

ROCIO FERNANDEZ BALLESTEROS

*Catedrático de Psicodiagnóstico
Universidad Autónoma de Madrid*

PSICODIAGNOSTICO
Concepto y Metodología

**cin
cel**
editorial cincel

Las técnicas

Como ya se ha dicho, una gran parte de la bibliografía sobre Psicodiagnóstico da cumplida referencia de lo que podríamos llamar «tecnología» de nuestra disciplina. Por otra parte, el análisis de los programas de Psicodiagnóstico impartidos en gran parte de nuestras universidades (a lo que ya se hizo referencia en la Primera Parte de este trabajo) dedican más de sus tres cuartas partes al tratamiento de las técnicas diagnósticas y evaluativas. Así pues, parece que el «grueso» de la metodología psicodiagnóstica se centra en el estudio y pormenorizada descripción de los tests y técnicas de evaluación y diagnóstico psicológico.

Con todo esto, es fácil suponer la dificultad de dar cuenta de todas ellas, entre otras razones, porque en los objetivos de este trabajo figuraba dejar de lado tal tecnología, por considerar que en castellano existía ya abundante bibliografía al respecto. Así pues, tan sólo vamos a intentar presentar una muestra de las más representativas, encuadradas en distintas categorías.

Nuevamente nos encontramos frente a una dificultad terminológica y clasificatoria. En la mayor parte de los manuales consultados se emplean los términos «tests», «instrumentos» y «técnicas», con una clara, pero no explicitada, equiparación. PELECHANO (1976) ha precisado certeramente las connotaciones de cada uno de ellos. Mientras que «instrumento» o «técnica» deben ser considerados como sinónimos y se refieren a «aquellas pruebas, procedimientos y técnicas, utilizadas tanto en el laboratorio como en el mundo social para la realización de un diagnóstico psicológico, no cuantificadas y no tipificadas» (p. 52), «test» implica «un instrumento sistemático y tipificado que compara la conducta de dos o más personas» (p. 52). Teniendo clara la significación de cada uno de estos términos, conviene ahora intentar una clasificación, lo más amplia y rigurosa posible, que englobe a todos ellos.

Son muchas las posibilidades de agrupación que se nos ofrecen. Podemos seguir ciertos criterios *prácticos* a través de los cuales clasificar las técnicas

de evaluación en razón a los objetivos del psicodiagnóstico (clínico, escolar, industrial, etc.). Podríamos seguir una perspectiva *procesual* en la medida en que las técnicas hayan sido construidas para evaluar determinadas funciones psicológicas (pensamiento, lenguaje, percepción, memoria, emociones, motivaciones, etc.). También podríamos hacerlo en virtud del *material* utilizado en ellas (lápiz y papel, ejecutivo, etc.) o en razón de las personas a las que va *dirigido* (adultos, niños). Un nuevo criterio provendría del *tipo de rendimiento* que se pide al sujeto, por ejemplo, CRONBACH (1971) divide los tests en los de rendimiento máximo y típico. Desde el enfoque conductual se pretende una división atendiendo a la *metodología* utilizada (FERNÁNDEZ BALLESTEROS 1979). Por último, como lo hace FISKE (1971), puede realizarse una clasificación en función de los múltiples criterios posibles, aunque éstos se solapen entre sí.

Hasta ahora, se ha intentado resaltar la importancia que para cualquier disciplina tiene su conexión con un marco referencial teórico y, también, la diversidad que el Psicodiagnóstico presenta a este respecto. Esto hace que intentemos, una vez más, relacionar las técnicas con los enfoques, tanto teóricos como metodológicos, y procurar que la clasificación elegida sea lo más amplia e integradora posible. Así, tratando de relacionar los tests y técnicas con sus respectivos enfoques teóricos y metodológicos, vamos a adoptar, en parte, la clasificación efectuada por PERVIN (1975), tratando de completarla y adaptarla —en la medida de lo posible— a nuestros objetivos.

Las ventajas que nos brinda la clasificación de Pervin son las siguientes: 1) Nos permite encuadrar las distintas técnicas y tests según enfoques teóricos. 2) Posibilita el acercamiento tecnológico a la metodología científica básica. 3) Permite la consideración del material de las técnicas o tests en cuestión según los criterios establecidos por CAMPBELL (1957) (*). 4) Se consigue también una integración de los datos en función de las estrategias utilizadas en su construcción.

No obstante, la clasificación de Pervin no permite una diferenciación ni procesual ni finalista. Es decir, a través de ella será difícil integrar tanto los procesos psicológicos evaluados como los objetivos del Psicodiagnóstico. La omisión de estos aspectos no nos parece tan relevante como la integración teórica y metodológica. Esta es la razón por lo que hemos adoptado dicha clasificación.

