y la validez del Rorschach han mejorado considerablemente.<sup>26,28</sup> En parte debido a la reciente mala prensa del Rorschach, se han establecido estándares de fiabilidad elevados que deben cumplirse para publicar investigaciones con este método.<sup>31-33</sup> El mínimo estándar para el  $\kappa$  interobservador es de 0,8 o más en todas las variables del Rorschach descritas en los estudios de investigación publicados. Como comparación, los  $\kappa$  en este intervalo son iguales o mejores que los reflejados en muchos estudios que usan diagnósticos del DSM por entrevistas estructuradas. Múltiples estudios de distintos grupos de investigadores han podido alcanzar o superar este punto de corte, proporcionando así muchos datos sobre la fiabilidad de la puntuación de los datos del Rorschach.

Al igual que las demás pruebas psicológicas revisadas en este capítulo, la interpretación de un Rorschach es un proceso multifásico. La determinación de que un protocolo es válido constituye el primer paso, y supone revisar el número de respuestas (R) y un cociente conocido como  $\lambda$ . Este refleja la disposición del individuo a participar en el proceso del Rorschach. Las puntuaciones  $\lambda$  más altas se asocian con defensa psicológica del paciente o falta de implicación en el proceso. Los protocolos con menos de 14 respuestas y una puntuación  $\lambda$  inferior a 1 se consideran no válidos (un protocolo se considera válido si tiene al menos 14 respuestas y  $\lambda \geq 1$ ). El siguiente paso en la interpretación es una revisión de las puntuaciones de los índices, que son una combinación de muchas de las variables individuales. Por ejemplo, las puntuaciones en el índice CON-S aportan información sobre el grado de capacidad autodestructiva del paciente y se ha demostrado que predicen conductas suicidas. Así pues, los protocolos con una CON-S positiva identifican pacientes que tienen un riesgo elevado de autolesionarse de forma significativa. Una vez revisados los índices, la interpretación del Rorschach prosigue con la revisión de agrupaciones de variables

específicas, como organización afectiva, capacidad de control y tolerancia al estrés, mediación cognitiva, ideación, procesamiento de la información, percepción interpersonal, autopercepción y estrés situacional. Los protocolos válidos incluirán puntuaciones o características que componen todas estas agrupaciones, permitiendo que el evaluador valore el funcionamiento del paciente en cada una de esas áreas.

El Rorschach es asimismo un instrumento sobresaliente para valorar la calidad del pensamiento de un individuo y la presencia de trastornos del pensamiento. La valoración de la calidad del pensamiento incluye determinar la congruencia o «encaje» entre lo que vio el paciente y los contornos de la mancha (precisión perceptiva) y si es correcta la lógica usada para justificar la respuesta. La calidad de la forma (CF) refleja el encaje entre lo percibido y la mancha, y se valora en una escala de cuatro puntos (superior [+], ordinario [o], inusual [i] y menos [-]). En el extremo superior de esta escala (+, o y hasta cierto punto i), los pacientes usan características apropiadas de la mancha y lo que ven se corresponde por lo general con la forma de la mancha donde vieron la respuesta. Una mala calidad de forma (respuestas menos) indica que la respuesta del paciente no encaja en el contorno de la mancha y representa una percepción distorsionada o forzada. Por ejemplo, imaginemos que se entrega una mancha de tinta a un paciente y este responde así:

Paciente: Parece una araña enorme.

Evaluador (fase de encuesta): ¿Qué hace que parezca eso?

Paciente: Las patas que salen de la parte de arriba y de abajo, y los grandes ojos rojos saliendo de arriba.

Evaluador: ¿Qué hace que parezcan ojos?

Paciente: Tienen puntos negros.

