

TESTS DE INTELIGENCIA

Durante los inicios del siglo XX, una gran cantidad de aspirantes a psicólogos descubrieron que aplicando pruebas de inteligencia podían ganarse la vida en su profesión con algo distinto a la docencia y la investigación. Por ello, las pruebas de inteligencia en ocasiones han sido llamadas “el pan y la mantequilla de la psicología”. En la actualidad las *pruebas de Binet* ya no son la única ocupación de los especialistas en psicología aplicada, pero la evaluación de aptitudes cognitivas todavía forma parte de las actividades de los psicólogos en los ámbitos clínicos, educativos y empresariales.

HISTORIA, DEFINICIONES Y TEORÍAS

El término *inteligencia*, común ahora en el vocabulario de la mayoría de las personas, era casi desconocido en el habla cotidiana de hace un siglo. Durante la última parte del siglo XIX, muchos académicos y científicos fueron atraídos por la teoría de Charles Darwin de que las diferencias entre las especies evolucionaban mediante selección natural. Dos de estos estudiosos, el filósofo Herbert Spencer y Francis Galton, el científico caballero primo de Charles Darwin, se interesaron por las diferencias dentro de las especies en cuanto a características mentales y comportamiento. Ambos, junto con sus seguidores, sostenían que entre los seres humanos existe un grado innato de habilidad mental general, a la que se refirieron como *inteligencia*.

A diferencia de Spencer, Galton no se contentaba simplemente con especular y discutir sobre la naturaleza de la inteligencia. Intentando demostrar que la inteligencia tiene una base hereditaria, estudió árboles genealógicos y diseñó varias pruebas de discriminación sensorial y tiempo de reacción para medir sus componentes. Éstas y otras pruebas sensoriomotrices (velocidad de movimiento, fuerza muscular, sensibilidad al dolor, discriminación de peso y otras similares) fueron estudiadas ampliamente por el psicólogo estadounidense J. McKeen Cattell. Desafortunadamente, las pruebas resultaron relativamente inútiles para predecir el desempeño en tareas escolares y otras actividades que supuestamente requieren de inteligencia.

El enfoque del psicólogo francés Alfred Binet fue radicalmente distinto al procedimiento analítico de tratar de medir los componentes de la inteligencia. Binet sostenía que la inteligencia se manifiesta en el desempeño en diversas tareas y que podía medirse mediante respuestas a una muestra de dichas tareas. Debido a que el trabajo de Binet al diseñar las primeras pruebas de inteligencia con éxito fue motivado por el problema de identificar niños con retraso mental en el sistema escolar de París, es natural que la muestra de pruebas seleccionada por él estuviera plagada de tareas de tipo escolar.

En 1905 Binet y su socio, el doctor Théodore Simon, publicaron su primera serie de pruebas de inteligencia, 30 pruebas breves ordenadas desde la más sencilla hasta la más difícil. Al

proseguir su trabajo, publicaron en 1908 una escala modificada Binet-Simon que consistía en 58 tareas dispuestas por niveles de edad de 3 a 13 años. Las tareas se agruparon por edad cronológica de acuerdo con lo que había indicado la investigación que podrían realizar los niños normales de una edad determinada. La edad mental (MA [EM]) de un niño se establecía por la cantidad de subpruebas aprobadas en cada nivel, y una edad mental notablemente inferior a la edad mental del niño se consideraba indicativa de retraso mental. En 1911 se publicó una última versión modificada de la escala (tabla 7.1), pero después de la muerte prematura de Binet en ese mismo año, la escena de los posteriores desarrollos en cuanto a pruebas de inteligencia se mudó a Estados Unidos y Gran Bretaña.

Definición de la inteligencia

Desde que Binet y Simon produjeron las primeras pruebas prácticas de inteligencia, los psicólogos han intentado formular una definición viable del concepto. La explicación de Binet destacaba el juicio, el entendimiento y el razonamiento. Otras definiciones describían la inteligencia como la habilidad de pensar en forma abstracta, la habilidad de aprender o la habilidad de adaptarse al medio ambiente. Sin embargo, todas estas definiciones fueron criticadas por una u otra razón. La habilidad obviamente es necesaria para la sobrevivencia, pero resulta una definición de la inteligencia demasiado amplia. Por otra parte, la definición de inteligencia de Lewis Terman como la habilidad de tener pensamiento abstracto es demasiado estrecha. La habilidad para el pensamiento abstracto es un aspecto importante de la inteligencia, pero ciertamente no es el único. Por último, la concepción popular de inteligencia como la habilidad de aprender es inadecuada si se aceptan las pruebas de inteligencia como medida de ésta. Los aciertos en tales pruebas no están correlacionados en alto grado con el ritmo o la velocidad de aprender cosas nuevas, aunque sí están más relacionados con el nivel o la cantidad de aprendizaje.

Más que intentar formular una definición universalmente aceptable de la inteligencia, algunos psicólogos han sugerido que podría ser mejor abandonar el término por completo. Si se requiere un término alternativo, tal vez sería preferible utilizar *habilidad mental general*, o *habilidad académica*. Los dos últimos términos son un reconocimiento al hecho de que las pruebas de inteligencia tradicional son sobre todo predictores del éxito en el trabajo escolar. Sin importar lo intensa que pueda ser la oposición al término *inteligencia*, es ciertamente menos fuerte que la oposición al coeficiente intelectual (CI). Debido a la controversia existente sobre el CI y a la implicación de que es una medida fija de habilidad cognoscitiva, ciertos psicólogos que han dedicado gran parte de sus vidas profesionales al estudio de la inteligencia han expresado una disposición a abandonar por completo el término CI (Vernon, 1979).

No todos los instrumentos examinados en este capítulo tienen la etiqueta específica de *prueba de inteligencia*; más bien se han propuesto como medidas de habilidad mental *general*. En este sentido, deben distinguirse de las medidas de *habilidades especiales* consideradas en el capítulo 10. Sin embargo, no está clara la distinción entre pruebas de habilidad mental general (inteligencia) y pruebas de habilidades especiales, y ciertas pruebas de habilidad académica analizadas en este capítulo podrían corresponder igualmente bien al capítulo 10.

Teorías de la inteligencia

Las teorías de la inteligencia, o más bien del comportamiento inteligente, se han basado en modelos psicométricos de desarrollo y procesamiento de información. Los primeros dos tipos de teorías son enfoques tradicionales, la tercera teoría es de origen más reciente.

TABLA 7.1 Las cincuenta y cuatro subpruebas de la Escala de Inteligencia Binet-Simon de 1911**3 años de edad**

Señala sus ojos, nariz y manos.
 Repite dos dígitos.
 Enumera objetos de una imagen.
 Dice su apellido.
 Repite una oración de seis sílabas.

4 años de edad

Dice su sexo.
 Nombra llave, cuchillo, dinero.
 Repite tres dígitos.
 Compara dos líneas.

5 años de edad

Compara dos pesos.
 Copia un cuadrado.
 Repite una oración de diez sílabas.
 Cuenta cuatro centavos.
 Une las mitades de un rectángulo dividido.

6 años de edad

Distingue entre mañana y tarde.
 Define palabras familiares en términos de uso.
 Copia un rombo.
 Cuenta 13 monedas.
 Distingue dibujos de rostros feos y hermosos.

7 años de edad

Muestra su mano derecha y su oreja izquierda.
 Describe un dibujo.
 Ejecuta tres órdenes dadas simultáneamente.
 Cuenta el valor de seis centavos, tres de los cuales son dobles.
 Nombra cuatro colores principales.

8 años de edad

Compara dos objetos de memoria.
 Cuenta de 20 a cero.
 Señala omisiones en dibujos.
 Da el día y la fecha.
 Repite cinco dígitos.

9 años de edad

Da cambio de 20 centavos.
 Define palabras familiares en términos superiores al uso.
 Reconoce todas las monedas (nueve).
 Nombra los meses del año en orden.
 Contesta o comprende "preguntas fáciles".

10 años de edad

Ordena cinco bloques por peso.
 Copia dos dibujos de memoria.
 Critica afirmaciones absurdas.
 Contesta o comprende "preguntas difíciles".
 Usa tres palabras dadas en no más de dos enunciados.

12 años de edad

Resiste sugerencias sobre la extensión de líneas.
 Compone una oración con tres palabras dadas.
 Menciona 60 palabras en 3 minutos.
 Define tres palabras abstractas.
 Descubre el sentido de una oración desordenada.

15 años de edad

Repite siete dígitos.
 Encuentra tres rimas para una palabra dada en un minuto.
 Repite una oración de 26 sílabas.
 Interpreta imágenes.
 Interpreta hechos dados.

Adulto

Resuelve el test del papel cortado.
 Reacomoda un triángulo en la imaginación.
 Menciona diferencias entre pares de términos abstractos.
 Da tres diferencias entre un presidente y un rey.
 Encuentra la idea principal en un párrafo que ha leído.

Teorías psicométricas. El método psicométrico, que ha dado origen a muchas pruebas de inteligencia y diversos métodos estadísticos para analizar las calificaciones de estas pruebas, se centra en las diferencias individuales en cuanto a habilidades cognoscitivas y en la búsqueda de las causas de estas diferencias. Entre las teorías o modelos de habilidades cognoscitivas basadas en el método psicométrico y originadas sobre todo de los resultados del análisis factorial (vea apéndice A), figuran la teoría bifactorial de Spearman (1927) (que consiste en un factor general

más varios factores específicos para cada prueba), la teoría multifactorial de siete habilidades mentales básicas de Thurstone (Ekstrom, French y Harman, 1979), el modelo de estructura del intelecto de Guilford (1985) y el modelo jerárquico de Vernon (1960). El modelo de Vernon consiste en un factor general en el primer nivel, factores verbales-educacionales y práctico-mecánico-espaciales en el segundo nivel, y varios factores de un grupo menor en un tercer nivel (vea la figura 7.1). La teoría de Cattell (1963) de dos tipos de inteligencia, fluida y cristalizada, también se basa en los resultados del análisis factorial y se relaciona con la distinción de Hebb (1949) entre Inteligencia A e Inteligencia B.

Teorías sobre el desarrollo. Las teorías sobre el desarrollo de las habilidades cognoscitivas que provienen de la investigación sobre psicología del desarrollo humano, subrayan la uniformidad o las similitudes interindividuales en la evolución cognoscitiva más que las diferencias individuales. Un ejemplo primordial es la idea de Piaget de que la cognición se desarrolla a partir de las acciones de asimilación y acomodamiento en el mundo exterior. La *asimilación* consiste en ajustar las nuevas experiencias en las estructuras cognoscitivas preexistentes (esquemas *schemata*); el *acomodamiento* es la modificación de estos *schemata* como resultado de la experiencia. Al interactuar con el ambiente, un niño en crecimiento crea *schemata* de modo que funcionen como mapas explicativos y guías para el comportamiento. De acuerdo con Piaget, por lo regular los niños se desarrollan intelectualmente a través de una serie de etapas progresivas: sensoriomotriz (del nacimiento a 2 años de edad), preoperativa (de 2 a 7 años de edad), operativa concreta (de 7 a 11 años de edad) y operativa formal (de 11 a 15 años de edad). Piaget pensaba que el aumento de la inteligencia se detenía a la edad aproximada de 15 años, pero varios investigadores han objetado esta afirmación.

Teorías sobre el procesamiento de información. Las teorías sobre procesamiento de información, o modelos de resolución de problemas y razonamiento, se ocupan de identificar los procesos cognoscitivos u operacionales mediante los cuales el cerebro maneja la información. La investigación sobre atención y velocidad de procesamiento ha recibido un énfasis particular desde una perspectiva de procesamiento de información. Resultan ilustrativas de las teorías de procesamiento de información las teorías triárquicas o de proceso componencial de Sternberg

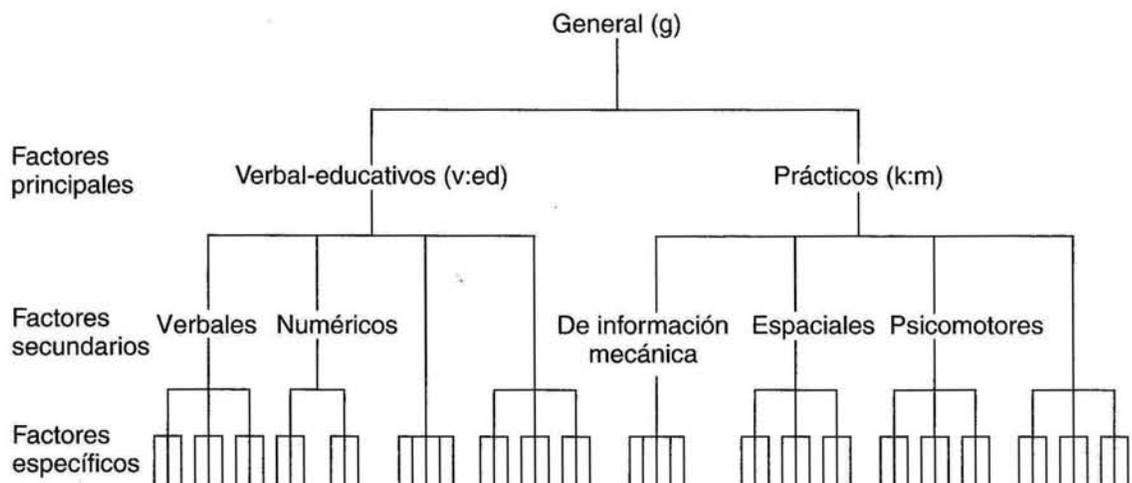


FIGURA 7.1 Modelo jerárquico de Vernon de las habilidades intelectuales.

(De acuerdo con Vernon, 1960, p. 22. Reproducida con autorización de la editorial Routledge.)

(1982), la teoría de inteligencias múltiples de Gardner (1983), y el modelo PASS de Das, Naglieri y Kirby (1994).

