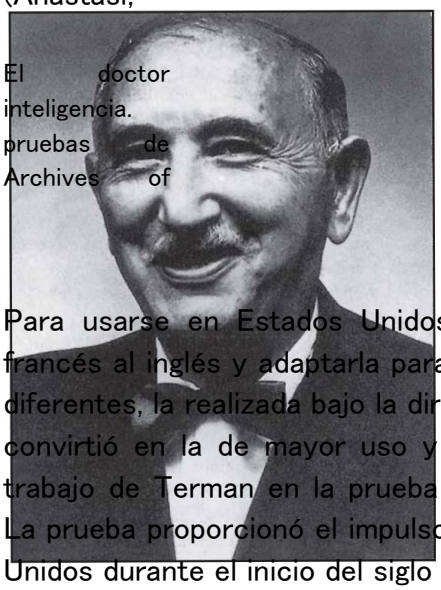
MEDICIÓN DE LA INTELIGENCIA

Antecedentes históricos

A finales del siglo XIX en Francia, Alfred Binet y sus colegas examinaron a niños de edad escolar para identificar a aquellos que probablemente encontrarían dificultades en la escuela, con el fin de proporcionarles las experiencias necesarias que les ayudarán a tener éxito. La teoría de la inteligencia de Binet incluía los elementos de dirección (saber qué tiene que hacerse y cómo debe cumplirse), adaptación (la selección y supervisión de la estrategia personal durante el curso del desempeño de la tarea), y crítica (la habilidad para criticar los propios pensamientos y acciones; Anastasi, 1988; Sattler, 1988). A partir de esos primeros intentos por medir y cuantificar el funcionamiento intelectual, el trabajo de Binet y Simon surgió como el estudio que ha dominado los métodos para examinar la inteligencia en la sociedad occidental hasta la actualidad. La definición y operacionalización de la inteligencia que domina el campo de la psicología todavía es un resultado directo de la petición que se hizo a Binet de que desarrollara un método para predecir el desempeño exitoso en la escuela (Sternberg, 1990). Si se hubiera pedido a Binet que predijera otro criterio (por ejemplo, desempeño en el trabajo, creatividad, habilidades artísticas), la definición actual y los métodos para la medición de la inteligencia serían muy diferentes.

Binet y sus colegas pasaron muchos años probando una variedad de métodos para medir la inteligencia, incluyendo la medición de los rasgos craneales y faciales, así como el análisis de la escritura. Sin embargo, los resultados llevaron a la convicción creciente de que la medición directa, incluso cruda, de las funciones intelectuales complejas era la más prometedora. Fue en relación a esto que Binet, en colaboración con Simon, preparó la primera escala de Binet-Simon (Binet y Simon, 1905). La versión de 1905 de su prueba consistía de 30 problemas o pruebas organizadas en orden de dificultad creciente. El nivel de dificultad se determinó aplicando las pruebas a 50 niños normales de tres a 11 años y a una pequeña muestra de niños y adultos con retardo mental. Las pruebas se diseñaron para cubrir una amplia variedad de funciones, destacando el juicio, la comprensión y el razonamiento (Sattler, 1988). En la segunda versión de la escala, desarrollada en 1908, se aumentó el número de pruebas y se les agrupó en niveles de edad con base en el desempeño de aproximadamente 300 niños normales entre las edades de tres y 13 años. La calificación del niño en la prueba entera podía expresarse, entonces, como un nivel mental correspondiente a la edad de los niños normales cuyo desempeño igualó (Anastasi, 1988).



El doctor David Wechsler fue un pionero en el campo de la examinación de la inteligencia. Desarrolló las escalas de inteligencia Wechsler para niños y adultos, las pruebas de inteligencia de mayor uso en la actualidad. (Fotografía cortesía de Archives of the History of American Psychology .The University of Akron.)

Para usarse en Estados Unidos, obviamente era necesario traducir la escala de Binet del francés al inglés y adaptarla para su uso en ese país. Aunque se desarrollaron varias versiones diferentes, la realizada bajo la dirección del psicólogo Lewis Terman en la Stanford University se convirtió en la de mayor uso y se conoce como la Stanford-Binet (Terman et al., 1916). El trabajo de Terman en la prueba incluyó vastas modificaciones, extensiones y estandarización. La prueba proporcionó el impulso para el movimiento masivo de examinación mental en Estados Unidos durante el inicio del siglo XX (Sattler, 1988).

En la actualidad, las pruebas de inteligencia de mayor uso en Estados Unidos son las desarrolladas originalmente por el psicólogo David Wechsler durante las décadas de 1940 y 1950. Basándose en las pruebas existentes en esa época –incluyendo la Stanford-Binet, las pruebas Alfa y Beta del ejército y la Escala de inteligencia de Bellevue Wechsler desarrolló primero una prueba individual de inteligencia para adultos, seguida de una prueba de estructura similar para niños de edad escolar hasta la edad de 16 años, y finalmente una prueba para niños de edad preescolar. En esas pruebas subyacía la creencia de Wechsler de que existe un nivel total o global de capacidad intelectual que puede medirse. De esta forma, esas pruebas arrojaban una calificación que representa la inteligencia total de la persona. Las pruebas desarrolladas por Wechsler también reflejan la metáfora geográfica de la inteligencia descrita por Sternberg (1990). Las pruebas de Wechsler, más que cualquier otra, han moldeado las percepciones que tienen los psicólogos de que el funcionamiento intelectual está compuesto por habilidades verbales y de desempeño (no verbales), las cuales están separadas pero relacionadas.