En primer lugar, y sobre todo, la mancha apenas se parece a una araña si se mira distraídamente. Aunque puede haber ciertas protrusiones de negro en la parte superior izquierda y derecha, y, en menor grado, en la inferior, es difícil ver cómo una primera respuesta frecuente para la mayoría de las personas serían las patas de una araña. Si esta fuera una lámina real del Rorschach, la base de datos del SC tendría una colección de muchas respuestas frecuentes obtenidas de la muestra normativa que ayudaran a dirigir el juicio del clínico. Recuerde, las instrucciones son «¿qué podría ser esto?», y no «¿a qué le recuerda esto?». Los «ojos rojos» que describe el paciente son las áreas rojas en el centro de la parte superior con puntos en medio. Aunque estas podrían verse fácilmente como ojos, no se parecen en absoluto a los ojos de una araña. Una respuesta que usa más correctamente los contornos de la mancha podría ser la siguiente:

Paciente: La parte blanca del medio me parece una seta.

Evaluador (fase de encuesta): ¿Qué hace que parezca eso?

Paciente: Bueno... porque es blanca se parece a una seta, y aquí está el tallo, y se ensancha aquí, en la parte más alta.

Esta respuesta, aunque no usa toda la mancha, es bastante fácil de reconocer mirando el área blanca en el centro de la mancha. La CF para esta respuesta probablemente sería puntuada como ordinaria, «o», y la respuesta previa de una araña recibiría la consideración de menos, «-». La valoración de la calidad del pensamiento de una persona (asociaciones) se basa en la revisión sistemática de sus verbalizaciones en busca de un número de transgresiones de la lógica específicas (llamadas puntuaciones especiales), que incluye lo siguiente: características improbables (p. ej., «una mariposa con ojos en la parte de atrás de la cabeza»), razonamiento basado en la posición o el tamaño (p. ej., «hay una garra arriba, así que tiene que ser una langosta»), mezcla incorrecta o fundido de objetos (p. ej., «parece una araña-

seta») o dependencia excesiva del conocimiento subjetivo/personal para explicar una percepción (p. ej., «he visto muchas arañas y a mí me parece que esta es una»). La combinación de las medidas de precisión perceptiva y calidad de asociación proporciona una evaluación sólida del contacto con la realidad y el razonamiento formal.

## **Prueba de apercepción temática**

La TAT resulta útil para poner de manifiesto las motivaciones dominantes, las emociones y los conflictos personales centrales del paciente.<sup>27</sup> La TAT consiste en una serie de 20 láminas que muestran personas en distintas interacciones interpersonales creadas intencionadamente para ser ambiguas. La TAT se administra presentando 8-10 láminas, de una en una, con las siguientes instrucciones: «Invéntese una historia sobre este dibujo. Como todas las buenas historias, debe tener una introducción, un nudo y un desenlace. Dígame cómo se sienten las personas y qué están pensando». Al igual que el Rorschach, los evaluadores escriben cada respuesta al pie de la letra. Como las láminas de la TAT son mucho menos ambiguas que las del Rorschach, es muy infrecuente que una persona no sea capaz de generar algún tipo de respuesta, y solo es necesaria una respuesta por lámina. El tiempo de administración habitual para un conjunto de ocho láminas es de unos 30-45 min, según la capacidad verbal y la motivación del paciente. En la actualidad no hay ningún método de puntuación aceptado para la TAT, lo que la convierte en una «técnica clínica» más que en una prueba psicológica de verdad, según la definición recogida al inicio de este capítulo. A pesar de todo, las administraciones que incluyan al menos ocho láminas y consigan historias razonablemente detalladas pueden obtener una información fiable y clínicamente útil.