Pasamos ahora a la descripción de las técnicas y tests más utilizados en Psicodiagnóstico, encuadrados en cuatro categorías fundamentales: Psicométricas, Proyectivas, Subjetivas y Objetivas.

1. Psicométricas

Podemos englobar en esta categoría, todos aquellos tests de evaluación y diagnóstico que han sido elaborados a través de procedimientos estadísticos altamente sofisticados, con material rigurosamente estandarizado y tipificación en sus tres fases fundamentales: aplicación, corrección e interpreta-

ción. Estos tests son producto de estrategias empíricas y factoriales (KELLY, 1967) y pueden ser considerados como los tests prototípicos.

Dentro de esta categoría, los tests pueden dirigirse tanto a la exploración de funciones cognitivas como emocionales. Es decir, pueden ser incluidos aquí los tests llamados por CRONBACH (1971) de «rendimiento máximo» (tests de aptitudes) así como los de «rendimiento típico» (tests que evalúan determinados atributos de la personalidad). Ambos tipos de instrumentos —desde el punto de vista de las estrategias utilizadas en su construcción— pueden ser considerados dentro de esta categoría que relatamos. Por otra parte, también hay que tener en cuenta que a través de las técnicas psicométricas se están midiendo determinados «rasgos» (aptitudinales, temperamentales, actitudinales, etc.) que implican constructos psicológicos con diferentes niveles de inferencia y que se dirigen a la evaluación del sujeto en su sistema biológico, conductual y social.

Las características materiales de estos tests son de máxima estructuración del contenido y máxima voluntariedad de participación del sujeto en examen.

El avance y la proliferación de los tests adscritos a esta categoría se debe, fundamentalmente, a las elaboraciones estadísticas y más concretamente al análisis factorial. En este sentido y como señala Pervin, podemos decir que las teorías factorialistas, tanto de las aptitudes como de la personalidad, forman el sustrato teórico de este tipo de técnicas evaluativas (Thurstone, Eysenck, Cattell, Guilford).

Por otra parte, otros tests psicométricos proceden de estrategias empíricas. Así, los tests de intereses, como el de Strong o los dedicados a medir la inteligencia, como las Escalas de Wechsler, o los expresamente diseñados para el diagnóstico clínico, como el MMPI. Así también otros tests de evaluación psiconeurológica como la Batería de Halstadt-Reitan o el Test de Retención Visual de Benton. Por último, podrían ser incluidos en esta categoría instrumentos de evaluación de los subsistemas relacionales, tales como los Tests de Clima Social de Moos.

Son probablemente este tipo de técnicas psicodiagnósticas las que mayor auge han tenido durante toda la historia de nuestra disciplina, y en la actualidad serían las que requerirían de una mayor reflexión y balance, no sólo de los criterios en los que se basan o sobre otras condiciones que puedan restarle validez, sino sobre la naturaleza de los propios constructos por ellos evaluados, así como de las habilidades básicas subyacentes. La proliferación de «características» y de técnicas para «medirlas» han llegado a ser tan desmesuradas que una reflexión y autocrítica constructiva se hacen absolutamente necesarias.

2. Proyectivas

El sustrato teórico de las técnicas encuadradas en esta categoría lo está, a su vez, en las teorías dinámicas de la personalidad. El énfasis principal se establece en una visión de la personalidad global, idiosincrásica y en la aceptación de un sustrato inconsciente en el que residen impulsos, tendencias, conflictos, necesidades, etc., todas ellas inferidas del comportamiento de los individuos humanos.

(*) CAMPBELL distingue entre los tests estructurados y no estructurados, enmascarados y no enmascarados.

LINDZEY (1961) define las técnicas proyectivas como «instrumentos que han sido considerados como especialmente sensibles para revelar aspectos inconscientes de la conducta, permiten provocar una amplia variedad de respuestas subjetivas, son altamente multidimensionales y evocan respuestas y datos inusualmente ricos con un mínimo de conocimiento por parte del sujeto del objetivo del Test» (p. 45). La utilización del término «proyectivo» para estos tests procede de FRANK (1939) que les dio este calificativo, ya que, según él, estas técnicas «elicitan la proyección del mundo privado de los sujetos al tener que organizar el campo, interpretar el material y reaccionar frente a él afectivamente» (p. 403). La discusión sobre la idoneidad del apelativo ha sido abundante (ABT 1965, BELLACK 1965) y fundamentalmente se ha producido en base a que con el término «proyectivo» se hace referencia a un mecanismo de defensa psicoanalítico. No es el momento de detenernos en estas cuestiones, puesto que la utilización de este adjetivo para este tipo de técnicas es general y no parece procedente enmendarles la plana a los especialistas del tema.