Wechsler (1939) insistió en que una prueba de CI mide la inteligencia funcional, no la inteligencia en sí misma. La inteligencia funcional se ve influida por factores no intelectuales, incluyendo la motivación, la configuración de habilidades específicas y el ajuste emocional. De acuerdo con Wechsler, una puntuación en una prueba CI es un reflejo de lo que se ha aprendido, lo cual es función de las oportunidades a las que se ha estado expuesto y de la habilidad que se tiene para obtener ventaja de esas oportunidades. Los subtests en las pruebas de Wechsler representan muestras de conducta, pero no son exhaustivas.

La necesidad de adaptación continua en la examinación de la inteligencia se representa en la historia de las escalas de Wechsler, porque las tres versiones han pasado por revisiones sustanciales desde su inicio e incluso después de la muerte de su creador. Esos cambios han tomado dos formas: cambios en los reactivos de las pruebas para actualizarlos y hacerlos más apropiados para las nuevas generaciones, y la examinación de nuevas muestras normativas para proporcionar fuentes actualizadas con las que puedan realizarse comparaciones normativas en la generación de puntuaciones.

PRINCIPIOS GENERALES DE LA EXAMINACIÓN DE LA INTELIGENCIA

Para algunos estudiantes de psicología clínica resulta una sorpresa que existan, literalmente, docenas de pruebas de inteligencia de entre las cuales elegir. ¿Cómo sabe un psicólogo cuál es la prueba correcta para un individuo específico? Primero, el psicólogo debe entender el modelo o metáfora de inteligencia que se refleja en la escala (Sternberg, 1992), lo cual asegurará que comprenda el significado de las respuestas y puntuaciones para este individuo. Por ejemplo, la Escala de inteligencia para niños de Wechsler, tercera edición (1992), se basa en la metáfora geográfica del funcionamiento mental. Otra escala de inteligencia para niños, la Batería de evaluación para niños de Kaufman (1984), se basa en la metáfora computacional que se interesa en las formas en que los individuos procesan la información. La interpretación de las puntuaciones de esas dos pruebas requiere un marco teórico muy diferente para entender la inteligencia.

Una de las metas al aplicar una prueba individual de inteligencia es comparar el desempeño del examinado con el desempeño de una muestra representativa de otras personas de la misma edad.

Para realizar tales comparaciones, es esencial que la prueba se aplique al individuo de la misma manera en que se administró a aquellos que fueron examinados como parte de la muestra de estandarización. Por lo tanto, las pruebas estandarizadas de inteligencia incluyen instrucciones

detalladas para la aplicación del instrumento. Esas instrucciones contienen un orden especificado para la administración de todos los reactivos y subtests, una formulación específica para todos los reactivos verbales, un arreglo particular para la presentación de todos los estímulos no verbales, y un conjunto de respuestas preparado de antemano para responder a las preguntas que pueda formular un examinado. Para el psicólogo clínico representa un desafío desarrollar un estilo de examinación que sea suficientemente cálido y atractivo para construir rapport con el examinado sin transgredir los requerimientos de la aplicación estandarizada.

Las desviaciones del procedimiento de aplicación pueden contribuir a cambios en el desempeño del individuo y resultar en una interpretación errónea de la calificación total. Por ejemplo, todos los reactivos y subtests se aplican en un orden estandarizado. Es probable que el desempeño en los reactivos iniciales de cualquier prueba se vea influido por la ansiedad del individuo y su poca familiaridad con la situación de prueba, mientras que los últimos reactivos de la prueba pueden ser afectados por la fatiga y, en algunos casos, por el aburrimiento con la tarea. En la medida en que los reactivos y subtests se apliquen en un orden estándar, los efectos de la ansiedad y de la fatiga se mantienen relativamente constantes entre los individuos. Sin embargo, si se cambia el orden es imposible explicar cómo pueden haber afectados esos factores el desempeño en los reactivos y subtests entre individuos.

Deben realizarse esfuerzos rigurosos para asegurar que, al igual que la aplicación de la prueba, la calificación e interpretación de las puntuaciones también se realicen de manera estandarizada. En disputa está el hecho de si diferentes examinadores pueden establecer niveles elevados de confiabilidad entre calificadores al calificar un conjunto de respuestas en una prueba de CI. En un esfuerzo por abordar este tema, todas las pruebas de CI de uso generalizado incluyen instrucciones detalladas para calificar las respuestas a cada reactivo aplicado. Es más fácil establecer un alto acuerdo sobre las respuestas a tareas que implican el uso de materiales físicos (por ejemplo, arreglar un conjunto de cubos de colores para igualar un patrón presentado) que lograr consenso sobre las respuestas a tareas verbales más complejas (por ejemplo, la definición exacta de una palabra del vocabulario).