Se han desarrollado varios métodos de puntuación estandarizados para la TAT y todos se limitan a aspectos específicos del funcionamiento psicológico, como nivel de operaciones defensivas y grado de madurez psicológica. Un método en particular que ha recibido atención recientemente es la escala de cognición social y relaciones de objetos (SCORS, Social Cognition and Objects Relations Scale), desarrollada por Drew Westen para puntuar todos los tipos de datos narrativos.<sup>34-36</sup> La SCORS consiste en ocho variables puntuadas en una escala tipo Likert de siete puntos, en la que las puntuaciones más bajas indican niveles mayores de alteración. Las ocho variables de la SCORS son complejidad de la representación, calidad afectiva de la representación, inversión emocional en las relaciones, inversión emocional en valores y normas morales, comprensión de la causalidad social, experiencia y gestión de los impulsos agresivos, autoestima, e identidad y coherencia del yo. Cada respuesta de la TAT se puntúa según las variables de la SCORS y los datos normativos de diversas poblaciones clínicas y comunitarias, lo que permite comparar las puntuaciones medias. La interpretación de la SCORS se basa en la puntuación promedio de cada una de las ocho variables en todas las respuestas. Por ejemplo, una puntuación de 1 o 2 en la escala de complejidad de las representaciones indica un individuo que tiene poca capacidad de ver a las personas como seres integrados (con rasgos de personalidad deseables e indeseables). Las historias de la TAT de estos sujetos tienden a incluir personas que son «totalmente buenas» o «totalmente malas». Los psicólogos valoran típicamente las historias de la TAT en busca de temas emocionales, grado de integración emocional y cognitiva, estilo de relaciones interpersonales y visión del mundo (p. ej., se contempla como un lugar útil o dañino). Este tipo de datos pueden ser especialmente útiles para predecir la respuesta de un paciente a la psicoterapia y al psicoterapeuta. En muchos casos es posible abordar

directamente con el paciente la información de la TAT, porque muchos temas de sus historias pueden repetirse en la relación terapéutica.

## **Dibujos proyectivos**

En ocasiones, los psicólogos emplean dibujos proyectivos (dibujos a mano alzada de figuras humanas o de otros objetos; p. ej., casa-árbol-persona) como técnica de valoración complementaria. Al igual que la TAT, los dibujos proyectivos representan una técnica clínica más que una prueba formal, porque no hay métodos de puntuación formales. De hecho, la interpretación de estos dibujos a menudo se basa en gran medida (y estrictamente) en la teoría psicoanalítica. A pesar de la mala calidad de sus propiedades psicométricas, los dibujos proyectivos resultan a veces útiles clínicamente. Por ejemplo, las personas psicóticas pueden dibujar figuras humanas transparentes con los órganos internos visibles. Las técnicas de dibujo proyectivo se usan mucho más con los niños, especialmente cuando la comunicación verbal es limitada o inexistente. Es importante recordar que los dibujos proyectivos son menos fiables y válidos que las demás pruebas revisadas en este capítulo.

## **Proceso e informe de la consulta de evaluación**

### **Derivación a la consulta de evaluación**

Derivar a un paciente a una consulta de evaluación es similar a remitirlo a cualquier colega profesional. Las pruebas psicológicas no pueden hacerse «a ciegas». El psicólogo querrá tener información relevante sobre el caso y explorará a qué pregunta (o preguntas) debe responderse (esto se denomina pregunta de derivación) en la consulta. Este proceso puede ser más eficiente si el clínico que procede a la derivación ha considerado las áreas de funcionamiento o las cuestiones clínicas que le gustaría que evaluaran las pruebas. Sin embargo, en muchos casos una breve charla entre el clínico remitente y el consultor de la evaluación será suficiente para generar preguntas de derivación importantes. Preparar al paciente para las pruebas psicológicas puede resultar útil tanto al paciente como al evaluador, al revisar por qué es deseable la consulta e indicar que probablemente llevará varias horas. Una evaluación psicológica eficaz incluye la evaluación oportuna de un paciente y el aporte de información verbal al remitente (en los días siguientes a las pruebas). Una buena evaluación psicológica también proporciona información al paciente. A menos que quien deriva esté completamente familiarizado con las pruebas usadas en una evaluación, probablemente será mejor que el evaluador de la prueba sea quien aporte esta información. Las evaluaciones también deben incluir un informe escrito enviado al remitente y al paciente (si así lo solicita). Según la carga asistencial del evaluador y el objetivo de la evaluación, el informe escrito puede llevar cierto tiempo. No hay que dudar en contactar con el evaluador si no se ha recibido el informe en las 2 semanas siguientes a la evaluación. Para evitar la frustración de esperar un informe tardío, hay que plantear cuánto tardará en llegar el informe en el primer abordaje del caso.