En un principio, Sternberg (1982) formuló la hipótesis de que existen cinco clases de *procesos componenciales* mediante los cuales el cerebro opera sobre la información y resuelve problemas, a saber: metacomponentes, componentes de desempeño, componentes de adquisición, componentes de retención y componentes de transferencia. Entre los diversos componentes de estas cinco clases, la codificación y la comparación son especialmente críticas para lograr una resolución efectiva de los problemas. En una extensión de su teoría de procesos componenciales, Sternberg (1985, 1986) propuso una *teoría triárquica* que incluye tres subteorías: componencial, experiencial y contextual. La subteoría componencial consiste en metacomponentes, componentes de desempeño y componentes de adquisición de conocimiento. La subteoría experiencial se ocupa de la de formular nuevas ideas combinando factores o información aparentemente no relacionados. La subteoría contextual aborda la de adaptarse a condiciones ambientales cambiantes y conformar el entorno de tal modo que nuestras ventajas se incrementen y nuestras desventajas se compensen. En una modificación posterior de su teoría, Sternberg (1988) propuso el concepto de *autocontrol mental*, que representa un intento por combinar el concepto de inteligencia con el de personalidad. Las maneras en que los tres tipos de inteligencia delineados por la teoría triárquica —componencial, experiencial y contextual— se ponen en práctica en la resolución de los problemas cotidianos, se caracterizan como *estilos intelectuales*. La efectividad de un estilo intelectual en particular depende de la medida en que se ajuste a la capacidad intelectual de la persona, su estilo preferido y el problema inmediato por resolver.

De acuerdo con la teoría de Gardner de inteligencias múltiples, la cognición y el procesamiento de información en los humanos implica el despliegue de varios sistemas simbólicos que son formas características de percepción, memoria y aprendizaje. Gardner propuso que hay siete formas de inteligencia: lingüística, lógico-matemática, espacial, musical, kinestésica corporal, y dos formas de inteligencia personal (intrapersonal e interpersonal). Sostuvo que sólo las primeras tres formas se miden mediante tests de inteligencia convencionales, y que la cultura occidental ha puesto demasiado énfasis en la primera de éstas, la lingüística. Sin embargo, Gardner advierte que las otras dos formas de inteligencia (lógico-matemática y espacial) son más valiosas en muchas sociedades y circunstancias.

El modelo de inteligencia PASS (planeación, atención, procesamiento simultáneo, procesamiento sucesivo) se basa en la teoría de Aleksandr Luria de que el cerebro humano está funcionalmente dividido en tres unidades. La *primera unidad funcional*, que se asocia con el tallo cerebral superior y el sistema límbico, es responsable de la estimulación y la atención. La *segunda unidad funcional* está asociada con las regiones posteriores de los hemisferios cerebrales, incluyendo las áreas visual (occipital), auditiva (temporal) y sensorial general (parietal); es responsable de la recepción, el análisis y almacenamiento de la información mediante procesos de razonamiento simultáneos y sucesivos. La *tercera unidad funcional* está asociada con las partes anteriores del hemisferio cerebral, en particular con la región prefrontal; es responsable de planear, regular y verificar la actividad cognoscitiva. Para efectuar el procesamiento cognoscitivo de información, la base de conocimiento del individuo debe estar integrada con los procesos de *planeación* (tercera unidad funcional), *atención* (primera unidad funcional), y procesos *simultáneos* y *sucesivos* (segunda unidad funcional) como lo requiere una tarea en particular. El resultado de semejante proceso cognoscitivo incluye hablar, escribir u otras actividades motoras (Das, Naglieri y Kirby, 1994).

A pesar de éstos y otros intentos interesantes y valerosos, ningún método teórico ha logrado proporcionar una explicación totalmente satisfactoria sobre cómo la inteligencia se desarro-

lla y cambia, las causas de las diferencias individuales en la inteligencia, o los procesos cognoscitivos y fisiológicos específicos que son responsables de la actividad intelectual. Al parecer, todas las corrientes actuales son correctas en cierta medida, pero de seguro ninguna proporciona una explicación completa, empíricamente verificada, sobre la estructura y el funcionamiento cognoscitivo. Por el momento, parece que las teorías sobre procesamiento de información ofrecen la mejor oportunidad de lograr una concepción lógica y con base empírica de las habilidades cognoscitivas, pero la situación podría cambiar al avanzar las investigaciones. De cualquier modo, algo es cierto: surgirán otras teorías sobre la inteligencia, y su valor se determinará por su eficacia para predecir y explicar el aprendizaje y el pensamiento humanos.

Aplicaciones de las evaluaciones de inteligencia

En contraste con otras definiciones más teóricas, las definiciones operativas de la inteligencia se centran en su medición y en las aplicaciones relacionadas. Tal vez la más operativa de dichas definiciones fue la sugerida por E. G. Boring, quien propuso definir la inteligencia como “aquello que se mide por medio de un test de inteligencia”. Lo que sea que midan los tests de inteligencia, estas pruebas se han usado para varios fines prácticos, incluyendo (1) el diagnóstico de la habilidad mental alta y baja y la ubicación de los retrasados mentales o los superdotados en programas o clases especiales; (2) la selección (sondeo), colocación y clasificación de estudiantes en instituciones de educación superior, empleados en organizaciones de negocios o industriales y personal en dependencias militares y gubernamentales; (3) la determinación y el diagnóstico de discapacidades relacionadas con el trabajo por demandas de seguros; (4) la asesoría y rehabilitación vocacional y educativa; (5) el psicodiagnóstico de niños y adultos en contextos clínicos o psiquiátricos; (6) la evaluación de la efectividad de tratamientos psicológicos e intervenciones en el medio ambiente, y (7) los estudios sobre habilidades cognoscitivas y personalidad.

Pruebas individuales colectivas

A pesar del objetivo común de medir una habilidad unitaria, los formatos de todos los tests de inteligencia general no son idénticos. En algunos hay reactivos de distintos tipos mezclados o alternados, y aumenta su dificultad a lo largo de la prueba. Los reactivos de otros tests de inteligencia se agrupan como conjuntos de subpruebas programadas en forma separada.

La forma más común de clasificar las pruebas de inteligencia es mediante la dicotomía *individual versus colectiva o de en grupo*. Los tests de inteligencia individual, que se aplican a una persona a la vez, tienen un enfoque algo distinto que los tests de inteligencia colectiva, los cuales pueden administrarse a muchas personas simultáneamente. El énfasis de las pruebas individuales es más global u holístico: su principal función es evaluar una habilidad cognoscitiva general. Por otra parte, el enfoque del test colectivo tiende a ser más reducido: a predecir el desempeño académico o laboral. Además, administrar un test de inteligencia individual suele ser más laborioso que administrar una prueba. Una ventaja de las pruebas individuales es que los examinadores pueden prestar más atención a los sujetos de examen. El enfoque del examinado a la prueba y otros comportamientos —angustia, confianza, estrategias para resolver problemas, frustraciones, distracción y aspectos similares— pueden observarse más de cerca cuando se examina a una persona a la vez, y el desempeño puede estimularse y recompensarse en forma más efectiva. Asimismo, las calificaciones de pruebas individuales no dependen tanto de la capacidad de lectura como las calificaciones de pruebas aplicadas colectivamente.

La mayor economía de administrar una prueba en grupo en ciertas situaciones ocasiona que se administren más pruebas en grupo que individuales. Además, a pesar de lo que en ocasiones han sostenido los defensores de las pruebas individuales, ciertas pruebas de inteligencia aplicadas en forma colectiva grupal pueden incluso tener mayores coeficientes de validez que sus contrapartes individuales.

Los tests de inteligencia colectivos grupales se usan con mayor frecuencia para una selección inicial en situaciones educativas y laborales, que es seguida por una evaluación individual cuando el examinado obtiene una calificación deficiente en una prueba colectiva y/o se requiere más información sobre sus cualidades y fallas cognoscitivas. También es más probable que los tests de inteligencia individuales se usen en clínicas, hospitales y otros sitios donde se realizan diagnósticos clínicos. En dichos lugares las pruebas sirven no sólo como medidas de la habilidad mental general, sino también como medio de comprender más a fondo el funcionamiento de la personalidad y las discapacidades cognoscitivas específicas.

TESTS DE INTELIGENCIA INDIVIDUALES

Los instrumentos que provienen del trabajo de Lewis Terman y David Wechsler han sido las pruebas de inteligencia individuales más comunes. Con el paso del tiempo, estos tests se han usado para evaluar las habilidades intelectuales de niños y adultos en muchos contextos diferentes. Otras pruebas individuales, algunas de las cuales constituyen variantes o extensiones de los tests de Terman y de Wechsler, se han diseñado específicamente para evaluar las habilidades mentales de niños pequeños y personas con desventajas lingüísticas y/o físicas.

Otras ediciones de la Escala de Stanford-Binet

Hubo tres traducciones y adaptaciones de la escala Binet-Simon en Estados Unidos. Una fue preparada por H. H. Goddard de la Escuela de Capacitación Vineland, otra por Frederic Kuhlmann de la Universidad de Minnesota, y una tercera por Lewis Terman de la Universidad Stanford. La más popular de estas revisiones, la Escala de Inteligencia Stanford-Binet, fue publicada por Terman en 1916.

La Escala de 1916. Al igual que las anteriores escalas de Binet-Simon, la Stanford-Binet de 1916 era una escala de edad donde las subpruebas se agrupaban en niveles de edad cronológica. Terman seleccionó reactivos de las escalas de Binet-Simon, así como reactivos totalmente nuevos que representaban una muestra amplia de las tareas que supuestamente requerían capacidades intelectuales aprovechadas. También se realizaron esfuerzos para incluir tareas que no eran tan dependientes de experiencias de aprendizaje escolares específicas.

Un criterio para incluir un reactivo en la escala de Stanford-Binet era que un porcentaje creciente de niños en niveles de edad sucesivos deberían ser capaces de responder el reactivo en forma correcta. Por algunas razones estadísticas que tienen que ver con mantener una escala de cociente de inteligencia bastante estable a través de los niveles de edad, el porcentaje de aprobados requerido se estableció más bajo en reactivos incluidos en subpruebas en niveles de años superiores que en reactivos de niveles de años inferiores. De cualquier modo, el criterio del porcentaje de aprobados sirvió como un medio objetivo de asegurarse que cada reactivo de la prueba se ubicara en un nivel de edad adecuado.

La *edad mental* (EM) y el *cociente de inteligencia* de un examinado en la escala Stanford-Binet dependían de la cantidad de subpruebas aprobadas en los niveles de edad sucesivos. El cociente de inteligencia se determinaba dividiendo la edad mental del examinado (EM), la cantidad

total de crédito de meses obtenida en la prueba, por su edad cronológica (EC) en meses y multiplicando el cociente resultante por 100. En símbolos, esta *razón de CI* se calculaba como:

$$CI = 100 \frac{MA}{CS} \quad (7.1)$$

Durante muchos años, la Escala de Inteligencia de Stanford-Binet funcionó como un estándar con respecto al cual se evaluaban otros tests de inteligencia. Sin embargo, tenía varias desventajas. Por ejemplo, la versión de 1916 sólo se estandarizó en 1,000 niños y 400 adultos. De acuerdo con las normas actuales, la muestra no se seleccionó con cuidado y no era representativa de la población estadounidense de la época. Otras dos desventajas fueron la inadecuación al evaluar adultos y niños muy pequeños, y la falta de una segunda forma para permitir la reevaluación. Por lo tanto, en 1937, Terman y su socia, Maud Merrill, publicaron una versión revisada, actualizada y reestandarizada de la escala.

La Escala de 1937. La versión de 1937 de la Escala de Inteligencia de Stanford-Binet tenía un límite *inferior* menor y uno *superior* mayor que la escala de 1916, dos formas paralelas (L y M) y una mejor estandarización. La escala de 1937 fue estandarizada de manera estratificada en 100 niños, con un intervalo por cada medio año de edad, desde el año y medio hasta los cinco y medio años; 200 niños con intervalos por cada año de edad desde los 6 hasta los 14 años, y 100 niños con intervalos por cada año de edad desde los 15 hasta los 18 años. Se administró la prueba a un número igual de niñas y niños en 17 comunidades de 11 estados, pero la muestra se limitó a individuos blancos nativos, quienes, como grupo, estaban en cierta medida por encima del promedio en cuanto a situación socioeconómica. En consecuencia, la muestra no era verdaderamente representativa de toda la población de Estados Unidos.

Se usaron tres criterios para incluir un reactivo en la escala: (1) el reactivo se consideró como una medida de comportamiento inteligente; (2) el porcentaje de niños que pasaban el reactivo aumentaba con la edad cronológica, y (3) los niños que aprobaron el reactivo tenían una edad mental media superior que la de quienes fracasaron en el reactivo. Los reactivos se agruparon en intervalos de medio año (niveles) del Año II al Año V, y en intervalos de un año desde el Año VI hasta el Año XIV; también había nivel Promedio de Adultos y tres niveles Superiores de Adulto (Adulto Superior I, II y III). Cada una de las seis subpruebas por nivel desde el Año II hasta el Año V recibió un mes de crédito, y las seis subpruebas en niveles Superiores de Adulto I, II y III tuvieron 4-, 5- y 6- meses de crédito, respectivamente.

Al evaluar a un niño con la Escala Stanford-Binet, el examinador primero determinaba la *edad basal* del niño. La edad basal era el nivel de años más alto en que el niño pasaba todas las subpruebas. La evaluación continuaba entonces hasta la *edad tope*, el nivel de años inferior en que el niño fallaba en todas las pruebas. La edad mental se calculaba añadiendo a la edad basal el número de meses de crédito recibido por pasar cada subprueba hasta la edad tope. Entonces se calculaba el CI mediante la fórmula 7.1.