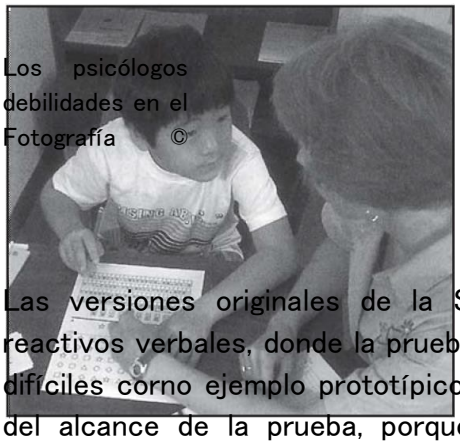
PRUEBAS DE INTELIGENCIA DE GRAN USO EN LA ACTUALIDAD

Ahora se revisarán en forma breve las características de cuatro pruebas de inteligencia que se utilizan actualmente en la evaluación clínica. Las pruebas difieren un poco en su base conceptual y en los grupos de edad en los que se utilizan. El uso clínico de las pruebas de inteligencia se presenta en referencia a dos de los estudios de caso, Jason y María (véanse cuadros 8.1 y 8.2 y la figura 8.2).

Stanford-Binet (4a. edición) El trabajo de Binet, Simon y Terman vive hoy en un sentido muy real en la cuarta edición de la prueba de inteligencia Stanford-Binet. La prueba ha pasado por muchos cambios en el transcurso de los años en un esfuerzo por asegurar que el instrumento refleje los avances modernos en las teorías de la inteligencia y de proporcionar datos normativos actuales para propósitos de calificación e interpretación.

La Stanford-Binet es una prueba de inteligencia de administración individual compuesta por una serie de tareas que reflejan el funcionamiento verbal y el no verbal. Cada tarea o subtest está formado por una serie de preguntas o problemas de nivel creciente de dificultad. No a todas las edades se le administran los mismos subtests; al aumentar la edad se agregan varias escalas mientras que otras se omiten. Dado que la meta original de esta prueba era distinguir a los niños con retardo mental de los no retardados, se centró la atención en crear una escala CI completa y no en generar puntuación es de subtest (aunque esas divisiones se han realizado

después). El primer paso es establecer un nivel basal (el nivel más alto en el cual todas las pruebas son aprobadas y ninguna prueba ha sido reprobada por debajo de ese nivel), y luego se continúa la examinación hasta que se establece un nivel de techo (el nivel más bajo en el cual todas las pruebas son falladas cuando ninguna prueba ha sido aprobada por encima de ese nivel).



Los psicólogos clínicos con frecuencia aplican pruebas para evaluar las fortalezas y debilidades en el funcionamiento intelectual de los niños. (Fotografía © Bob Daemmrich The Image Works.)

Las versiones originales de la Stanford-Binet contenían una fuerte carga de subtests y reactivos verbales, donde la prueba de vocabulario pedía definiciones de palabras cada vez más difíciles como ejemplo prototípico (Sattler, 1988). La cuarta edición representa una expansión del alcance de la prueba, porque ahora se incluyen muchos más reactivos y subtests no verbales. Los 15 subtests han sido organizados en cuatro grandes áreas de contenido. Razonamiento verbal (Vocabulario, Comprensión, Absurdos, Relaciones verbales), Razonamiento Visual/Abstracto (Análisis de patrones, Copia de matrices, Doblado y cortado de papel), Razonamiento cuantitativo (Cuantitativo, Series de números, Construcción de ecuaciones), y Memoria a corto plazo (Memoria de abalorios, Memoria para oraciones, Memoria para dígitos, Memoria para objetos).

La calificación de la cuarta edición de la Stanford-Binet se basa en una muestra de estandarización de más de 5 000 individuos entre las edades de dos y 23 años de 47 estados y la ciudad de Washington, D.C. en Estados Unidos. La muestra fue estratificada para igualar el censo de 1980 en Estados Unidos de acuerdo a la distribución de la población por región geográfica, sexo y origen étnico. De esta forma, el desempeño de los individuos en la versión actual de la Binet puede compararse con una muestra grande representativa de la población de Estados Unidos.

Las confiabilidades por consistencia interna para la escala completa va de .95 a .99 dependiendo de la edad. Las confiabilidades de los subtests varían de .80 a .97 para cada una de las cuatro áreas. En los subtests, la mayoría de las confiabilidades caen entre la parte superior de .80 y la parte inferior de .90. Esos datos indican que la Stanford-Binet es una prueba muy confiable. Como sucede con la validez de la mayoría de las pruebas de inteligencia, la validez de la Stanford-Binet se ha examinado de dos maneras, mediante correlaciones con otras pruebas de inteligencia y por correlaciones con pruebas de aprovechamiento escolar o con el desempeño/logro en el trabajo. Se ha demostrado que la Stanford-Binet correlaciona de manera moderada con otras medidas de inteligencia, con el logro y las calificaciones escolares, y con el logro de los adultos en el trabajo (Sattler, 1988).

La Stanford-Binet se utiliza sobre todo como un indicador del funcionamiento intelectual global. Las puntuaciones CI de la escala completa de los individuos se interpretan en términos del grado en que se desvían de las normas para la edad. Aunque la determinación de cualquier trastorno clínico o condición especial requiere más que la información en una sola medición, como una prueba de CI, la Stanford-Binet puede proporcionar información importante en la identificación del retardo mental o del talento intelectual. Como la Stanford-Binet requiere que el examinador aplique un gran número de reactivos para establecer los puntos más alto y más

bajo del funcionamiento de un niño, es más sensible que otras pruebas de inteligencia para discriminar entre puntuaciones extremas en los extremos alto y bajo de la distribución.