### **Uso de la evaluación psicológica para mejorar la asistencia**

Hay que plantearse revisar los hallazgos relevantes con el paciente, aunque se haya dado información al clínico remitente y al paciente por separado. Esto ayuda a confirmar al paciente el valor que el clínico otorga a las pruebas y al tiempo que el paciente ha invertido en ellas. Muchas personas tienen reservas y temores infundados sobre las pruebas psicológicas. Los clínicos pueden asegurar a los pacientes que no existen pruebas de «locura» y que los resultados por sí mismos no forzarán el ingreso del paciente. Preguntar (p. ej., qué le parecieron las pruebas al paciente, qué le preocupa de los resultados, si los resultados coinciden con cómo se valora a sí mismo, y si ha aprendido algo gracias a las pruebas) puede ayudar realmente a fortalecer la relación terapéutica y demostrar el interés por el bienestar del paciente.<sup>37,38</sup> Una gran cantidad de trabajos teóricos y empíricos han explorado las similitudes entre el proceso de evaluación y la terapia. Esto puede emplearlo a su favor el clínico escuchando las historias que el paciente cuenta sobre sus experiencias en la evaluación y observando si pueden relacionarse con temas prominentes que hayan surgido durante el trabajo con él. En muchos casos, señalar estos temas ayuda al paciente a delimitar mejor cómo interacciona con el mundo. Si surgen cuestiones más específicas sobre las pruebas durante este abordaje, hay que plantearse consultar al evaluador para que las aclare. Por último, se deben considerar minuciosamente las recomendaciones del informe. En muchos casos, el evaluador ha valorado cuidadosamente lo que sería útil para el paciente y el terapeuta, o tiene experiencia previa evaluando a personas con dificultades similares a las del paciente y tal vez tenga acceso a recursos previamente desconocidos.

## **Comprensión del informe de evaluación**

El informe es una declaración escrita de los hallazgos del psicólogo. Debe ser comprensible, y recoger y responder sencillamente las preguntas de derivación. El informe ha de contener los siguientes datos: información relevante del contexto, lista de las técnicas usadas en la consulta, resumen sobre la validez de los resultados de pruebas y confianza que tiene el psicólogo en los hallazgos, descripción detallada del paciente de acuerdo con los datos de las pruebas y recomendaciones extraídas de los hallazgos de las pruebas. Los hallazgos de las pruebas deben presentarse de una forma lógica que proporcione una descripción integrada y rica del paciente (no una descripción de los resultados de pruebas individuales). También debería contener ciertos datos brutos (p. ej., cifras de CI) y una explicación de la medida de las puntuaciones (p. ej., las puntuaciones en los índices de la WAIS-IV tienen una media de 100 y una DE de 15). Esto permitirá que posibles pruebas de seguimiento sean comparadas significativamente con los hallazgos presentes. Debe terminar con una lista de recomendaciones. En un grado considerable, la calidad de un informe (y de la consulta) puede juzgarse por las recomendaciones facilitadas. Muchos clínicos no familiarizados con las pruebas psicológicas se sienten forzados a leer solo el resumen de los hallazgos de las pruebas y las recomendaciones. Un informe eficaz es aquel que debería escribirse para incluir una gran cantidad de información pertinente en toda su extensión. Al igual que los pacientes no son la suma de sus síntomas psiquiátricos, un resumen del informe de las pruebas solo es un elemento de un proceso de consulta muy complejo.