El material de estas técnicas varía extraordinariamente, aunque todas ellas son enmascaradas y no voluntarias. Puede tratarse de estímulos visuales de mayor o menor estructuración (como en el caso del TAT o el Rorschach), de estímulos verbales (Test de Frases Incompletas o Asociación de Palabras) o de otros tipos de material de ejecución con los que el sujeto debe realizar tareas constructivas o expresivas. Esta diversidad de estímulos requiere una nueva clasificación de las técnicas proyectivas: Las «estructurales» (Ej.: Rorschach, Z-Test), ante las cuales el sujeto ha de decir lo que ve ante estímulos visuales muy poco estructurados. Las técnicas «temáticas» (ej. TAT, Relaciones Objetivas de Phillipson, PN de Corman), que consisten en estímulos visuales con distintos grados de estructuración ante los que el sujeto debe narrar una historia. Las técnicas «expresivas» (Dibujo de la Figura Humana de Machover, Test del Animal de Levy, Test de la Familia de Corman, etc.), en las que se le pide al sujeto que realice un dibujo. Las «constructivas», en las que el sujeto debe construir «algo» con muy diversos materiales (ej.: Test del Pueblo, Juego, etc.). Por último, las «asociativas» (ej.: Asociación de palabras, Frases Incompletas, etc.), en las que el sujeto debe manifestar verbalmente sus asociaciones a palabras, frases, cuentos.

En algunas de estas técnicas, como en el caso del Psicodiagnóstico de Rorschach, se partió en principio de estrategias empíricas para su construcción, pero el desarrollo posterior del instrumento derivó —de lo que fueron objetivos científicos del autor— hacia derroteros fundamentalmente psicodinámicos. Por nuestra parte, pensamos que el Rorschach es una técnica fundamentalmente perceptiva que podría ser incluida, con todo rigor, entre las técnicas Objetivas en el caso de que las respuestas de los sujetos ante sus láminas fueran utilizadas como una «muestra» de conducta perceptiva. La investigación en este sentido se hace totalmente necesaria.

En la mayor parte de estos instrumentos también existe, junto con el sustrato racional-teórico dinámico, un bagaje, más o menos nutrido, de evidencia empírica que, por desgracia y en la mayor parte de los casos, es incongruente.

Centrándonos en algunas de las técnicas más utilizadas, como el TAT y el Rorschach, la discusión actual plantea, tres importantes alternativas. De un lado, si el valor principal de estas técnicas está en los contenidos verbales

que elicitan en el sujeto, por lo que el análisis debería recaer sobre la investigación de tales contenidos (SARASON 1972, ARONOW, REZNIKOFF y RAUCHWAY, 1979). Por otra parte, otros autores piensan que la investigación debe recaer, más bien, en el análisis de los estímulos y su calificación, apoyándose en presupuestos cognitivistas (FOULKERSON 1965, EPSTEIN 1966). En este caso, las respuestas de los sujetos deberían ser analizadas en función de los estímulos. Por último, tal vez la más nutrida fuente de trabajos esté en el intento de presentar evidencia empírica suficiente de tales técnicas (EXNER 1974, 1978) a través de un basamento psicométrico. En definitiva y sin intentar profundizar, en modo alguno, en estas cuestiones, lo que sí está claro es que en los últimos años se está intentando desproveer a gran parte de estas técnicas de sus concatenaciones dinámicas, fundamentalmente por la falta de rigor metodológico de sus planteamientos. Lo que se está produciendo es un número cada vez más crecido de estudios a través de métodos experimentales y correlacionales, tanto del material de las técnicas como de las categorías de respuesta que son posibles.

Las técnicas Proyectivas se dirigen, en líneas generales, a la evaluación de las funciones óréticas y cognitivas implicadas en el Sistema Conductual de nuestro modelo. Por otra parte, en los últimos años se han desarrollado instrumentos dirigidos a la evaluación de algunos de los subsistemas relacionales (por ejemplo el «Group Personality Projective Test» de Cassel y Kahn).

3. Subjetivas

Incluimos en esta categoría técnicas de muy distinta índole. Fundamentalmente nos referimos aquí a instrumentos cuyo material no está estandarizado y su corrección no es objetiva. El sustrato teórico también es vario y oscila entre enfoques fenomenológicos —como en el caso de algunos tipos de entrevista y el análisis de documentos personales— y las teorías cognitivas, como en el caso del REP de KELLY (1955).