Escala de inteligencia para niños de Wechsler, tercera edición (WISC-III, Wechsler Intelligence Scale for Children, tercera edición) Wechsler y sus colegas desarrollaron dos pruebas para usarlas con niños: la Escala de inteligencia preescolar y primaria de Wechsler (WPPSI, Wechsler Preschool and Primary Scale of Children) para edades de 2 a 6 años, y la Escala de inteligencia para niños de Wechsler (WISC, Wechsler Intelligence Scale for Children) para edades de 6 a 16 años. La WPPSI ha pasado por una revisión hasta su versión actual, la WPPSI-R (1989), y la WISC-III (1991) representa la segunda revisión de esta prueba de gran uso. Este texto se centrará en el análisis de la WISC-III; en Sattler (1992) puede obtenerse mayor información sobre la WPPSI-R. La WISC-III es la tercera versión de la prueba para niños de edad escolar que fue originalmente diseñada en 1949 y previamente revisada en 1974.

La WISC-III es una prueba de administración individual que requiere aproximadamente de una a dos horas para aplicarse, dependiendo de la edad y nivel intelectual del niño. Está compuesta por 13 subtests que tocan diferentes aspectos del funcionamiento intelectual. En congruencia con la metáfora geográfica de la inteligencia que se refleja en el trabajo de Wechsler, esas pruebas se dividen en dos amplias categorías de subtests verbales y subtests de desempeño. Los subtests verbales evalúan el conocimiento básico que el niño ha adquirido acerca del mundo y que se basa en el procesamiento y almacenamiento de la información verbal. Por ejemplo, uno de esos subtests es una lista de palabras que se pide al niño que defina. Las pruebas de desempeño examinan las habilidades de solución de problemas del niño en tareas que se basan menos en el lenguaje y las habilidades verbales. Por ejemplo, un subtest de desempeño presenta al niño una serie de diseños geométricos y el niño intenta reconstruirlos usando cubos rojos y blancos.

Los subtests de la WISC-III también se han sometido a análisis factorial (el examen de las correlaciones entre los 13 subtests), y los resultados han proporcionado apoyo empírico parcial a la distinción original de Wechsler entre tareas verbales y tareas de desempeño (Sattler, l 992b). El análisis factorial de la WISC-III identificó los siguientes cuatro factores: comprensión verbal (información, semejanzas, vocabulario, comprensión), organización perceptual (figuras incompletas, ordenación de dibujos, diseño con cubos, composición de objetos), libertad de distracciones (aritmética, retención de dígitos) y velocidad de procesamiento (codificación, búsqueda de símbolos). De esta forma, existen dos factores (comprensión verbal, organización perceptual) que corresponden a la conceptualización original de Wechsler de habilidades verbales y de desempeño. Por otro lado, hay otros dos subgrupos de pruebas que no tienen una correlación elevada con esos dos grandes factores, y es útil considerarlos por separado cuando se interpreta la prueba. (En la figura 8.1 encontrará ejemplos de reactivos de la WISC-III.)

La confiabilidad, validez y muestra normativa de la WISC-III son ejemplares dentro del campo de las pruebas psicológicas. La muestra de estandarización de 1991 estuvo compuesta por 2200 niños y adolescentes, 200 de cada uno de 11 grupos de edad -de seis a 16 años basados en el censo estadounidense de 1988: 70 por ciento de angloamericanos, 15 por ciento de afroamericanos, 11 por ciento de hispanoamericanos y 4 por ciento de otros grupos . Los coeficientes de confiabilidad test-retest y de consistencia interna para las escalas verbal, de desempeño y global de la WISC-III son tan altas como las obtenidas por cualquier otra prueba psicológica (véase tabla 8.4). Sin embargo, la confiabilidad de los subtests individuales (por ejemplo, vocabulario, diseño con cubos) por lo general es más baja que la confiabilidad para las escalas más amplias, las cuales están compuestas por los subtests individuales. La menor

confiabilidad de los subtests individuales puede atribuirse directamente al hecho de que éstos contienen un número menor de reactivos que las puntuaciones de la escala más amplia (la confiabilidad aumenta como función del tamaño de la muestra de la conducta que se toma). En aquellos subtests que requieren que el examinador formule juicios al calificar (laberintos, vocabulario, semejanzas, comprensión), la confiabilidad entre calificadores en la calificación va de .90 a .94.

TABLA 8.4
Confiabilidad de la WISC-III

Escala o subtest	Confiabilidad por consistencia Escala o subtest interna	Confiabilidad test-retest
Escala completa de CI	.96	.92 a .95
CI verbal	.95	.90 a .94
CI de desempeño	.91	.87 a .92
Índice de comprensión verbal	.94	.89 a .93
Índice de organización perceptual	.90	.86 a .87
Índice de libertad de distracción	.87	.74 a .86
Índice de velocidad de procesamiento	.85	.80 a .85

Nota: Manual para la escala de inteligencia para niños de Wechsler, tercera edición (WISC-III) (Wechsler, 1991).