Acceda online a las preguntas de opción múltiple (en inglés) en <https://expertconsult.inkling.com>

# Bibliografía

1. Meyer G. Psychological testing and psychological assessment. *Am Psychol.* 2001;58:128–165.
2. Roid GH. Stanford Binet Intelligence Scales. ed 5 Interpretive manual: expanded guide to the interpretation of the SB5. Itasca: Riverside Publishing; 2003.
3. Matarazzo JD, Herman DO. Clinical uses of the WAIS-R: base rates of differences between VIQ and PIQ in the WAIS-R standardization sample. In: Wolman BB, ed. *Handbook of intelligence: theories, measurement, and applications.* New York: John Wiley & Sons; 1985.
4. Wechsler D. *Wechsler Intelligence Scale for Children—Fourth Edition: Administration and Scoring Manual.* San Antonio, TX: Pearson Publishers; 2003.
5. Wechsler D. *Wechsler Adult Intelligence Scale—Fourth Edition: Administration and Scoring Manual.* San Antonio, TX: Pearson Publishers; 2008.
6. Wechsler D. *Wechsler Abbreviated Scale of Intelligence—Second Edition: Manual.* Bloomington, MN: Pearson Publishers; 2011.
7. Hathaway SR, McKinley JC. *Minnesota Multiphasic Personality Inventory manual.* New York: Psychological Corporation; 1967.
8. Butcher JN, Dahlstrom WG, Graham JR, et al. *Manual for the administration and scoring of the MMPI-2.* Minneapolis: University of Minnesota Press; 1987.
9. Greene R. *The MMPI-2/MMPI: an interpretive manual.* Boston: Allyn & Bacon; 1991.
10. Graham JR. *MMPI-2: assessing personality and psychopathology.* New York: Oxford University Press; 1993.
11. Blanchard DD, McGrath RE, Pogge DL, et al. A comparison of the PAI and MMPI-2 as predictors of faking bad in college students. *J Pers Assess.* 2003;80:197–205.
12. Harris R, Lingoes J. *Subscales for the Minnesota Multiphasic Personality Inventory. Mimeographed materials.* Langley Porter Clinic; 1955.
13. Harris R, Lingoes J. *Subscales for the Minnesota Multiphasic Personality Inventory. Mimeographed materials.* Langley Porter Clinic; 1968.
14. Butcher JN, Graham JR, Williams CL, et al. *Development and use of the MMPI-2 content scales.* Minneapolis: University of Minnesota Press; 1990.

15. Millon T. *The Millon Clinical Multiaxial Inventory–III manual*. ed 2 Minneapolis: National Computer Systems; 1997.
16. Millon T, Meagher S, The Millon Clinical Multiaxial Inventory–III (MCMI-III). Hilsenroth MJ, Segal DL, eds. *Comprehensive handbook of psychological assessment*, vol 2. Hoboken, NJ: John Wiley & Sons; 2004: Personality assessment.
17. Morey LC. *The Personality Assessment Inventory professional manual*. Odessa: Psychological Assessment Resources; 1991.
18. Siefert CJ, Sinclair SJ, Kehl-Fie KA, et al. An item level psychometric analysis of the personality assessment inventory clinical scales. *Assessment*. 2009;16:373–383.
19. Sinclair SJ, Siefert CJ, Shorey H, et al. A psychometric evaluation of the personality assessment inventory—short form (PAI-SF) clinical scales in an inpatient psychiatric sample. *Psychiatry Res*. 2009;170:262–266.
20. Sinclair SJ, Antonius D, Shiva A, et al. The psychometric properties of the personality assessment inventory-short form (PAI-SF) in inpatient forensic and civil samples. *J Psychopathol Behav Assess*. 2010;40:406–415.
21. Slavin-Mulford J, Sinclair SJ, Blais MA. Empirical correlates of the Personality Assessment Inventory (PAI) in a clinical sample. *J Pers Assess*. 2012;94:593–600.
22. Slavin-Mulford J, Sinclair SJ, Stein MB, et al. External correlates of the Personality Assessment Inventory higher order structures. *J Pers Assess*. 2013;doi: 10.1080/00223891.2013.767820.
23. Blais MA, Baity MR. Exploring the psychometric properties and clinical utility of the Modified Mini-Mental State Examination (3MS) in a medical psychiatric sample. *Assessment*. 2005;12:1–7.
24. Sinclair SJ, Bello I, Nyer M, et al. The Suicide (SPI) and Violence Potential Indices (VPI) from the Personality Assessment Inventory: A preliminary exploration of validity in an outpatient psychiatric sample. *J Psychopathol Behav Assess*. 2012;34:423–431.
25. Sinclair SJ, Slavin-Mulford J, Antonius D, et al. Development and preliminary validation of the Level of Care Index (LOCI) from the personality assessment inventory in a psychiatric sample. *Psychol Assess*. 2013;doi: 10.1037/a0032085.
26. Exner JE. *The Rorschach: a comprehensive system*. ed 3 Basic foundations and principles of interpretation. New York: John Wiley & Sons; 1993.
27. Murray HA. *Thematic Apperception Test: manual*. Cambridge: Harvard University Press; 1943.