La Escala de 1960. La tercera edición de la Escala de Inteligencia de Stanford-Binet, publicada en 1960, consistía en una actualización de los mejores reactivos de las formas L y M. Al igual que sus predecesoras, la tercera edición se usaba para medir la inteligencia de individuos desde la edad de dos años hasta la adultez. El procedimiento para administrar la prueba era similar al de la escala de 1937, pero se introdujeron algunos cambios. Uno de éstos consistía en una subprueba alternativa en cada nivel de edad para usarla cuando alguna de las subpruebas no se

aplicaba o se aplicaba de modo incorrecto. El tiempo de la prueba también podía reducirse en ciertos casos administrando sólo cuatro subpruebas seleccionadas en lugar de seis en cada nivel de un año. Otro cambio fue la disposición para prevenir desviaciones del CI. La razón del CI, al igual que cualquier otra norma de edad, no satisfizo el requisito de igualdad de unidades de edad. Asimismo, no tenía sentido cuando se aplicaba a adultos, porque no había una respuesta satisfactoria a la pregunta sobre qué edad cronológica debía usarse como denominador de la relación MA/CA al evaluar adultos. Se han propuesto las edades de 14, 16 y 18 años como la edad en que el crecimiento mental se detiene y, por lo tanto, cualquiera de esas edades puede ser un denominador adecuado para calcular el CI. Debido a los problemas para determinar la razón del CI, se tomó la decisión de cambiar de un CI de razón a una calificación estándar escala de *desviación CI*, con una media de 100 y desviación estándar de 16. Ocasionalmente se siguió reportando la razón de CI antigua y se incluían tablas para calcularla en el manual de Stanford-Binet de 1960.

La muestra de estandarización para la Forma 1960 L-M de la Escala Stanford-Binet consistió en 4,500 niños, de entre 2½ y 18 años de edad, que habían tomado cualquiera de las formas L o M de la Escala de 1937 entre 1951 y 1954. Tomando en cuenta la necesidad de normas actualizadas, el editor hizo adaptaciones para la prueba al ser administrada en 1972 a una muestra nacional estratificada de 2,100 niños (100 niños por cada intervalo de medio año desde los 2 hasta los 5½ años, y por cada intervalo de un año también 100 niños, éstos de 6 a 18 años). La muestra era más representativa que las anteriores muestras normativas de la población general de Estados Unidos. Con base en la estandarización de 1972, se publicó un manual revisado para la tercera edición (Terman y Merrill, 1973). El manual incluía coeficientes de confiabilidad de test-retest de más de .90 y, como en las dos primeras ediciones, correlaciones moderadas con grados escolares y calificaciones de pruebas de aprovechamiento (.40 a .75).

Cuarta edición de la Escala Stanford-Binet

La cuarta edición de la Escala de Inteligencia Stanford-Binet (SB-IV) (por Riverside Publishing) se elaboró considerando las necesidades de psicólogos clínicos, escolares y otros psicólogos que usan la información de los tests de inteligencia. SB-IV mantuvo la continuidad histórica con las versiones anteriores de la escala, pero representó una marcada separación de sus predecesoras en cuanto a sus bases teóricas y psicométricas, su contenido y el procedimiento de administración. Al igual que muchas pruebas modernas, SB-IV fue desarrollada usando procedimientos psicométricos complejos, tales como la teoría de respuesta al ítem (escala de Rasch) y análisis de sesgo étnico. Además estaba diseñada no sólo para ayudar a identificar individuos con retraso mental o superdotados, sino también a proporcionar información diagnóstica sobre discapacidades de aprendizaje específicas. Con respecto al sesgo por sexo y etnia, se omitieron los reactivos considerados injustos o que mostraban diferencias estadísticas atípicas entre sexos o grupos étnicos.

Modelo teórico y pruebas. Como se diagrama en la figura 7.2, el modelo en que se basó la escala SB-IV consiste en una jerarquía de tres niveles con un factor de inteligencia general (*g*) en el primer nivel, tres factores amplios (habilidades cristalizadas, habilidades fluido-analíticas y memoria de corto plazo) en el segundo nivel, y tres factores (razonamientos verbal, cuantitativo y abstracto-visual) en el tercer nivel. Los factores de razonamiento cuantitativo y verbal comprenden el factor de habilidades cristalizadas en el segundo nivel, y el factor abstracto-visual en el tercer nivel comprende el factor de habilidades fluido-analíticas en el segundo nivel.

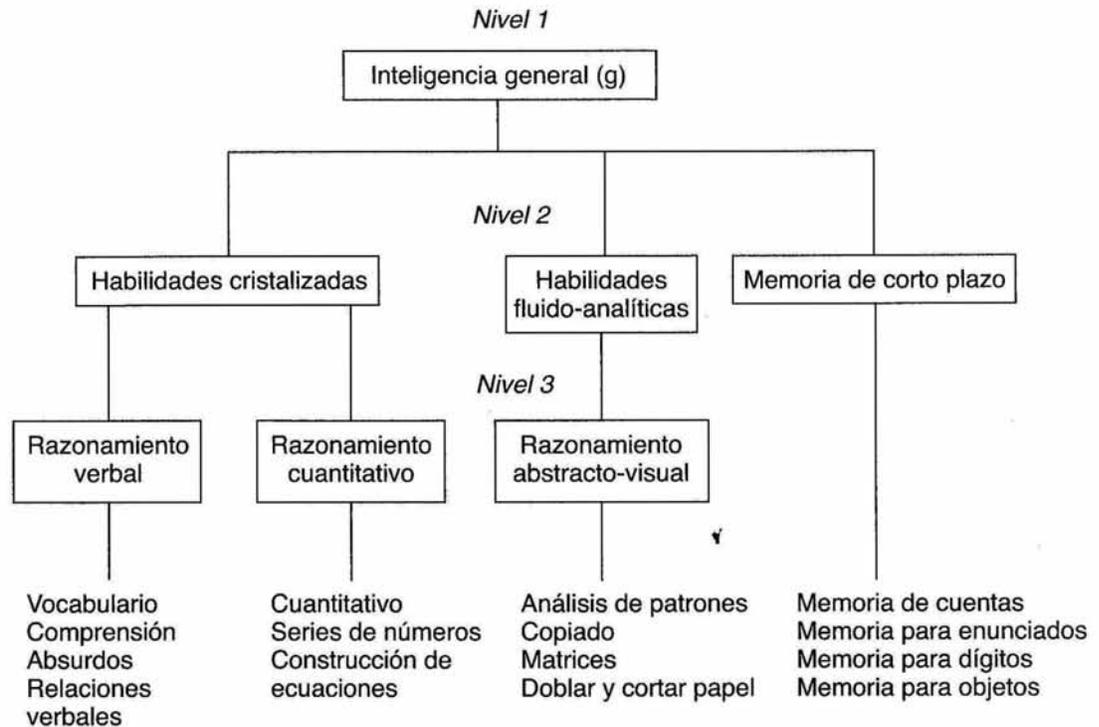


FIGURA 7.2 Modelo teórico y pruebas para la escala Stanford-Binet IV.

(Derechos Reservados 1986 por Riverside Publishing Company. Reproducido de *The Stanford-Binet Intelligence Scale*, cuarta edición, por Robert L. Thorndike, Elizabeth P. Hagen y Jerome M. Sattler, con autorización del editor.)

Al igual que sus antecesoras, la escala SB-IV fue diseñada para medir la inteligencia desde los 2 años hasta la edad adulta. Hay 15 tests: tres o cuatro tests en cada una de las tres categorías más amplias del Nivel 3 (Razonamiento Verbal, Razonamiento Cuantitativo, Razonamiento Abstracto-Visual), además de cuatro tests de Memoria de Corto Plazo (vea la figura 7.2). Cada prueba se acomoda en una serie de niveles que consisten en dos reactivos cada uno. Casi todas las pruebas incluyen reactivos de muestra para familiarizar a los examinados con el carácter de la tarea específica.

Aplicación. El tiempo de administración para toda la escala SB-IV es de aproximadamente 75 minutos, y varía de acuerdo con la edad del examinado y la cantidad de pruebas administradas. El carácter adaptativo, o de múltiples etapas, de la prueba exige administrar el Test de Ruta (Vocabulario) primero para determinar el nivel inicial en los demás tests. El nivel de entrada en el Test de Ruta se determina por la edad cronológica del examinado. La administración del Test de Ruta continúa mientras el examinado no falle en tres o cuatro reactivos en dos niveles consecutivos, el más alto de los cuales es el *nivel crítico*. El nivel inicial para las 14 pruebas restantes se establece a partir de la tabla mediante una combinación del nivel crítico del Test de Ruta y la edad cronológica del examinado (*nivel basal*) y hacia arriba hasta que falla en tres o cuatro reactivos en dos niveles consecutivos. El más alto de estos niveles es la *edad tope* del examinado para esa prueba.

Calificación. Las puntuaciones crudas en cada una de las 15 pruebas son iguales a la cantidad de reactivos aprobados. Estas puntuaciones se convierten, dentro de cada grupo de edad, en calificaciones normalizadas de *escala de edad estándar* (SAS) con una media de 50 y desviación estándar de 8. Las puntuaciones crudas en cada una de las cuatro áreas (Razonamiento Verbal, Razonamiento Abstracto-Visual, Razonamiento Cuantitativo, Memoria de Corto Plazo) son iguales a la suma de las puntuaciones crudas en las tres o cuatro pruebas que comprenden esa área. Estas puntuaciones de área se convierten en calificaciones de escala estándar (*calificaciones de área SAS*) con una media de 100 y desviación estándar de 16. Por último, una *calificación compuesta* que consiste en la suma de las cuatro puntuaciones de área se convierte a una escala de calificaciones estándar con una media de 100 y desviación estándar de 16. El rango de las calificaciones compuestas generales es de 36 a 164, que es el equivalente a un rango de calificaciones z de -4 a $+4$.

Estandarización. En Estados Unidos, la escala SB-IV fue estandarizada en 5,013 individuos de entre 2 y 23 años 11 meses de edad en 47 estados y el Distrito de Columbia. La muestra de estandarización fue estratificada por género y raza-etnia, y los estudiantes también fueron estratificados de acuerdo con la posición relativa en su clase. A pesar de los esfuerzos por seleccionar una muestra de estandarización que fuese verdaderamente representativa de la población estadounidense, la muestra contenía cantidades desproporcionadas de individuos de los niveles socioeconómicos y educativos más altos. Se intentó corregir este error al calificar las pruebas, pero el esfuerzo no fue del todo exitoso. Otros problemas son que los factores medidos por la escala no son uniformes en todos los niveles de edad y la información de confiabilidad del manual es inadecuada. Sin embargo, los coeficientes de división por mitad y de test-retest, calculados en medidas obtenidas a lo largo de un intervalo de 2 a 8 meses indican que las confiabilidades de las 15 pruebas, las cuatro áreas, y el conjunto son satisfactorias.

Las pruebas de Wechsle

Aunque las subpruebas en el nivel adulto se han incluido en la escala Stanford-Binet desde la revisión de 1937, nunca ha habido una medida muy satisfactoria de la inteligencia en adultos. Por consiguiente, en 1939 David Wechsler, un psicólogo del Hospital Bellevue en Nueva York, publicó un test de inteligencia individual diseñado específicamente para adultos. Para esta prueba, la Forma I de la Escala de Inteligencia de Wechsler-Bellevue, Wechsler añadió una segunda forma en 1947, la Forma II de la Escala de Inteligencia de Wechsler-Bellevue. Una revisión completa y reestandarización de la Forma I se publicó en 1955 como la Escala de Inteligencia para Adultos de Wechsler (WAIS). La WAIS misma fue modificada, reestandarizada y reeditada por The Psychological Corporation en 1981 como la Escala de Inteligencia para Adultos de Wechsler Revisada para evaluar la inteligencia de adultos entre 16 y 74 años de edad.

Escala de Inteligencia para Adultos de Wechsler, Revisada. Las seis subpruebas verbales (V) y cinco subpruebas de ejecución (E) de la Escala de Inteligencia para Adultos de Wechsler Revisada (WAIS-R), por orden de administración, se describen en la tabla 7.2. Las subpruebas Verbal y de Ejecución se administran alternadamente, y dentro de cada subprueba los reactivos se presentan en orden de dificultad creciente. Se requieren alrededor de 75 minutos para administrar las once subpruebas, y la administración en una subprueba en particular se discontinúa cuando el examinado falla en una cantidad específica de reactivos sucesivos.

Calificación. Las puntuaciones crudas en las once subpruebas del WAIS-R se convierten a una escala de calificación normalizada con una media de 10 y desviación estándar de 3. Entonces,

TABLA 7.2 Subpruebas de la Escala de Inteligencia para Adultos de Wechsler Revisada

Información (V): 33 preguntas sobre información general que deben contestarse en pocas palabras o números.

Completamiento de dibujos (E): 27 dibujos en tarjetas, cada uno con una parte faltante; el examinado tiene 20 segundos para indicar lo que hace falta en la imagen.

Serie de dígitos (V): 7 series de dígitos que deben recitarse hacia adelante y 7 series para repetirse hacia atrás.

Ordenamiento de dibujos (E): 10 series de tarjetas, cada una con una pequeña imagen; se pide al examinando que ordene los dibujos de cada serie de tarjetas para construir una historia coherente.

Vocabulario (V): se presentan 37 palabras en orden de dificultad creciente que deben definirse.

Diseño de cubos (E): 10 diseños geométricos en rojo y blanco en tarjetas y nueve bloques de los mismos colores; se solicita al examinado que copie cada uno de los diseños usando 4 o 9 bloques.

Aritmética (V): se presentan 15 problemas aritméticos en orden de dificultad creciente.

Ensamble de objetos (E): se presentan 4 rompecabezas de cartón al examinado en un formato preestablecido; se pide al examinado que una las piezas para armar algo.

Comprensión (V): 18 preguntas que requieren respuestas detalladas se presentan en orden de dificultad creciente.

Símbolos en dígitos (E): 93 casillas que deben llenarse con el símbolo codificado correcto correspondiente al número que aparece sobre la casilla.