La WISC-III puede utilizarse para abordar una serie de cuestiones importantes acerca del funcionamiento intelectual de los niños individuales. Primero, y ante todo, la prueba puede usarse como un indicador del funcionamiento cognitivo actual de un niño en tareas relacionadas con la escuela con respecto a sus compañeros de la misma edad. De esta forma, puede proporcionar información útil para determinar si un niño está funcionando a un nivel que coincide con las expectativas para un niño de una determinada edad en la escuela. ¿Está el funcionamiento intelectual de este niño de nueve años a un nivel en el que pudiera esperarse que fuera capaz de tener éxito en las tareas que se asignan a los niños de cuarto grado? ¿O muestra este niño deficiencias en su funcionamiento actual a un grado que se necesitan servicios especiales para ayudarlo a tener éxito en la escuela? De esta forma, las puntuaciones en la WISC-III pueden usarse como una pieza de información (junto con la evaluación del funcionamiento adaptativo) para determinar si un niño cumple los criterios del retardo mental.
Además de proporcionar información acerca del funcionamiento intelectual global de un niño, la WISC-III también puede utilizarse para considerar variaciones en el funcionamiento de un niño en varios aspectos de la inteligencia. ¿Es este niño relativamente más competente en las tareas verbales o en las tareas basadas en el desempeño? ¿Muestra este niño debilidades específicas en tareas que requieren concentración o procesamiento rápido de la información? Para abordar esas preguntas, las puntuaciones en la WISC-III se interpretan en una serie de pasos, en los cuales los datos empíricos dirigen las comparaciones específicas entre las escalas y subtests. Específicamente, los psicólogos determinan si las puntuaciones de un niño en las escalas o subtests de la WISC-III difieren entre sí de una forma que sea estadísticamente significativa y rara. La significancia estadística de la diferencia entre las escalas o subtests establece si la diferencia es confiable (es decir, ¿esperaría ver una diferencia de este tamaño o mayor si aplicara la prueba de nuevo a este niño?). La información de tasa base se utiliza para determinar si una diferencia de esta magnitud entre puntuaciones es relativamente rara o frecuente en la población normativa. El uso cuidadoso de los datos de significancia y tasa base

proporciona una base empírica adecuada para la interpretación de las puntuaciones de la WISC-III y ayuda a los psicólogos a evitar los tipos de errores inferenciales que suelen cometer cuando interpretan datos de una prueba de manera muy subjetiva.

Considérese a una niña de 11 años de edad que obtiene un CI verbal de 127 y un CI de desempeño de 112. La diferencia de 15 puntos entre sus CI verbal y de desempeño parece ser considerable porque es equivalente a una desviación estándar completa en la WISC-III. Los datos generados en la estandarización de la prueba y proporcionados en el manual indican que esta diferencia es significativa, con una probabilidad menor a 5 por ciento de que ocurra por azar. No obstante, un análisis de la tasa base de esta diferencia en la muestra normativa indica que una discrepancia de este tamaño entre los CI verbal y de desempeño puede esperarse en más de 25 por ciento de los niños que presentaron la prueba. En contraste, considérese a un niño de ocho años que logra un CI verbal de 102 y un CI de desempeño de 73. La diferencia de 29 puntos en esas puntuaciones es confiable (una probabilidad menor a 1 por ciento de que ocurra por azar) y bastante rara, con menos de 3 por ciento de la muestra normativa con una diferencia de esta magnitud entre sus CI verbal y de desempeño. Este tipo de aproximación cuantitativa cuidadosa a la interpretación de las diferencias en las puntuaciones de la WISC-III es congruente con la aproximación estadística a la interpretación de la prueba (Dawes, Faust y Meehl, 1989). (Véase figura 8.2 donde se encontrará una presentación de los resultados de Jason en la WISCIII.)

Batería de evaluación para niños de Kaufman A finales de la década de 1970 y principios de la década de 1980, Allen y Nadine Kaufman, dos psicólogos que habían participado de manera activa en la investigación sobre la WISC y la Stanford Binet, desarrollaron una nueva medida de la función cognitiva. Los objetivos principales al desarrollar la Batería de evaluación para niños de Kaufman (K-ABC, Kaufman Assessment Battery for Children) fueron los siguientes: medir la inteligencia desde una base teórica y de investigación sólida, separar el conocimiento adquirido de la habilidad para resolver problemas no familiares, producir puntuaciones que se tradujeran en intervención educativa y ser sensible a las diversas necesidades de los niños preescolares, pertenecientes a minorías y excepcionales (Kaufman y Kaufman, 1984).

La K-ABC difiere de las escalas de Wechsler en que la primera se basa en una distinción entre dos estilos de procesamiento de la información, procesamiento secuencial y simultáneo, también denominado procesamiento secuencial y paralelo o procesamiento analítico y Gestalt/holístico (Kaufman, 1983). En el procesamiento secuencial de la información, cada reactivo de la prueba presenta un problema que debe ser resuelto organizando la información en un orden secuencial o serial. Cada línea se relaciona lineal o temporalmente con la precedente. Los ejemplos de tareas incluyen repetir números en el orden dicho por el examinador y reproducir una serie ordenada de movimientos manuales hechos por el examinador. Se supone que esas tareas se relacionan con una variedad de tareas cotidianas, de orientación escolar, como la memorización de números o el deletreo de palabras y el aprendizaje de asociaciones entre las letras y los sonidos que producen. Las tareas del procesamiento simultáneo de la información son de naturaleza espacial u organizacional. La información tiene que integrarse y sintetizarse de manera simultánea para producir la solución apropiada. Los ejemplos incluyen recordar la localización espacial de los estímulos, identificar el objeto en un dibujo parcialmente completo y construir un diseño abstracto con varios triángulos. Se propone que el procesamiento simultáneo se relaciona con tareas que son de naturaleza sobre todo perceptual, como aprender las formas de letras y números, así como funciones intelectuales de nivel superior porque involucran la capacidad para integrar información de diferentes fuentes.