28. Exner JE. *The Rorschach: a comprehensive system. Basic foundations and principles of interpretation*. ed 4 New York: John Wiley & Sons; 2002.
29. Ganellen R. Weighing evidence for the Rorschach's validity: a response to Wood et al. (1999). *J Pers Assess*. 2001;77:1–15.
30. Wood JM, Lilienfeld SO, Nezworski MT, et al. Coming to grips with negative evidence for the Comprehensive System for the Rorschach: a comment on Gacono, Loving, and Bodholdt; Ganellan; and Bornstein. *J Pers Assess*. 2001;77:48–70.
31. Garb HN, Wood JM, Lilienfeld SO, et al. Roots of the Rorschach controversy. *Clin Psychol Rev*. 2005;25:97–118.
32. Wood JM, Nezworski MT, Stejskal WJ. The Comprehensive System for the Rorschach: a critical examination. *Psychol Sci*. 1996;7:3–10.
33. Wood JM, Nezworski MT, Lilienfeld SO, et al. *What's wrong with the Rorschach? Science confronts the controversial inkblot test*. San Francisco: Jossey-Bass; 2003.
34. Hilsenroth MJ, Stein M, Pinkster J. *Social Cognition and Object Relations Scale: Global rating method (SCORS-G)*. Garden City, NY: Derner Institute of Advanced Psychological Studies, Adelphi University; 2004: Unpublished manuscript.
35. Westen D. *Social Cognition and Object Relations Scale: Q-sort for Projective Series (SCORS-Q)*. Cambridge: Cambridge Hospital and Harvard Medical School; 1995: Unpublished manuscript.
36. Stein MB, Slavín-Mulford J, Sinclair SJ, et al. Exploring the construct validity of the Social Cognition and Object Relations Scale through clinical assessment. *J Pers Assess*. 2012;94:533–540.
37. Ackerman SJ, Hilsenroth MJ, Baity MR, et al. Interaction of therapeutic process and alliance during psychological assessment. *J Pers Assess*. 2000;75:82–109.
38. Finn SE, Tonsager ME. Information-gathering and therapeutic models of assessment: complementary paradigms. *Psychol Assess*. 1997;9:374–385.

|

(Stern, 20171024, pp. 72.e1-81.e1)

Stern, T. A. (20171024). *Massachusetts General Hospital. Tratado de Psiquiatría Clínica, 2nd Edition* [VitalSource Bookshelf version]. Retrieved from <https://bookshelf.vitalsource.com/books/9788491132301>

La cita indicada es una instrucción. Compruebe cada cita antes del uso para obtener una mayor precisión.