Semejanzas (V): 14 reactivos del tipo “¿En qué son similares A y B?”

V, subprueba verbal; E, subprueba de ejecución.

mediante la referencia a una tabla especial que viene dentro del manual de administración, la suma de las puntuaciones escaladas de la subpruebas en la Escala verbal puede convertirse en un CI Verbal, la suma de las puntuaciones escaladas de las subpruebas en la Escala de Desempeño, en un CI de Desempeño, y la suma de las puntuaciones escaladas de las once subpruebas, en una Escala Completa de CI. Éstas son desviaciones CI, expresadas en números en una escala de calificaciones estándar con una media de 100 y una desviación estándar de 15.

Estandarización. La escala WAIS-R se estandarizó en una muestra nacional, cuidadosamente seleccionada, de 1,880 adultos “normales” ubicados en nueve grupos de edad (16 a 17, 18 a 19, 20 a 24, 25 a 34, 35 a 44, 45 a 54, 55 a 64, 65 a 69 y 70 a 74) dentro del rango de 16 a 74 años. La muestra de cada categoría de edad se estratificó por sexo, región geográfica, blanco contra no blanco, educación y ocupación. Se controlaron otras características, tales como residencia urbana contra rural, pero no funcionaron como variables de estratificación. La estandarización de la WAIS-R difirió de la presentada originalmente por la de WAIS en 1955, sobre todo en cuanto a la estratificación de la muestra por grupo étnico y la provisión de muestras más representativas para adultos mayores.

Significado diagnóstico de las calificaciones de Wechsler. Al diseñar la escala WAIS, Wechsler planeó obtener más de un cálculo de la habilidad mental general de una persona. Se consideró que una diferencia significativa entre los CI Verbal y de Ejecución de una persona y el patrón de calificaciones (dispersión) en las once subpruebas era característica de cierto tipo de trastornos mentales y, por lo tanto, potencialmente útil para el diagnóstico clínico. Desafortunadamente, la investigación proporciona poco apoyo para las hipótesis de Wechsler en cuanto a la importancia diagnóstica de la dispersión de la calificación escalada en las diversas subpruebas.

Un problema al tratar de analizar la dispersión de la calificación de las subpruebas en las Escalas de Wechsler es que dichas calificaciones no son muy confiables y algunas subpruebas tienen correlaciones considerables entre sí. En consecuencia, la diferencia entre las calificaciones escaladas de una persona en dos subpruebas dadas debe ser muy grande antes de que pueda considerarse significativa. Las diferencias pronunciadas entre las calificaciones escaladas de subpruebas y entre CI Verbales y de Ejecución tienen cierto valor en el diagnóstico de daño cerebral orgánico y psicopatología y en la diferenciación entre inteligencia y oportunidad. Un CI Verbal considerablemente inferior al CI de Ejecución, por ejemplo, puede ser resultado de una experiencia lingüística limitada o de *carencia cultural*.

WAIS-III. Así como en otras pruebas de habilidades cognoscitivas, el contenido y las normas de los tests de inteligencia de algún modo pierden actualidad con los años. Por ello, una nueva edición de WAIS-R, la WAIS-III, se elaboró a mediados de la década de 1990 y fue publicada por The Psychological Corporation en 1997. Al elaborar la WAIS-III, se prestó particular atención a las subpruebas verbales, tales como Información, Vocabulario y Comprensión, las cuales, debido a que están más sujetas a cambios culturales, se vuelven obsoletas más pronto que otras subpruebas.

Además de las revisiones de las once subpruebas de la WAIS-R, se incluyeron tres nuevas subpruebas en la WAIS-III: Razonamiento de Matriz, Búsqueda de Símbolos y Secuencias de Letras y Números. El Razonamiento de Matriz consiste en una serie de imágenes de cinco formas geométricas; se requiere que los examinados nombren o señalen la forma correcta. En esta subprueba se incluyen cuatro tipos de reactivos, Completar Patrones, Clasificación, Razonamiento por Analogía y Razonamiento Serial. La subprueba de Búsqueda de Símbolos comprende un conjunto de grupos pareados, en el que cada par consiste en un grupo meta y un grupo de búsqueda. Los examinados marcan la casilla adecuada para indicar si cada símbolo meta aparece en el grupo de búsqueda. La subprueba de Secuencias de Letras y Números es una serie de letras y números presentados oralmente en desorden. Los examinados reordenan y repiten la lista diciendo los números en orden ascendente y repitiendo luego las letras en orden alfabético.

El tiempo de aplicación de la WAIS-III es menor que el de la WAIS-R, y el nivel inferior para la mayoría de las subpruebas se ha reducido con el propósito de lograr un mejor cálculo del funcionamiento cognoscitivo de los individuos con retraso mental. Además de los CI tradicionales, Verbal, de Ejecución y de Escala Completa, se obtienen cuatro calificaciones de índice de factores (Comprensión Verbal, Memoria de Trabajo, Organización Perceptual y Velocidad de Procesamiento).

La WAIS-III se estandarizó en una muestra de 2,450 adultos de edades entre 16 y 89 años. La muestra fue estratificada por raza-etnia (blancos, afroamericanos, latinos, otros), sexo, nivel educativo y región geográfica en cada grupo de edad. Las correlaciones entre las calificaciones WAIS-III y las de otras pruebas de la familia Wechsler, así como la cuarta edición de la Escala Stanford-Binet y las Matrices Progresivas de Raven, se incluyen en el manual. También se proporcionan estadísticas basadas en diversos grupos clínicos. Las escalas de la WAIS-III se ajustaron a las normas de la tercera edición de la Escala de Memoria Wechsler (WMS-III), lo que permite efectuar un examen de la relación entre el funcionamiento intelectual de una persona y su memoria.

Escala de Inteligencia para Niños de Wechsler, tercera edición. La Escala de Inteligencia para Niños de Wechsler (WISC), una extensión hacia abajo de la Forma I de la Escala de Wechsler-Bellevue, fue publicada por The Psychological Corporation en 1949. En 1974 se publicó una

revisión de la WISC, la WISC-R, y en 1991 apareció publicada la Escala de Inteligencia para Niños de Wechsler, tercera edición (WISC-III). Esta prueba, diseñada para niños de entre 6 y 16 años 11 meses, consiste en las siguientes seis subpruebas Verbales y siete subpruebas de Ejecución:

SUBPRUEBAS VERBALES	SUBPRUEBAS DE EJECUCION
Información	Completamiento de dibujos e imágenes
Semejanzas	Codificación
Aritmética	Acomodamiento de dibujos
Vocabulario	Diseño de cubos
Comprensión	Ensamble de objetos
Intervalo de dígitos (complementaria)	Búsqueda de símbolos (complementaria)
	Laberintos (complementaria)

Las diez subpruebas principales (no complementarias) pueden administrarse en un lapso de entre 50 y 70 minutos, y las subpruebas complementarias en otros 10 o 15 minutos adicionales. Así como en la WAIS-R, las subpruebas Verbales y de Ejecución de WISC-III se administran alternadamente. Los CI Verbal, de Ejecución y de Escala Completa, basados en la misma escala de calificaciones normalizadas que los de la WAIS-III, se determinan añadiendo la calificación escalada de las cinco subpruebas Verbales y las cinco de Ejecución que se aplican. El WISC-III también puede calificarse para cuatro factores: Comprensión Verbal, Organización Perceptual, Libertad y Distracción y Velocidad de Procesamiento.

La WISC-III fue estandarizada en muestras representativas de estadounidenses de 100 niños y 100 niñas en cada uno de once grupos de edad de los 6 a los 16 años. Las muestras también se estratificaron por región geográfica, nivel educativo de los padres y raza. Se evaluaron otras muestras de niños con WISC-III y con la WAIS-R o la WPPSI-R, dependiendo de sus edades. Las confiabilidades de test-retest de la WISC-III, obtenidas al readministrar la escala después de 4 a 8 semanas, son satisfactorias. Asimismo, se han llevado a cabo diversos estudios de validación con varios grupos clínicos de niños.

Escala de Inteligencia para Nivel Preescolar y Primaria de Wechsler Revisada. Una tercera prueba de Wechsler, la Escala de Inteligencia para Nivel Preescolar y Primaria de Wechsler (WPPSI), fue publicada por The Psychological Corporation en 1967 y una revisión, la WPPSI-R, en 1989. Las seis subpruebas Verbales (V) y las seis de Ejecución (E) de la WPPSI-R, en orden de aplicación, son: Ensamble de Objetos (E), Información (V), Diseño Geométrico (E), Comprensión (V), Diseño de Bloques (E), Aritmética (V), Laberintos (E), Vocabulario (V), Completar Imágenes (E), Semejanzas (V), Piezas con forma de animales (E), y Enunciados (V). Las últimas dos son subpruebas complementarias. Diseñada para niños de entre 3 y 7 años de edad, la WPPSI-R fue estandarizada a fines de la década de 1980 sobre una muestra nacional de niños estadounidenses de entre 3 y 7 años de edad. Estratificar la muestra por género, etnia y nivel educativo y ocupacional de los padres, la hizo más representativa de la población de Estados Unidos en este rango de edad. Del mismo modo que la WAIS-R y la WISC-III, la WPPSI-R produce CI Verbal, de Ejecución y de Escala Completa por separado, con base en una escala de calificación estándar con una media de 100 y desviación estándar de 15.

Escala de Inteligencia de Wechsler Abreviada. En contextos clínicos y educativos, la necesidad de una medición confiable de la inteligencia que pudiera realizarse en forma más rápida que la WAIS-III y la WISC-III condujo a la creación de la Escala de Inteligencia de Wechsler

Abreviada (WASI). Las subpruebas de la WASI se construyeron independientemente de las correspondientes subpruebas de la WAIS-III y la WISC-III, pero en forma paralela. La forma de cuatro subpruebas de la WASI consiste en subpruebas de Vocabulario, Semejanzas, Diseño de Cubos y Razonamiento de Matriz. Las primeras dos constituyen la Escala Verbal y las últimas dos la Escala de Ejecución de la WASI. La forma de dos subpruebas de la WASI incluyen Vocabulario y Razonamiento de Matriz. La forma de cuatro subpruebas requiere de aproximadamente 30 minutos y la de dos subpruebas toma alrededor de 15 minutos en administrarse.

Otras pruebas de inteligencia individuales de rango amplio

Aunque son las pruebas de inteligencia individuales más populares en Estados Unidos, la Stanford-Binet y la de Wechsler de ningún modo son las únicas baterías de amplio rango para evaluar la habilidad mental general. Tampoco son las pruebas más populares de habilidad mental en otros países. De particular relevancia en el Reino Unido son las Escalas de Habilidad Británicas (BAS), que fueron revisadas por The Psychological Corporation y reestandarizadas en Estados Unidos como Escalas de Habilidad Diferencial (DAS).

Escalas de habilidad diferencial. El objetivo de las Escalas de Habilidad Diferencial (DAS) (de The Psychological Corporation) es proporcionar perfiles de habilidad para analizar y diagnosticar problemas de aprendizaje en los niños, evaluar cambios en las habilidades con el tiempo e identificar, seleccionar y clasificar a los niños (de entre 2¹/₂ y 17 años de edad) con problemas de aprendizaje. Las DAS consisten en 20 subpruebas, incluyendo 12 subpruebas principales, 5 subpruebas de diagnóstico y 3 subpruebas de aprovechamiento. Las tres subpruebas de aprovechamiento (Habilidades Numéricas, Ortografía, Lectura de Palabras) son útiles para evaluar habilidades académicas básicas, pero las subpruebas centrales y de diagnóstico proporcionan el principal medio de evaluar las cognoscitivas. A cada examinando se le aplican de cuatro a seis subpruebas centrales, de los 2 años 6 meses a los 17 años 11 meses. Se combinan las calificaciones de varias subpruebas centrales para obtener índices generales de Habilidad Verbal, Habilidad de Razonamiento No Verbal y Habilidad Conceptual General, en una escala con una media de 100 y desviación estándar de 15. Aunque las subpruebas de diagnóstico no se usan para calcular los índices de habilidad, proporcionan información útil para comprender las ventajas y deficiencias cognoscitivas del niño.

Las normas de las DAS se basan en 3,475 niños estadounidenses; la muestra fue estratificada por edad, sexo, raza-etnia, educación de los padres, región geográfica e inscripción en educación preescolar. Los niños especiales (con trastornos de aprendizaje, dificultades de habla y lenguaje, retrasados mentales susceptibles de ser educados, superdotados, emocionalmente perturbados, con trastornos sensoriales o motrices) se incluyeron en la muestra.

Prueba Detroit de Habilidad de Aprendizaje. Otra batería relevante es la Prueba Detroit de Habilidad de Aprendizaje (de pro.ed). La administración de la principal edición de esta batería (DTLA-P-2), que fue diseñada para niños de entre 3 y 9 años de edad, dura entre 15 y 20 minutos. Las subpruebas incluyen Articulación, Compaginación Conceptual, Reproducción de Diseños, Secuencia de Dígitos, Dibujar una Persona, Secuencias de Letras, Instrucciones Motoras, Secuencias de Objetos, Instrucciones Orales, Imágenes Fragmentadas, Identificación de Imágenes, Imitación de Enunciados y Relaciones Simbólicas.

La cuarta edición de la Prueba Detroit de Habilidad de Aprendizaje (DTLA-4) fue diseñada para niños de 6 a 17 años y tarda de 50 a 90 minutos en administrarse. Las subpruebas de la DTLA-4 incluyen Palabras Opuestas, Secuencias de Diseño, Imitación de Enunciados, Letras Invertidas, Construcción de Historias, Reproducción de Diseños, Información Básica, Relaciones Simbó-

licas, Secuencias de Palabras y Secuencias de Historias. Calificaciones normalizadas, rangos percentilares y equivalentes de edad pueden determinarse para las diez subpruebas y las pruebas compuestas (General, Nivel Óptimo, Dominio, Teórico). Las pruebas compuestas de Dominio son Verbal, No Verbal, Aumento de la Atención, Reducción de la Atención, Motricidad Aumentada, Motricidad Reducida. Las compuestas teóricas son de inteligencia Fluida y Cristalizada, de Asociación y Cognoscitiva, Simultánea y Sucesiva, Verbal y de Ejecución. La DTLA-4 es un mejoramiento sobre sus predecesoras con respecto a la claridad, facilidad de administración, estandarización, confiabilidad, validez y otras características estadísticas.