La K-ABC también incluye por separado una escala de logro que pretende evaluar conocimiento factual y habilidades que, por lo general, se adquieren en la escuela. Las tareas que se evalúan en la escala de logro se basan principalmente en procesos verbales, muy similares a los reactivos en los subtests verbales de la WISC-III. La inclusión de escalas separadas de habilidades cognitivas y de logro pretendía proporcionar una comparación directa del aprendizaje potencial y la experiencia de aprendizaje. La K-ABC se centra en los procesos cognitivos, es decir, en cómo se procesa la información y no en qué tipo de información se procesa. Este énfasis contrasta con la dicotomía que subyace en las escalas de Wechsler, las cuales se basan en el contenido de los estímulos (contenido verbal contra contenido no verbal).
Al igual que la WISC-III, la K-ABC es fuerte en su confiabilidad y estandarización. La muestra normativa de estandarización constó de 2 000 niños de 2 a 12 años, distribuidos también por edad, sexo, ubicación geográfica y posición socioeconómica de los padres de acuerdo al censo de Estados Unidos. La consistencia interna de las escalas va de .96 para el procesamiento simultáneo a .93 para los subtests de logro. La validez de la K-ABC se ha evaluado de varias maneras. Kaufman y Kaufman (1984) reportan que los estudios analítico factoriales apoyan la validez de constructo del procesamiento secuencial-simultáneo y de las subescalas de logro.
La K-ABC refleja un modelo de la estructura de la inteligencia ligeramente diferente al reflejado por las escalas de Wechsler y la Stanford-Binet. Las puntuaciones generadas por la K-ABC representan competencias en el procesamiento simultáneo y secuencial de la información así como en el logro actual relacionado con la escuela. Esas puntuaciones pueden usarse para evaluar el modo preferido por el niño o el más fuerte de procesar la información, como manejar la información de una manera holística e integrada o hacerlo de una manera serial y lineal (Kaufman y Kaufman, 1984). Además, la K-ABC permite comparar entre esas habilidades de procesamiento de información y el logro actual. Por ejemplo, un niño que califica significativamente más alto en las escalas de procesamiento mental que en las escalas de logro sería evaluado como un niño de bajo rendimiento que puede estar experimentando alguna forma de problema o dificultad de aprendizaje (Kaufman y Kaufman, 1984).
Escala de inteligencia para adultos de Wechsler, tercera edición (WAIS-III) Se dispone de menos pruebas para medir la inteligencia en adultos, y la gran mayoría de los psicólogos utilizan la Escala de inteligencia para adultos de Wechsler, tercera edición (WAIS-III, Wechsler Adult Intelligence Scale, tercera edición), la tercera revisión de la escala de inteligencia para adultos de Wechsler. La versión original se publicó en 1939 como la Escala de inteligencia Wechsler-Bellevue, y en 1981 se hizo pública la WAIS-R. La WAIS-III refleja la misma estructura global que la WISC-III, con subtests organizados para generar CI verbal, de desempeño y de la escala completa. Aunque muchas de las tareas son similares a las de la WISC-III (por ejemplo, vocabulario y diseño con cubos), los reactivos son más complejos para distinguir entre adultos que varían en su funcionamiento intelectual. La prueba se aplica de manera individual, siguiendo un conjunto estandarizado de instrucciones para su aplicación.
La WAIS-III, la cual se revisó en 1997, se basó en una muestra normativa de 2 450 adultos que iban de 16 a 89 años y que fueron seleccionados para ser representativos de las características sociodemográficas de la población estadounidense, según lo reflejan las estadísticas del censo de 1995. Las confiabilidades para las puntuaciones de la escala WAIS-III son muy similares a las de la WISC-III. La consistencia interna está por encima de .90 para el CI de la escala completa, por encima de .90 para el CI verbal, y entre .80 y .90 para el CI de desempeño. La validez de la WAIS-III se ha establecido examinando las correlaciones de las puntuaciones en la prueba con otras medidas de inteligencia y de logro. El CI de la escala completa de la WAIS-III tiene una

elevada correlación con la puntuación de la escala completa de la Stanford-Binet, con los años de educación y con las puntuaciones en las pruebas de lectura, aritmética y ortografía (correlaciones de .60 a .76; Groth-Marnat, 1997).

La interpretación de la WAIS-III sigue los mismos procedimientos generales que la WISC-III. La diferencia principal en el uso de esas dos pruebas es el resultado de los diferentes tipos de preguntas de canalización que probablemente lleven al uso de pruebas de inteligencia en niños de edad escolar en oposición a los adultos. Mientras que en el caso de los niños las pruebas de inteligencia suelen utilizarse para arrojar luz en problemas relacionados con la escuela, el aprovechamiento escolar ya no es una pregunta para la mayoría de los adultos. En lugar de ello, la WAIS-III suele usarse con los adultos como parte de una batería de pruebas cuando existen preguntas concernientes a un deterioro neurológico o a una lesión cerebral como resultado de un trauma. (Véase cuadro 8.2 en el cual se presentan los resultados de María en la WAIS-111.)