Pruebas de inteligencia de Kaufman. La Batería de Kaufman de Evaluación para Niños (K-ABC) (del American Guidance Service) fue diseñada por A. S. Kaufman y N. L. Kaufman con el propósito de evaluar las habilidades de niños de entre 2¹/₂ y 12¹/₂ años de edad para resolver problemas que requieren de un procesamiento mental simultáneo y secuencial. La K-ABC también incluye una Escala de Aprovechamiento para medir habilidades adquiridas en lectura y aritmética. Basada en una extensa investigación sobre neuropsicología y psicología cognoscitiva, la K-ABC fue diseñada especialmente para niños de edad preescolar, menores de edad, y excepcionales. 13 de las 16 subpruebas tipo juego que comprende la K-ABC pueden administrarse en un lapso de 35 a 85 minutos. Las calificaciones se obtienen en cuatro áreas globales: Procesamiento Secuencial, Procesamiento Simultáneo, Compuesta de Procesamiento Mental (Secuencial más Simultánea) y Aprovechamiento.

La muestra de estandarización para la K-ABC, basada en estadísticas registradas en el censo de Estados Unidos de 1980, se estratificó por raza (blanca, negra, latina, asiática, indígena estadounidense) e incluía un grupo representativo de niños excepcionales. Se establecieron normas de rangos percentilares separados por raza y nivel socioeconómico para niños blancos y negros. Los coeficientes de confiabilidad de división por mitades para las cuatro escalas globales en la K-ABC están en los rangos que van del.80 y el.90. En el manual también se da información sobre la validez de constructo, concurrente y predictiva de la prueba.

Otras dos pruebas de inteligencia relevantes diseñadas por A. S. Kaufman y N. L. Kaufman, publicadas por el American Guidance Service, son la Prueba de Inteligencia de Kaufman para Adolescentes y adultos (KAIT) y la Prueba Breve de Inteligencia de Kaufman (K-BIT). Ambas pruebas se basan en la teoría de R. B. Cattell sobre la inteligencia fluida y cristalizada. La KAIT se diseñó para edades entre los 11 y 85+ y toma entre 60 y 90 minutos; la K-BIT está diseñada para edades de 4 a 90 años y dura de 15 a 20 minutos.

Pruebas Woodcock-Johnson III de Habilidades Cognoscitivas. La Woodcock-Johnson III (WJ III) (de Riverside Publishing) consiste en dos baterías co-normalizadas para medir la habilidad intelectual general, habilidades cognoscitivas específicas y el aprovechamiento académico. Una batería, las Pruebas de Habilidades Cognoscitivas Woodcock-Johnson III (WJ III), se basa en la teoría de habilidades cognoscitivas de Cattell-Horn-Carroll (CHC) (vea Woodcock, 1998). Esta batería consiste en una Batería Estándar de diez pruebas y una Batería Ampliada de diez pruebas adicionales. Las pruebas tienen un rango amplio de edad y grado (de 2 a 90+ años; desde jardín de niños hasta la universidad) y una duración de evaluación relativamente breve (aproximadamente cinco minutos por prueba).

Las calificaciones de seis grupos: Verbal-Estándar, de Pensamiento-Estándar, Eficiencia Cognoscitiva-Estándar, Percepción Fonémica, Memoria Funcional y Recuerdo Demorado, se determinan a partir de la Batería Estándar. Las calificaciones de catorce grupos adicionales se obtienen cuando se aplica la Batería Ampliada. Además de las calificaciones en los grupos separa-



dos, se calcula una calificación de Habilidad Intelectual General (GIA) al combinar las calificaciones de las primeras siete pruebas o una calificación GIA (Ampliada) administrando 14 pruebas cognoscitivas. Puede calcularse una calificación de Habilidad Intelectual Breve (BIA) combinando las calificaciones de las pruebas de Comprensión Verbal, Formación de Conceptos y Compaginación Visual. También pueden determinarse calificaciones en los siguientes factores CHC: Comprensión-Conocimiento (Gc), Recuerdo a Largo Plazo (Glr), Pensamiento Visual-Espacial (Gv), Procesamiento Auditivo (Ga), Razonamiento Fluido (Gf), Velocidad de Procesamiento (Gs) y Memoria de Corto Plazo (Gsm).

Sistema de Evaluación Cognoscitiva Das-Naglieri. Otra prueba de inteligencia reciente es el Sistema de Evaluación Cognoscitiva Das-Naglieri (CAS)(Naglieri y Das, 1997)(de Riverside Publishing). El CAS es similar a la Woodcock-Johnson III en cuanto a que está basado en una teoría cognoscitiva y lo publica la misma compañía (Riverside Publishing Co.). Al orientarse hacia niños en edad escolar y adolescentes, el rango de edad del CAS (de 5 años a 17 años 11 meses) es más estrecho que el de la WJ III.

El CAS se diseñó “para proporcionar una medida del procesamiento cognoscitivo que sea justa para niños menores de edad, eficaz para un diagnóstico diferencial y relacionada con la intervención”. Se basa en la teoría PASS (Planeación, Atención, Simultánea, Sucesiva) de Das-Naglieri sobre la cognición, descrita en los inicios de ese capítulo, y es adecuada para niños en edad escolar y adolescentes. El tiempo de evaluación es de 40 minutos para la Batería Básica y de 60 minutos para la Batería Estándar. Las subpruebas se agrupan en los cuatro procesos cognoscitivos del modelo PASS:

PLANEACIÓN	SIMULTÁNEA
Números Correspondientes	Matrices No Verbales
Códigos Planeados	Relaciones Verbal-Espaciales
Conexiones Planeadas	Recuerdo de Figuras
ATENCIÓN	SUCESIVA
Atención Expresiva	Series de Palabras
Detección de Números	Repetición de Enunciados
Atención Receptiva	Velocidad de Habla (de 5 a 7 años de edad)
	Preguntas de Enunciados (de 8 a 17 años de edad)

La Batería Básica consiste en dos subpruebas, y la Batería Estándar en tres subpruebas, a partir de cada una de estas cuatro categorías.

Además de las calificaciones en las pruebas separadas, las calificaciones normalizadas con una media de 100 y desviación estándar de 15 se obtienen al combinar las calificaciones de todas las escalas. Como la Woodcock-Johnson III, el CAS se estandarizó cuidadosamente y tiene confiabilidades aceptables. Las evidencias de investigación relativas a diversos tipos de validez (de constructo, concurrente, predictiva y discriminante) se registran en el manual de la prueba.

Pruebas no verbales para los discapacitados

Los instrumentos psicométricos que requieren de señalar, manipular objetos o de otra respuesta no verbal, antes que de hablar o escribir, se conocen como *pruebas no verbales*. El desempeño en algunas tareas de estas pruebas puede facilitarse con el lenguaje verbal, pero su uso es mínimo.

El hecho de que las escalas Wechsler contengan medidas verbales y de ejecución separadas las hace más adecuadas que las versiones anteriores de Stanford-Binet para examinar perso-

nas con diferencias físicas, lingüísticas y culturales. Las subpruebas de desempeño Wechsler tienden a ser medidas más precisas de la habilidad mental en niños con problemas de audición y culturalmente distintos, mientras que las subpruebas verbales son medidas más válidas para los ciegos y débiles visuales. Al evaluar a personas ciegas, en ocasiones se ha aplicado una serie de seis pruebas de desempeño especialmente diseñadas conocidas como la Escala de Inteligencia Haptic para Adultos Ciegos, en conjunto con la Escala Verbal del WAIS, como medida de la inteligencia de adultos ciegos y débiles visuales.

Pruebas de una única tarea. Una de las pruebas no verbales más antiguas, el Tablero de Formas Seguin, se introdujo en 1866. Sin embargo, no fue sino hasta la primera parte del siglo XX cuando Knox, Kohs, Porteus y otros psicólogos realizaron serios esfuerzos por estandarizar dichas pruebas. Para medir las habilidades mentales, adicionalmente a muchos tipos de tableros de formas, se han utilizado tareas no verbales como rompecabezas de diversos tipos, el golpeteo secuencial de cubos, problemas de emparejamiento, diseños de cubos, laberintos, dibujo de personas y señalamiento de imágenes.

Los laberintos se han usado en forma extensa tanto en laboratorios psicológicos y clínicas como en varias pruebas estandarizadas. Los Laberintos de Porteus, publicados inicialmente en 1914 y descritos por su diseñador como una medida de la capacidad de previsión y planeación, consisten en un conjunto de laberintos ordenados por dificultad creciente. En cada laberinto se instruye al examinado para que trace el camino más corto entre el punto de partida y el final, sin levantar el lápiz ni entrar en un callejón sin salida. Para quienes padecen algún trastorno verbal, los Laberintos de Porteus son particularmente adecuados como prueba breve (25 minutos), y se han empleado en varias investigaciones y estudios antropológicos sobre los efectos de las drogas y la neurocirugía.

Otra prueba de ejecución no verbal para los discapacitados consiste en diseños de cubos tales como los de las escalas de Wechsler y las Escalas de Habilidad Diferencial. Una de las pruebas más antiguas de este tipo es el Diseño de Cubos de Kohs. Los materiales de la prueba de Kohs son 16 cubos de color y 17 tarjetas con diseños coloreados que el examinado debe copiar. El Diseño de Cubos se consideraba especialmente apropiado para niños con discapacidad de lenguaje y audición, pero ahora su aplicación es muy esporádica.

La Escala de Madurez Mental de Columbia (CMMS) es otra prueba de una única tarea que sólo requiere de señalar. Esta prueba se diseñó originalmente para evaluar niños con parálisis cerebral, pero puede administrarse a otros niños con verbales y motrices disminuidas (discapacidades visuales, trastornos del habla, de la audición, retraso mental) así como a niños hiperactivos. Los materiales de prueba consisten en 92 reactivos (una serie de dibujos) impresos en tarjetas de 15 × 47.5 cm. Se pide al niño (de entre 3½ y 10 años de edad) que seleccione, a partir de una serie de dibujos presentados en cada tarjeta, la imagen que no pertenece al grupo. Al seleccionar, el niño usa discriminación perceptiva y clasificatoria o habilidades de razonamiento general que incluyen color, forma, tamaño, uso, número, partes faltantes y material simbólico. Los 92 reactivos de la CMMS están dispuestos en ocho niveles traslapados, pero sólo entre 51 y 65 reactivos se aplican de hecho a un examinado determinado. La prueba dura entre 15 y 20 minutos, y las instrucciones se dan en inglés o en español. La ejecución se expresa en términos de calificaciones de desviación de edad desde 50 hasta 150, así como en rangos percentilares, estatinas e índices de madurez.

Baterías de pruebas de ejecución. La primera batería de pruebas de ejecución estandarizadas que se distribuyó comercialmente fue la Escala Pintner-Paterson de Pruebas de Ejecución (1917). Igualmente conocida es la Escala Puntual Arthur de Pruebas de Ejecución, publicada inicialmente por Grace Arthur en 1925. Dos baterías de pruebas de ejecución que se han usado am-

pliamente en niños con discapacidades de habla y de audición, y que todavía están disponibles, son la Escala Leiter de Desempeño Internacional y las Pruebas Hiskey-Nebraska de Aprendizaje. También son interesantes algunas baterías recientemente publicadas, tales como la Prueba Comprensiva de Inteligencia No Verbal, la Prueba de Habilidad No Verbal de Naglieri y la Prueba Universal de Inteligencia No Verbal.

Prueba Hiskey-Nebraska de Habilidad de Aprendizaje. Esta prueba se diseñó específicamente para evaluar las capacidades cognoscitivas de niños con discapacidad auditiva. Consiste en 12 subpruebas no verbales aplicadas mediante instrucciones en pantomima a niños sordos o con instrucciones verbales a niños normales. La prueba se aplica en forma no acelerada y proporciona la edad mental y un cociente de inteligencia. En el momento de escribir el presente libro, la prueba Hiskey-Nebraska se estaba reestructurando en cuanto a sus normas por Slosson Educational Publications, de modo que las características demográficas de la muestra de estandarización se ajustaran a las de la población actual de Estados Unidos.

Leiter-R. La versión revisada de la Escala Leiter de Desempeño Internacional (Leiter-R) (de Stoelting) se promueve como una medida de las habilidades cognoscitivas que es justa para la cultura y adecuada para personas de varios contextos culturales. Tiene un rango de edad de 2 a 21 años y puede administrarse sin lenguaje verbal a niños con problemas de audición o con otros trastornos lingüísticos de expresión o de recepción y culturalmente diferentes, con discapacidades motrices, autistas e incluso a niños superdotados. Se solicita a los examinados que unan una serie de cartas de respuesta coloreadas con las ilustraciones correspondientes presentadas en un caballete. Las cuatro subpruebas de Razonamiento y las seis de Visualización de la batería de Visualización y Razonamiento requieren de un total de 40 minutos para administrar. Las ocho subpruebas de Memoria y las dos de Atención en la correspondiente batería toman 35 minutos. Evaluar el CI o LD/ADHD puede realizarse en 25 minutos administrando una batería incompleta; la Evaluación de Selección de Superdotados requiere de 35 minutos. La Leiter-R fue estandarizada en 1993 en 1,719 niños típicos y 692 atípicos de entre 2 y 12 años de edad. Las evidencias de confiabilidad y validez registradas en el manual indican que Leiter-R es un instrumento bastante seguro psicométricamente.

Prueba Comprensiva de Inteligencia No Verbal. La Prueba Comprensiva de Inteligencia No Verbal (CTONI)(de pro.ed) tiene un rango de edad muy amplio (de 6-0 a 90-11) y puede administrarse en alrededor de una hora. Es particularmente apropiada para calcular la inteligencia de niños y adultos con problemas de lenguaje o de habilidades motoras finas. Pueden ser personas que hablen una lengua distinta al inglés, tengan desventajas socioeconómicas o sean sordos, o que sufran algún trastorno del lenguaje, una discapacidad motora o un problema neurológico.

Las seis subpruebas de CTONI son Analogías Pictóricas, Categorías de Imágenes, Secuencias Pictóricas, Analogías Geométricas, Categorías Geométricas y Secuencias Geométricas. Estas subpruebas se diseñaron para medir el razonamiento analógico, las calificaciones categóricas y las habilidades de razonamiento secuencial, como lo revelan las respuestas a las imágenes de objetos familiares (animales, personas, juguetes y reactivos similares) y diseños geométricos (dibujos, esbozos inusuales, etc.). Así como en la prueba Leiter-R, en la CTONI los examinados indican sus respuestas señalando opciones alternativas. Las respuestas se califican entonces y las calificaciones se combinan para obtener tres cocientes compuestos: Cociente de Inteligencia No Verbal (CINV), Cociente de Inteligencia No Verbal Pictórica (CINVP) y Cociente de Inteligencia No Verbal Geométrica (CINVG).

La CTONI se estandarizó en 25 estados de Estados Unidos, Canadá y Panamá. Aunque bastante pequeñas, las muestras se estratificaron por género, regiones geográficas, etnia, raza, residen-

cia urbana-rural y discapacidad. Esta prueba reporta coeficientes de confiabilidad de .80 o mayores, y en el manual también se da cierta evidencia para la validez con referencia a criterios, de contenido y de constructo. De especial relevancia son los esfuerzos de los diseñadores de la CTONI para detectar y eliminar sesgos culturales, de género, raciales y lingüísticos en los reactivos.

Prueba Universal de Inteligencia No Verbal (UNIT). A diferencia de muchas pruebas no verbales que se caracterizan por una modalidad no verbal ya sea en la administración o bien en los formatos de respuesta, a fin de garantizar la justicia sin importar la cultura, la etnia, el género o la habilidad auditiva, la UNIT se desarrolló con ambas modalidades simultáneamente. La administración de esta batería de prueba implica múltiples modos de respuesta, incluyendo el uso de herramientas de manipulación, lápiz y papel así como señalamiento. El examinador usa ocho gestos universales de manos y cuerpo para explicar las tareas de la prueba al examinado. Además de estos gestos, la aplicación de la prueba incluye demostraciones por parte del examinador, reactivos de muestra, respuestas correctoras, reactivos de transición en puntos de verificación y reactivos que no permiten la retroalimentación del examinador.

La UNIT (de Riverside Publishing) es adecuada para individuos con impedimentos de habla, lenguaje o audición, así como para aquellos que no son comunicativos verbalmente o provienen de distintos contextos culturales o lingüísticos. Los materiales de la prueba se seleccionaron considerando que fueran relativamente independientes de las culturas particulares y de interés para niños con distintos ambientes culturales.

Hay seis subpruebas en la Batería Ampliada de la UNIT: Memoria Simbólica, Memoria de Objeto, Memoria Espacial, Razonamiento Analógico, Diseño de Cubos y Laberintos. Las puntuaciones crudas de estas subpruebas se convierten en calificaciones escaladas con una media de 10 y desviación estándar de 3. También se definen cinco cocientes, Cociente de Inteligencia de Escala Completa (FSIQ), Cociente de Memoria (MQ), Cociente de Razonamiento (RQ), Cociente Simbólico (SQ) y Cociente No Simbólico (NSQ), a partir de la combinación de las calificaciones obtenidas en seis subpruebas de la Batería Ampliada o en cuatro subpruebas de la Batería Estándar. La Batería Ampliada toma 45 minutos en su administración, mientras que la Batería Estándar sólo 30 minutos. Una Batería Abreviada de dos subpruebas, que puede usarse para seleccionar, tarda de 10 a 15 minutos en completarse.

La UNIT se estandarizó hacia mediados de la década de 1990 en una muestra nacional de 2100 niños y adolescentes (de entre 5 años y 17 años 11 meses de edad). Los datos de confiabilidad son satisfactorios, y la evidencia de investigación que corresponde a la validez concurrente, predictiva y discriminativa de este instrumento se proporciona en el manual.

TESTS DE INTELIGENCIA COLECTIVOS

Durante la segunda década del siglo xx, Lewis Terman impartía habitualmente un curso en la Universidad de Stanford sobre la Escala de Inteligencia de Stanford-Binet. Según se informó, en una sección de este curso un estudiante, Arthur Otis, tuvo la idea de adaptar tareas seleccionadas de la Stanford-Binet a un formato de lápiz y papel. Poco después, muchas de las tareas adaptadas por Otis y otros autores se combinaron como la primera prueba de inteligencia colectiva, el Examen Alfa del Ejército.

Los exámenes Alfa y Beta del Ejército, una prueba no verbal lingüística para no angloparlantes y analfabetos, se administraron a casi dos millones de reclutas del ejército estadounidense durante y después de la Primera Guerra Mundial con propósitos de selección militar y clasificación de puestos. El Examen Alfa del ejército consistía en reactivos que incluían analogías, problemas aritméticos, completamiento de series de números, sinónimos y antónimos, análisis de cubos, símbolos en dígitos, información y juicio práctico. Esto propició la aparición de

otras pruebas colectivas de inteligencia y de aptitudes académicas, y funcionó como su modelo después de la guerra. Arthur Otis y otros psicólogos empezaron a publicar sus propias pruebas de inteligencia colectivas, y hacia la década de 1930 había disponibles comercialmente muchos más instrumentos de este tipo.

Una prueba colectiva grupal de inteligencia típica puede constar de un conjunto de preguntas de opción múltiple dispuestas en un formato en espiral-ómnibus o de una serie de subpruebas en momentos separados. En el *formato colectivo en espiral* se mezclan los diversos tipos de reactivos que comprende la prueba y se ordenan por dificultad creciente; los reactivos con el mismo grado de dificultad se agrupan juntos.

Aplicación, calificación e informes

Las pruebas de inteligencia colectivas pueden administrarse a pequeñas cantidades de niños desde los 5 o 6 años de edad o a grupos más numerosos de adultos. Al evaluar niños pequeños, los examinadores deben tener particular cuidado en asegurarse de que los examinados comprendan las instrucciones, pasen a la página correcta, comiencen y terminen a tiempo, entre otros aspectos. Al calificar pruebas de inteligencia colectivas, las puntuaciones crudas, ya sea parciales o globales, pueden convertirse en rangos percentilares, calificaciones estándar y otras unidades numéricas.

Incluso más que en pruebas individuales, las calificaciones de pruebas aplicadas de manera colectiva deben interpretarse con precaución, tomando en cuenta otros datos (grados escolares e información obtenida mediante entrevistas o la observación) sobre el examinado. El informe 7.1 que aparece en la página siguiente ilustra la manera en que los hallazgos de una prueba colectiva grupal de inteligencia pueden registrarse e interpretarse, junto con más información relevante sobre el examinado. Asimismo, pueden prepararse perfiles de calificaciones interpretativos a través de un servicio de calificación de pruebas. Los examinados con calificaciones muy bajas deben ser sometidos a otras pruebas, de preferencia individuales, antes de tomar decisiones sobre diagnóstico o colocación.

Ejemplos de pruebas de inteligencia colectivas

Tres de las pruebas de inteligencia colectivas más populares son la Prueba Otis-Lennon de Habilidad Escolar, la Prueba de Aptitudes Cognoscitivas y la Prueba de Personal Wonderlic.

Prueba Otis-Lennon de Habilidad Escolar. Esta prueba (de Harcourt Brace) es una revisión de las Pruebas Autoadministrables de Otis de Habilidad Mental (OLSAT), la Prueba Otis-Lennon de Habilidad Mental y las Pruebas Otis de Habilidad Mental de Calificación Rápida. Igual que sus predecesoras, la séptima edición de la OLSAT consiste en diversos reactivos de imágenes, verbales, de figuras y cuantitativos, a fin de medir Comprensión Verbal, Razonamiento Verbal, Razonamiento de Imágenes, Razonamiento de Figuras y Razonamiento Cuantitativo, desde la etapa preescolar hasta el 12° grado. Hay dos formas y siete niveles de la OLSAT, cada una de las cuales puede administrarse en 60 o 75 minutos. Las normas, que se basan en una muestra nacional amplia, se expresan como rangos percentilares, calificaciones estandarizadas y NCE por grado. De igual manera pueden realizarse comparaciones entre la habilidad y el logro cuando se aplica la OLSAT con la Serie de Pruebas de Aprovechamiento de Stanford, en su novena edición.

Prueba de Habilidades Cognoscitivas. La quinta edición de la Prueba de Habilidades Cognoscitivas (CogAT) (de Riverside Publishing) evalúa las habilidades de los niños para razonar y resolver problemas usando símbolos verbales, cuantitativos y espaciales (no verbales). La CogAT es una prueba de niveles múltiples, con los niveles 1 y 2 para los grados K-3 y niveles de la A a la H para los grados 3-12; su administración dura aproximadamente 90 minutos. Cada ni-

REPORTE 7.1 Resultados de una prueba colectiva de inteligencia

Nombre del examinado: Jane N. Brown
Fecha de nacimiento: 11 de marzo de 1980
Dirección: 12449 Mount Olive Street
Thousand Oaks, CA

Sexo: Femenino
Edad: 21 años, 11 meses
Escolaridad: Licenciatura universitaria
Fecha de aplicación: 15 de abril de 2002

Prueba aplicada: Prueba Otis-Lennon de Habilidad Escolar, Forma avanzada R

Jane Brown, una joven de altura y peso promedios (1.65 m, 60 kg), se ofreció como voluntaria para someterse a la prueba de inteligencia debido a un interés personal en sus habilidades mentales y como un favor hacia el examinador. La prueba se administró como una tarea en Psicología 405 (Evaluación Psicológica) en Western College durante el semestre de primavera de 2002.

En la época del examen, Jane estaba en su último semestre de la especialización de contabilidad. Informó que su promedio de graduación era de 3.2, y señaló que le gustaría ir a la escuela de posgrado en administración para obtener un grado de maestría, pero que inmediatamente después de su graduación planeaba trabajar de tiempo completo en una empresa de contabilidad en el área de Los Ángeles.

El padre de Jane tiene un grado universitario, y su madre terminó dos años de educación universitaria. Ambos trabajan en el negocio familiar, una empresa de asesoría fiscal. Jane informa haber obtenido buenas calificaciones (B y A) en toda su educación escolar, pero confesó que "¡No soy ninguna académica!" Parece ser muy práctica en cuanto a sus intereses, como lo indica no sólo la licenciatura que eligió, sino también sus planes y otros comentarios que hizo al examinador.

Jane mostró un interés moderado en las preguntas de la prueba y se mostró relajada pero involucrada durante el proceso de evaluación. Trabajó con atención y sin interrupciones durante todos los 40 minutos. Las condiciones de la prueba fueron buenas; no hubo interrupciones ni distracciones.

Resultados e interpretación de la prueba

Jane terminó todas las preguntas de la prueba durante el tiempo estipulado (40 minutos). Obtuvo las siguientes calificaciones en la prueba Otis-Lennon:

Puntuación cruda = 65
Índice de habilidad escolar = 116
Rango percentilar (grupo de 18+ años de edad) = 84
Estanina = (grupo de 18+ años de edad) = 7

Estas calificaciones corresponden aproximadamente al promedio para los estudiantes que han completado la licenciatura universitaria, e indican una habilidad intelectual general en el rango del "Promedio superior" para la población general. Un breve análisis de los 16 reactivos que contestó Jane en forma incorrecta indica que en cierta medida tiene mayor dificultad con el razonamiento no verbal que con el verbal. Sin embargo, no hubo ningún patrón significativo en los errores que cometió; en general fueron bastante aleatorios.

Conclusiones y recomendaciones

En una entrevista posterior al examen, Jane señaló que había realizado su mejor esfuerzo en la prueba y que no tenía prisa por terminar a tiempo. Acabó la prueba en 35 minutos y dedicó los restantes 5 a verificar sus respuestas. Afirmó que el Índice de habilidad escolar, que el examinador le comunicó, se encontraba dentro del rango de 5 puntos de una calificación de CI que obtuvo en una prueba de inteligencia que había realizado en la preparatoria. No pudo recordar el nombre de dicha prueba.

Tomando en cuenta las condiciones de la evaluación, el comportamiento que se observó en el examinando y sus afirmaciones después de la prueba, los resultados se consideran válidos en este momento. Los planes y aspiraciones profesionales de Jane parecen adecuados a su habilidad intelectual, aunque tal vez tendrá que esforzarse con diligencia para obtener una maestría en alguna institución de prestigio.

Laura F. Green

Laura F. Green
Pasante de la Licenciatura en Psicología
Western College

vel contiene una Batería Verbal, una Batería Cuantitativa y una Batería No Verbal que consisten en dos o tres subpruebas. Las calificaciones separadas obtenidas en las tres baterías y una calificación compuesta general pueden convertirse a diversos tipos de calificación normalizadas (calificaciones de edad estándar, rangos percentilares de grado y edad nacionales, calificaciones estatinas de grado y edad, y equivalentes de curva normal) con base en una estandarización nacional llevada a cabo en 1992.