EVALUACIÓN NEUROPSICOLÓGICA

Si bien las pruebas de inteligencia proporcionan información importante acerca del funcionamiento cognitivo y, con ello, reflejan ciertos aspectos del funcionamiento cerebral, no cuentan la historia completa. Se ha desarrollado una serie de pruebas y procedimientos para evaluar el funcionamiento neuropsicológico en un sentido más amplio. Benton (1994) define la evaluación neuropsicológica de la siguiente manera: "El propósito principal de la evaluación neuropsicológica es inferir las características estructurales y funcionales del cerebro de una persona evaluando su conducta en situaciones estímulo-respuesta definidas".

El término examinación neuropsicológica con frecuencia se malinterpreta para reflejar técnicas que se emplean para obtener una medición directa de las funciones subyacentes del cerebro (Sperry, 1961). Éste no es el caso. Más bien, las pruebas y procedimientos neuropsicológicos se utilizan para medir conductas observables que reflejan o son influidas por la estructura y función subyacentes del cerebro y el sistema nervioso central. Muy parecidas a las pruebas de CI, pues son representaciones funcionales del constructo subyacente (pero no observable) de inteligencia, las pruebas neuropsicológicas son indicaciones funcionales de los procesos subyacentes en el cerebro y el sistema nervioso central. Por lo general, se aplica una batería de pruebas para evaluar competencias y debilidades en tareas verbales y no verbales que pueden revelar patrones específicos de lesión o disfunción subyacentes (Lezak, 1995). Esas pruebas pueden proporcionar información sobre el lenguaje receptivo y expresivo, la memoria a corto y a largo plazo, el procesamiento de información visual y espacial, las funciones táctiles, motoras y cognitivas de orden superior.

La examinación neuropsicológica clínica, por lo general, se conduce para entender los efectos de trauma en la cabeza (lesiones en la cabeza ocurridas en accidentes), apoplejía, tumores cerebrales, enfermedades degenerativas (por ejemplo, enfermedad de Alzheimer, enfermedad de Parkinson), trastornos nutricionales y los efectos del abuso crónico del alcohol o las drogas. La evaluación neuropsicológica se basa en las relaciones entre las funciones de áreas específicas del cerebro y la conducta. Por ejemplo, en la mayoría de la gente la porción izquierda de la corteza frontal es responsable del lenguaje y la corteza frontal controla las habilidades viso espaciales. Por ello, si se observa una pérdida de lenguaje después de un accidente o una apoplejía, puede suponerse que el daño principal ha ocurrido en el hemisferio izquierdo del cerebro. La examinación neuropsicológica clínica ha sido descrita en términos de las tres L: detección de la lesión (identificación de las lesiones o daños en el cerebro), localización (identificar dónde se encuentra el daño en el cerebro), y lateralización (si el daño se encuentra en el lado izquierdo o derecho del cerebro).

Se ha creado una amplia gama de pruebas para medir el funcionamiento neuropsicológico. Éstas incluyen pruebas de inteligencia, lenguaje, atención, razonamiento abstracto, memoria a corto y a largo plazo y procesamiento viso espacial. Además de la aplicación de una prueba de inteligencia, como la forma apropiada para la edad de las escalas de Wechsler, la examinación neuropsicológica por lo general incluye baterías de pruebas (o parte de baterías) como la Batería Halstead-Reitan o la Batería neuropsicológica Luria-Nebraska). El desempeño de un individuo se compara entonces con una muestra normativa de individuos así como con muestras clínicas de individuos que se sabe padecían lesiones o trastornos cerebrales cuando presentaron la prueba. El objetivo del clínico es determinar si las respuestas del individuo en la prueba son similares a las de los individuos con una forma conocida de lesión o trastorno.

CUADR0 8.1

JASON: RESULTADOS INTERPRETACIÓN DE LA PRUEBA WISC-III

Como se recordará, Jason fue llevado a la clínica debido a los problemas de conducta que estaba presentando en la escuela y en el hogar y por preguntas acerca de trastorno por déficit de atención con hiperactividad (AqHD, Attention Deficit/Hyperactivity Disorder). Además, su desempeño académico en la escuela había declinado a pesar de que sus maestros sentían que era un niño brillante. Como parte del proceso de evaluación, el psicólogo pensó que podría ser importante obtener una medida de funcionamiento intelectual de Jason. Se aplica la SC-III para proporcionar un indicador del funcionamiento actual de Jason y para ofrecer datos adicionales que pudieran estar relacionados con la evaluación del ADHD.

Los resultados de Jasón en la WISC-III se presentan en la figura 8.1. En el momento del examen, Jason tenía ocho años y seis meses de edad. Por ello, su desempeño se comparó con una muestra de otros niños de la misma edad de la muestra normativa de comparación. Obtuvo un CI de escala completa de 95, un CI verbal de 90 y un CI de desempeño de 102. La diferencia de 12 puntos entre sus CI verbal y de desempeño fue estadísticamente significativa ($p < .05$), pero no fue extraordinariamente rara porque ocurrió en 36 por ciento de la muestra normativa. Los resultados sugieren que el funcionamiento intelectual global de Jason es promedio para su edad, pero es significativamente más fuerte en tareas no verbales que en tareas verbales.