Prueba de Personal Wonderlic. La Prueba de Personal Wonderlic (de Wonderlic) es un instrumento breve (12 minutos) de 50 reactivos basada originalmente en la Prueba Autoaplicable de Otis de Habilidad Mental. Las preguntas de la Wonderlic, cuyos ejemplos se muestran en la figura 7.3, consisten en analogías, definiciones, problemas lógicos y aritméticos, relaciones espaciales, comparaciones entre palabras y ubicación de dirección. Esta prueba se ha usado ampliamente como herramienta de selección en situaciones laborales durante muchos años, y la

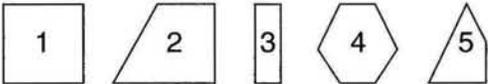
**Observe la lista de números que se presenta a continuación.
¿Qué número debe seguir?**

8 4 2 1 1/2 1/4 ?

**Suponiendo que las dos primeras afirmaciones son ciertas, ¿la última de ellas es:
(1) cierta, (2) falsa, (3) incierta?**

El niño juega béisbol. Todos los jugadores de béisbol usan sombrero.
El niño usa sombrero.

Una de las siguientes figuras numeradas es la que más se diferencia de las otras. ¿Qué número tiene dicha figura?



Un tren recorre 60 metros en 1/5 de segundo. A la misma velocidad, ¿qué distancia recorrerá en tres segundos?

¿Cuántos de los seis pares de cifras de la siguiente lista son duplicados exactos?

3421	1243
21212	21212
558956	558956
10120210	10120210
612986896	612986896
356471201	356571201

Las horas de luz diurna y de oscuridad en SEPTIEMBRE son más similares a las horas de luz diurna y oscuridad de:

(1) Junio (2) Marzo (3) Mayo (4) Noviembre

FIGURA 7.3 Muestra de reactivos de la Prueba de Personal Wonderlic.

(Reproducida con autorización de Wonderlic Personnel Test, Inc., Libertyville, IL.)

investigación indica que es un dispositivo justo y válido para la selección en un amplio rango de puestos. A pesar de la brevedad de la prueba Wonderlic, hay registros de que sus coeficientes de confiabilidad y sus correlaciones con calificaciones de otras medidas de inteligencia llegan a .90.

Pruebas de inteligencia colectivas grupales no verbales y justas para las culturas

Las pruebas de ejecución aplicables individualmente y diseñadas como medidas de las habilidades intelectuales de personas con desventajas lingüísticas o culturales ya se trataron en este capítulo. Ahora veamos cómo también se han elaborado instrumentos complementarios que pueden administrarse en forma colectiva para evaluar la inteligencia de individuos con desventajas físicas o culturales. El antecesor de estas pruebas no verbales fue el Examen Army Beta del Ejército aplicado a los reclutas estadounidenses de la Primera Guerra Mundial, el cual incluía tareas como análisis de cubos, símbolos en dígitos, construcciones geométricas, laberintos y completamiento de imágenes. Esta prueba también resultó útil para evaluar a trabajadores civiles no capacitados y fue actualizada, reestandarizada y reeditada, por The Psychological Corporation, en 1978 como Examen Beta Revisado, segunda edición, y de nuevo en 1999 como Beta III.

Test de Dibujo de Goodenough-Harris. Otra prueba no verbal adecuada para su administración colectiva (o individual) es el Test de Dibujo Goodenough-Harris (de The Psychological Corporation). A diferencia de Beta III, que es una prueba de tareas múltiples, la Goodenough-Harris sólo requiere que el examinando realice dibujos de un hombre, una mujer y de sí mismo. Más que calificarse por su mérito artístico, los dibujos se evalúan comparándolos con los doce dibujos modelo y por la presencia de 73 características específicas (por ejemplo, detalles corporales y de vestimenta, proporcionalidad de cabeza y tronco). La prueba no tiene límite de tiempo, pero suele durar entre 10 y 15 minutos. Las normas para niños de entre 3 y 15 años de edad se registran como calificaciones y rangos percentilares, en forma separada por sexo. También es interesante un sistema de calificación cuantitativa, el Dibuja una Persona: QSS, desarrollado por J. A. Naglieri que hace más objetiva la calificación del dibujo de personas.

Pruebas justas para las culturas. Durante muchos años, los diseñadores de pruebas de inteligencia han sido atacados por la crítica de que estos instrumentos están repletos de sesgos culturales de la sociedad occidental de clase media. Goodenough y Harris tenían la esperanza de que su prueba mediría la inteligencia básica relativamente al margen de influencias culturales, pero ha quedado claro que la tarea de dibujar una figura humana está considerablemente alterada por las experiencias socioculturales específicas. Ha habido varios intentos relevantes por elaborar una prueba de inteligencia independiente de la cultura, pero esos esfuerzos no han tenido éxito alguno. Por consiguiente, el objetivo se modificó después por el de desarrollar una prueba de inteligencia que resultara justa para las culturas. Al diseñar una *prueba de inteligencia justa para las culturas*, se intenta usar sólo reactivos relacionados con experiencias comunes a un amplio espectro de culturas. Se omiten reactivos que incluyan construcciones lingüísticas particulares y otras tareas embebidas de cultura, tales como la velocidad de respuesta. En este sentido, el test de Goodenough-Harris es culturalmente justo. Otras pruebas muy utilizadas que probablemente también están cerca de resultar justas para la cultura son las Matrices Progresivas de Raven y la prueba de Inteligencia Justa para la Cultura.

Matrices Progresivas de Raven. Esta prueba, que puede administrarse ya sea en forma individual o colectiva, demanda al examinado indicar cuál de diversas figuras o diseños pertenece a

una matriz dada. Desarrollada en Gran Bretaña como una medida del factor de inteligencia general de Spearman, la prueba de Raven está disponible en The Psychological Corporation en formas de matrices progresivas Estándar, Coloreada y Avanzada. La Forma Estándar, para edades de 6 a 80 años, incluye cinco conjuntos en blanco y negro de 12 problemas cada uno y se termina en un lapso de 20 a 45 minutos. La Forma Coloreada, para niños de 5 a 11 años, individuos de edad avanzada y personas mental y físicamente impedidas, toma de 15 a 30 minutos en terminarse. La Forma Avanzada tiene un rango de los 11 años a la edad adulta y dura entre 40 y 60 minutos. Las normas más recientes, basadas en muestras británicas y estadounidenses, están disponibles en la Forma Avanzada, pero las tres formas requieren de reestandarización.

Una prueba similar, pero más reciente que las Matrices Progresivas de Raven, es la Prueba de Analogías de Matriz-Forma Ampliada. Consiste en reactivos de razonamiento no verbal en cuatro categorías: Completamiento de Patrones, Razonamiento por Analogía, Razonamiento en Serie y Visualización Espacial. Los examinados (de entre 5 y 17 años de edad) son evaluados en un lapso de 20 a 25 minutos con 64 diseños abstractos del tipo de la matriz progresiva estándar, con un diseño por página. Las normas están basadas en una muestra representativa numerosa de individuos de 5 a 17 años de edad, residentes en Estados Unidos. Las puntuaciones crudas se convierten en calificaciones estándar, rangos percentilares y estandines por intervalos de medio año y en equivalentes de edad de los 5 años a los 17 años 11 meses. The Psychological Corporation también tiene disponible una Prueba de Analogías de Matriz-Forma Abreviada, que consiste en 34 reactivos.

Pruebas de Inteligencia Justas para las Culturas. Estas pruebas (de IPAT) están compuestas por tres escalas: la Escala 1, para niños de 4 a 8 años de edad y adultos retrasados mentales; la Escala 2, para niños entre 8 y 14 años y adultos de inteligencia promedio, y la Escala 3 para estudiantes universitarios, ejecutivos y otras personas de inteligencia superior al promedio. Cada escala contiene cuatro subpruebas (Series, Clasificaciones, Matrices y Condiciones) para medir la habilidad para percibir relaciones. Además de estas cuatro subpruebas justas para las culturas, la Escala 1 contiene cuatro subpruebas para evaluar información cultural y comprensión verbal. La Escala 1 no tiene límite de tiempo, pero toma alrededor de 22 minutos resolverla; las escalas 2 y 3 se llevan 12½ minutos cada una.

Prueba Naglieri de Habilidad No Verbal. La Prueba Naglieri de Habilidad No Verbal-Forma Multinivel (NNAT) (The Psychological Corporation) es similar a la de Raven en cuanto a su diseño de matrices. En la figura 7.4 se presentan ejemplos de los reactivos de la NNAT. El objetivo de esta prueba, así como el de otras pruebas no verbales, es proporcionar una medida no sesgada de la habilidad mental general de individuos con habilidades limitadas para la lengua inglesa o con otros problemas de aprendizaje. La NNAT-Forma Multinivel es adecuada para alumnos desde preescolar hasta el 12° grado, y tarda aproximadamente 30 minutos en completarse. Se proporcionan las Calificaciones del Índice de Habilidad No Verbal y otras calificaciones establecidas basadas en una muestra de estandarización de más de cien mil alumnos.

Instrumentos tales como las Matrices Progresivas de Raven, las Pruebas de Inteligencia Justas para las Culturas, la Prueba Naglieri de Habilidad No Verbal y la Prueba de Inteligencia No Verbal Universal, representan esfuerzos encomiables por elaborar pruebas en que los distintos grupos culturales obtengan calificaciones iguales. No obstante, ahora se reconoce que tal vez sea imposible construir una prueba que mida las habilidades cognoscitivas independientemente de la experiencia. En cualquier caso, los resultados de la investigación realizada en países en vías de desarrollo indica que las diferencias en cuanto a los índices generales de alfabetismo y escolaridad son más importantes que la lengua, el país, la raza o la etnia para determinar diferencias “culturales” en las calificaciones de pruebas de inteligencia (vea Frisby, 1999).

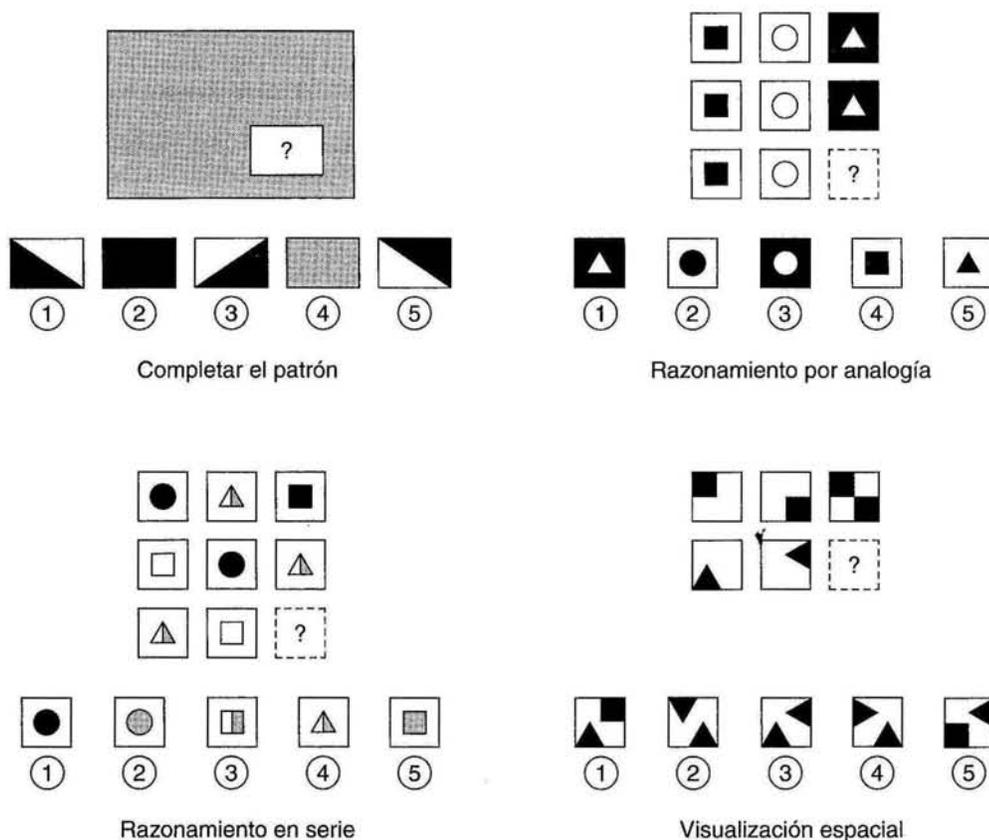


FIGURA 7.4 Ejemplos de reactivos de la Prueba Naglieri de Habilidad No Verbal-Forma Multinivel.

(Derechos Reservados © 1996 por Harcourt, Inc. Reproducido con autorización. Reservados todos los derechos, "Naglieri Nonverbal Ability Test" y "NNAT" son marcas registradas propiedad de The Psychological Corporation e inscritas en Estados Unidos de Norteamérica y otras jurisdicciones.)

Como se observa en la Prueba de Inteligencia No Verbal Universal, continúan los esfuerzos por desarrollar pruebas de habilidades cognitivas que sean justas para personas de distintas culturas, pero ello no significa que muchas antiguas pruebas de inteligencia muy sesgadas culturalmente deban abandonarse. Es digno de mención que en otros países el mercado de pruebas de inteligencia tradicionales es mucho mayor que el de pruebas justas para las culturas (Oakland y Hu, 1993). ¡Aparentemente, las personas de países no occidentales se preocupan menos que los estadounidenses de clase media por lo justo para las culturas que resulten las pruebas tradicionales de inteligencia tipo Binet!

Pruebas de aptitud académica y de admisión

Muchas pruebas de inteligencia colectivas se han diseñado específicamente con el propósito de medir la aptitud para el trabajo académico y se conocen como *pruebas de aptitud académica*. Algunas pruebas de inteligencia en grupo tienen un enfoque más amplio que éste, pero aun así su contenido es similar al de las medidas de habilidad académica: tienen un gran contenido de reactivos verbales, numéricos y otros de tipo escolar.

A lo largo del tiempo se han usado muchas pruebas distintas con propósitos de admisión a las universidades, incluyendo el Examen Psicológico del Consejo Estadounidense sobre Educación

(ACE), las Pruebas de Habilidad Universitaria y Escolar (SCAT), la Prueba de Habilidad Académica del Consejo de Exámenes de Ingreso a la Universidad (ahora denominada Prueba de Evaluación Académica, o SAT), y la Evaluación del Programa de Pruebas Universitarias Estadounidenses (ACT). Debido a su extenso uso, se describirán con cierto detalle las últimas dos de estas baterías.

Prueba de Evaluación Académica (SAT). Antes de 1994, la SAT, anteriormente denominada Prueba de Habilidad Académica, consistía en dos secciones que producían dos calificaciones: Verbal (SAT-V) y Matemática (SAT-M). La sección verbal estaba compuesta por reactivos de analogías verbales, antónimos, información, comprensión de lectura y completamiento de enunciados; la sección matemática consistía en reactivos de aritmética, álgebra, geometría, cuadros y gráficas y razonamiento lógico. Ambas secciones se calificaron en una escala estándar con una media de 500 y desviación estándar de 100, con las calificaciones en un rango de 200 a 800. Aunque cada año se desarrollaron versiones nuevas de la SAT, las calificaciones de cada nueva forma se escalaron hacia el grupo de estandarización de 1941. Este grupo estuvo constituido por diez mil alumnos del noreste de Estados Unidos, en su mayoría varones de raza blanca y con nivel de ingresos alto, los cuales estaban solicitando su admisión a las escuelas de la Ivy League. Como es comprensible, los estudiantes de preparatoria de principios de la década de 1990 obtuvieron calificaciones algo inferiores a la media de 500 lograda por este grupo.

La versión actual de la SAT, que se administró primero a nivel nacional en marzo de 1994, está compuesta por dos partes, SAT I: Razonamiento, y SAT II: Pruebas de Materia. SAT I consiste en secciones de Razonamiento Verbal y Razonamiento Matemático con una duración de 75 minutos cada una. La sección de Razonamiento consta de 78 reactivos de opciones múltiples en Analogías, Completamiento de Enunciados y Lectura Crítica. La sección de Razonamiento Matemático está formada por 60 reactivos en Matemáticas Regulares, Comparaciones Cuantitativas y Respuestas Producidas por el Alumno. Se pide a los examinados que lleven al examen su propia calculadora de bolsillo, de modo que puedan calcular las respuestas de las subpruebas matemáticas.

Así como en versiones anteriores de la SAT, las puntuaciones crudas de las Pruebas de Razonamiento se convierten a una escala de calificación estándar que tiene una media de 500 y una desviación estándar de 100. La calificación de la versión revisada de la SAT, la Prueba de Evaluación Académica, se basa en el desempeño de más de un millón de estudiantes que se sometieron a la prueba en 1994. Las calificaciones se *recalaron* para reflejar la población estudiantil mayor y más diversa de la actualidad, lo que dio como resultado un aumento de la calificación promedio de Razonamiento Verbal de aproximadamente 80 puntos, y de la calificación promedio del Razonamiento Matemático en alrededor de 20 puntos. Además de las calificaciones estándar en las Pruebas de Razonamiento Verbal y Matemático, un informe de calificaciones de SAT da puntuaciones crudas y rangos percentilares para cada subprueba, rangos de calificaciones basadas en error estándar de medida de las pruebas y equivalentes de percentiles nacionales y estatales para estudiantes universitarios del último año. Los resultados de múltiples estudios indican que la SAT-I es un predictor válido para el desempeño en la universidad, específicamente de los promedios de grado durante el primer semestre universitario, pero también predice con eficacia los promedios posteriores y el desempeño en otros exámenes académicos.

Las 20 Pruebas de Materia SAT pertenecen a cinco áreas generales: Inglés, Historia y Estudios Sociales, Matemáticas, Ciencias y Lenguas. Se obtiene una muestra directa de las de redacción del examinado, y también se administran preguntas de opción múltiple sobre inglés escrito, dicción y expresión lógica. Así como sucede con las calificaciones de la SAT-I, las calificaciones de la SAT-II se registran en una escala de calificación estándar con una media de 500 y desviación estándar de 100.

Pruebas Universitarias Estadounidenses. El segundo examen de admisión a la universidad más usado es el constituido por las Pruebas Universitarias Estadounidenses (ACT), el cual se aplica cinco veces al año tanto en Estados Unidos como en otros países. Hay cuatro subpruebas en las ACT: Inglés, Matemáticas, Lectura y Razonamiento Científico. A quienes se someten a este examen se les entregan calificaciones de las cuatro subpruebas, así como una calificación compuesta (el promedio de las cuatro subpruebas redondeado al entero más cercano) y siete subcalificaciones. Las calificaciones compuestas y las de las subpruebas van de 1 a 36, con una media de 18; las siete subcalificaciones están entre 1 y 18, con una media de 9. Las confiabilidades de las cuatro subpruebas van desde .78 para Razonamiento Científico hasta .91 para Inglés, con coeficientes de consistencia interna algo más elevados que los coeficientes de formas paralelas. Como podría esperarse debido a su extensión más corta, las confiabilidades de las subcalificaciones son inferiores a las de las subpruebas, y están entre .67 para Geometría Plana/Trigonometría y .85 para Uso/Mecánica del Inglés.

Exámenes del Registro de Graduados. La prueba más popular para admisión en una escuela de posgrado es el Examen del Registro de Graduados (GRE). Consiste en una Prueba General que mide la aptitud para el trabajo de posgrado y una serie de Pruebas de Materia que miden el aprovechamiento en una materia en particular. La Prueba General contiene tres secciones calificadas: una sección Verbal (V) de 30 minutos con 30 preguntas, una sección Cuantitativa de 45 minutos con 28 preguntas, y una sección Analítica (A) de 60 minutos con 35 preguntas. Los reactivos de la sección Verbal consisten en analogías, antónimos, completamiento de enunciados y comprensión de lectura. Los reactivos de la sección Cuantitativa incluyen comparación cuantitativa, cuantitativa discreta y problemas de interpretación de datos. Los reactivos de las pruebas analíticas constan de ejercicios de razonamiento analítico y de razonamiento lógico. La Prueba General produce calificaciones separadas: Verbal (GRE-V), Cuantitativa (GRE-Q), y Análisis (GRE-A), con la misma escala de calificación estándar que la SAT. Las Pruebas de Materia GRE son exámenes de tres horas sobre materias en particular (vea el capítulo 6).

RESUMEN

Las definiciones y teorías de *inteligencia*, un antiguo término latino reintroducido durante el siglo XIX, se dieron en abundancia en el siglo pasado. Entre las teorías más prominentes sobre la inteligencia figuran las relacionadas con el enfoque psicométrico (Spearman, Thurstone, Guilford, Vernon, Cattell), el enfoque del desarrollo (Piaget), y el enfoque del procesamiento de información (Sternberg, Gardner, Das y Naglieri).

Varias pruebas sensoriomotrices se usaron en los primeros intentos por evaluar la inteligencia, pero la primera prueba práctica de inteligencia fue elaborada por Alfred Binet y Théodore Simon durante la primera década del siglo XX. La Escala Binet-Simon, una serie de tareas relacionadas con la escuela y ordenadas por dificultad creciente, proporciona una calificación de edad mental para cada examinando. Entre las múltiples traducciones y revisiones de la Escala Binet-Simon, la más popular fue la Escala de Inteligencia Stanford-Binet, que fue publicada por primera vez en 1916 y revisada en 1937 y 1960, y cuyo autor fue Lewis Terman. La prueba producía un valor llamado razón de CI, definida como

$$CI = 100 \left(\frac{MA}{CA} \right)$$

aunque una desviación del CI podía también calcularse en la revisión de 1960.

La cuarta edición de la Escala Stanford-Binet representó un considerable alejamiento de las ediciones previas. La elaboración de la cuarta edición, que incluía una teoría y una metodología psicométrica más complejas, proporcionaba calificaciones separadas en 15 pruebas y cuatro áreas, así como una calificación compuesta. El énfasis al diseñar la cuarta edición residía no sólo en identificar el retraso mental, sino también en proporcionar información para diagnosticar causas específicas de problemas de aprendizaje.

Durante muchos años, las primeras ediciones de la Escala Stanford-Binet funcionaron como un patrón contra el cual se comparaban otras pruebas de inteligencia. Sin embargo, en la segunda mitad del siglo pasado, las escalas de inteligencia de Wechsler se volvieron más populares. A diferencia de las subpruebas de la Stanford-Binet, que se agrupan de acuerdo con niveles de edad, las subpruebas de las escalas Wechsler se dividen en aproximadamente diez categorías de acuerdo con su contenido. Asimismo, las calificaciones de las escalas Wechsler proporcionan tres tipos de CI de desviación: Verbal, de Ejecución y de Escala Completa. Sumado a los tres CI, el patrón de calificaciones escaladas de las subpruebas en las pruebas de Wechsler puede proporcionar información clínica útil para el diagnóstico de ciertas características y trastornos de la personalidad.

Entre otras pruebas de inteligencia de rango amplio se encuentran las Escalas de Habilidad Diferencial (DAS), la Prueba Detroit de Habilidad para el Aprendizaje (DTLA), la Batería de Evaluación para Niños de Kaufman (K-ABC), las Pruebas de Habilidades Cognoscitivas de Woodcock-Johnson III (WJ III) y el Sistema de Evaluación Cognoscitiva de Das-Naglieri (CAS). Las últimas tres merecen especial mención por sus bases en la teoría psicométrica y neuropsicológica.

Como representantes de las múltiples pruebas de inteligencia con fines específicos que se aplican individualmente figuran las pruebas pictóricas como la Escala de Madurez Mental de Columbia y otras pruebas de tarea única como los Laberintos de Porteus y los Diseños de Cubos de Kohs. Las pruebas de inteligencia no verbales de aplicación individual, diseñadas para personas con desventajas lingüísticas o físicas incluyen las baterías de pruebas de ejecución como las Pruebas Hiskey-Nebraska de Habilidad de Aprendizaje, la Escala Leiter de Desempeño Internacional, la Prueba Comprensiva de Inteligencia No Verbal y la Prueba Universal de Inteligencia No Verbal.

De uso más extenso que las pruebas de inteligencia individuales en escuelas y otras organizaciones, son las pruebas de inteligencia de aplicación colectiva. Estas pruebas provienen de los Exámenes Army Alfa y Army Beta del Ejército, que se basan en el trabajo pionero de Arthur Otis y otros psicólogos durante la Primera Guerra Mundial. Como ejemplos de las pruebas de inteligencia colectivas actuales están disponibles también la Prueba de Habilidad Escolar de Otis-Lennon, la Prueba de Habilidades Cognoscitivas y la Prueba de Personal Wonderlic, así como las pruebas en grupo supuestamente justas para las culturas, tales como la Prueba de Dibujo de Goodenough-Harris, las Matrices Progresivas de Raven, las Pruebas de Inteligencia Justas para las Culturas y la Prueba de Habilidad No Verbal de Naglieri.

Algunas pruebas de habilidad académica, como la Prueba de Evaluación Académica (SAT), las Pruebas Universitarias Estadounidenses (ACT) y los Exámenes del Registro de Graduados (GRE), en particular, se usan ampliamente para el ingreso a universidades e instituciones profesionales.

PREGUNTAS Y ACTIVIDADES

1. Elija una de las teorías sobre la inteligencia examinadas en este capítulo para efectuar un análisis más profundo y escriba un breve ensayo sobre su investigación.

2. ¿Cuál es la razón CI para un niño de 8 años 9 meses si su calificación en la Escala de Inteligencia Stanford-Binet es igual a la edad mental de 6 años 5 meses?
3. ¿Por qué las calificaciones del CI de desviación se consideran psicométricamente superiores a las de la razón CI?
4. Explique el desarrollo de la Escala de Inteligencia de Stanford-Binet desde las primeras pruebas de Binet hasta la cuarta edición de la escala.
5. Enumere y describa las ediciones actuales de la serie de pruebas de inteligencia de Wechsler, incluyendo el rango de edad apropiado para cada una así como las subpruebas que abarcan.
6. Compare las escalas Wechsler con las ediciones anteriores y recientes de la escala Stanford-Binet en términos de rango de edad, tipos de habilidades medidas, justicia de las pruebas para las personas con desventajas físicas o culturales, y otras características relevantes.
7. ¿Qué pruebas de inteligencia recomendaría para administrar en cada uno de los siguientes individuos? (a) Un niño de 5 años de edad en quien se sospecha retraso mental; (b) un grupo de aborígenes de las Islas del Sur; (c) un niño de 10 años con parálisis cerebral; (d) un adulto normal angloparlante; (e) un niño de 7 años totalmente ciego; (f) un adulto esquizofrénico, y (g) un grupo de alumnos de la escuela elemental con desventajas culturales.
8. Elija una de las siguientes categorías de pruebas de inteligencia analizadas en este capítulo y un instrumento publicado que sea representativo de esa categoría: pruebas pictóricas individuales; escalas de desarrollo para niños pequeños; pruebas de inteligencia colectivas grupales de multinivel; pruebas de inteligencia colectivas grupales no verbales. Obtenga tanta información como le sea posible sobre dos pruebas de los libros de texto sobre evaluación *The Mental Measurements Yearbooks, Tests, Test Critiques* y otras fuentes (consulte *The Psychological Abstracts* y *Education Index* en particular). Escriba una reseña comparativa de esas pruebas centrándose en el diseño y el formato, los procedimientos de administración y calificación, las normas, la confiabilidad, la validez y la investigación relacionada (vea la pregunta 8 de la sección de Preguntas y Actividades en el capítulo 6). Deduzca conclusiones apropiadas en cuanto a los méritos relativos de las dos pruebas que haya seleccionado.
9. ¿Cuál es la diferencia entre una prueba independiente de la cultura y otra justa para las culturas? ¿Es posible desarrollar una prueba de alguno de estos dos tipos y, de ser así, para qué se usaría?