También se examinaron 11 las puntuaciones dice porque pueden proporcionar información adicional relevante para la evaluación del ADHD. Su puntuación en comprensión verbal fue 95, su puntuación índice en organización perceptual fue 105, su puntuación índice en velocidad del procesamiento fue 88 y su puntuación índice en la escala libre de distracción (LD) fue '69 (en el segundo percentil de la muestra normativa).

Su puntuación LD fue significativamente más baja que sus otras tres puntuaciones índice, y cayó más de dos desviaciones estándar por debajo de la media normativa. Esta puntuación indica que se desempeñó muy mal en dos subtests que requerían que escuchara y repitiera series de números (hacia adelante y hacia atrás) y que resolviera mentalmente (es decir, sin escribir el problema o la solución) una serie de problemas aritméticos. Esas tareas requieren prestar atención a lo que el examinador está diciendo cuando se presenta nulos problemas y que la memoria a corto plazo retenga los problemas en la mente. La investigación ha demostrado que los niños con ADHD califican más bajo en el índice LD que los niños sin ADHD, lo que indica que el pobre desempeño en esas tareas puede estar asociado con un deterioro en la atención sostenida. Si Bien una baja puntuación en el factor LD de la WISC-III no es diagnóstico del ADHD, la baja calificación de Jason es congruente con las observaciones de que tiene problemas para mantener la atención, especialmente en tareas relacionadas con la escuela.

Un ejemplo de evaluación neuropsicológica se encuentra en pruebas de habilidades del lenguaje que se han diseñado para emplearse como medidas de disfunción cerebral en general y, en particular, para evaluar la función del hemisferio izquierdo o dominante del lenguaje (Benton,

1994). Un ejemplo es la Prueba Token (De Renzi y Vignolo, 1962; De Renzi y Faglioni, 1978), la cual se diseñó para evaluar la comprensión oral verbal incluso en individuos con habilidades verbales muy limitadas. La Prueba Token utiliza estímulos que sólo requieren que se conozcan los nombres de los colores y ciertas formas familiares, así como el significado de términos como tocar, grande y pequeño. El examinador registra la precisión con la que se ejecutan órdenes de complejidad variable (por ejemplo, "toque el círculo amarillo y el cuadrado rojo"). Los estudios de pacientes con daño en un lado del cerebro han demostrado que el desempeño deteriorado en esta prueba es mucho más frecuente entre pacientes con lesiones en el hemisferio izquierdo que en aquellos con lesiones en el hemisferio derecho. Un ejemplo de procesos de atención es la prueba de rastreo. Esta prueba requiere que el cliente dibuje líneas que conectan círculos consecutivamente numerados y es seguida por una segunda tarea en la cual el cliente dibuja líneas que conectan círculos alternativamente numerados y designados por letras. Las puntuaciones se basan en el tiempo total que se lleva para completar cada tarea. Esas tareas involucran una serie de habilidades además de la atención, incluyendo escudriñación compleja, coordinación motora, seguimiento viso motor y rapidez en el procesamiento de la información.

La comprensión sobre el funcionamiento del cerebro ha aumentado de manera considerable con desarrollos revolucionarios en la ciencia cognitiva y la neurociencia. El desarrollo de métodos para medir la actividad cerebral usando tomografía por emisión de positrones (PET, positron emission tomography) e imagenología por resonancia magnética funcional (fMRI, functional magnetic resonance imaging) ha sentado los cimientos para entender el funcionamiento intelectual a un nivel que nunca antes fue posible. Más que reemplazar las medidas funcionales de inteligencia actuales, los métodos para evaluar la actividad cerebral complementarán las técnicas actuales de medición. La tomografía por emisión produce una imagen de la distribución de un radionúclido previamente administrado en una sección deseada del cuerpo (en este caso el cerebro). Las PET utilizan las propiedades únicas de la aniquilación de la radiación que se genera cuando los positrones son absorbidos en la materia para proporcionar una imagen de un área seleccionada de tejido. La PET se ha usado en humanos para medir el flujo sanguíneo del cerebro, el volumen sanguíneo y el metabolismo de la glucosa y el oxígeno.

La metodología usual consiste en hacer que el individuo ejecute una serie de tareas que se supone involucran diferentes áreas del cerebro y usar la PET para monitorear la actividad del cerebro durante esas tareas. La fMRI mide el flujo relativo de sangre oxigenada a las áreas del cerebro que se activan en respuesta a una tarea o estímulo –las áreas activas del cerebro están enriquecidas en oxígeno en comparación con las áreas que no están activas–. Es probable que a medida que las tecnologías de PET e MRI mejoren se integren cada vez más con las medidas actuales de desempeño cognitivo para proporcionar una comprensión más completa de los componentes conductuales y biológicos del funcionamiento cognitivo.

Al mismo tiempo que avanza en la medición de la estructura y funcionamiento del cerebro, el campo de la examinación de la inteligencia está llegando a apreciar la relevancia de la cultura y del ambiente en la conducta inteligente. El futuro de la examinación de la inteligencia y de la evaluación neuropsicológica reflejará la continua necesidad de evaluar el funcionamiento cognitivo en un contexto ambiental y de considerar los aspectos dependientes de la cultura y los aspectos culturalmente invariantes de la inteligencia.