Con respecto a las estrategias seguidas en la elaboración de estas técnicas, podemos decir que son, en su gran parte, racionales, bien teóricas o racionales de «sentido común» (KELLY 1967). Entre las primeras podemos citar el REP de KELLY o el Diferencial Semántico de Osgood; entre las segundas se encontraría la entrevista, entre otras.

Pasemos, brevemente a analizar algunos de los más representativos instrumentos encuadrables en este apartado. La entrevista puede ser considerada como la técnica más antigua y de más amplio espectro (medicina, psiquiatría, trabajo e investigación social, psicología, etc.). A pesar de su importancia, esta técnica no es tratada en una gran parte de textos de Psicodiagnóstico. Tal vez esto sea debido a la dificultad de objetivación de sus resultados, la falta de tipificación de su material y lo costoso de lograr una cuantificación de los datos que de ella se obtienen. No obstante, en la actualidad (MITTLER 1973, SARASON 1975, PELECHANO 1976, SUNDBERG 1977), existe la opinión muy generalizada de que representa uno de los principales instrumentos de carácter general en nuestra disciplina, ya que de un lado abarca muy distintos ámbitos del diagnóstico y la investigación psicológica y, de otro, es necesaria —aunque

sea mínimamente— en toda evaluación psicológica (recogida de información personal del sujeto/s).

MATARAZZO y WIENS (1972) definen la entrevista como «una forma de conversación entre dos o más personas, comprendidas en una interacción verbal y no verbal con una finalidad previamente establecida» (p. 32). Tanto si tomamos esta definición como cualquier otra, la entrevista entraña una serie de características: 1) Se refiere a un intercambio cara a cara entre una o más personas en el que una de ellas trata de obtener información de otra u otras. 2) Existe un objetivo previamente establecido (diagnóstico, orientación, selección, investigación, etc.). 3) Existen diferentes roles en este proceso interactivo. 4) Estos roles están en relación con que el entrevistador es el que provoca información del entrevistado o entrevistados. Los psicólogos se han preocupado de hacer distintas clasificaciones de esta técnica (MATARAZZO y WIENS 1972, MACCOBY y MACCOBY 1959, BERSOFF y GRIEGER 1971), que suelen hacer referencia a los distintos niveles de estructuración, a los objetivos de la misma o bien a la posibilidad de manipulación cuantitativa de los resultados. Los últimos trabajos sobre la entrevista (POPER 1979) demuestran cómo se está avanzando de manera importante y, por tanto, parece posible que en el futuro se consigan de ella datos fiables y válidos.

El análisis psicológico de los documentos personales ha sido fundamentalmente enfatizado por los psicólogos evaluadores de un enfoque fenomenológico (ALLPORT 1942, 1961). Algunos psicólogos (WEBB, CAMPBELL, SCHWARTZ y SECHREST 1966, ALLPORT 1971) han hecho especial hincapié en su importancia para la evaluación y diagnóstico psicológico en los últimos años, fundamentalmente como técnica auxiliar de la entrevista y la observación. El material de este tipo de técnica consiste en todo tipo de producciones del sujeto como colecciones personales, diarios, obras de arte por él producidas, cartas, etc. Desde luego los procedimientos de análisis de tales materiales, no están sistematizados, la cuantificación del contenido es difícil y los problemas de sensibilidad y ética que llevan aparejados son grandes.

La técnica «Q-sort» (STEPHENSON 1953) es un procedimiento de evaluación del autoconcepto muy utilizado tanto en clínica como en investigación psicológica básica (BLOCK 1971). La amplitud que dan las características de construcción de esta técnica permite su utilización desde muy variadas perspectivas. Por ejemplo, hoy está siendo una técnica sumamente útil en estudios sobre interaccionismo (BEM y FUNDER, 1978). El material de la prueba es flexible, puesto que consiste en una serie de tarjetas, en cada una de las cuales aparece una aseveración o cualidad sobre la manera de ser de las personas. El sujeto en exploración tan sólo ha de clasificarlas según lo apropiadas que sean a sí mismo. Esta técnica puede ser también utilizada por otras personas allegadas al sujeto evaluado. En los últimos años también se está utilizando, cada vez con más frecuencia, en la evaluación del cambio, es decir, como variable dependiente de tratamientos.

Por último, pueden ser incluidas entre las técnicas subjetivas construidas a partir de estrategias racionales teóricas, el REP de Kelly y el Diferencial Semántico de Osgood. El «Role Construct Repertory» de George KELLY (1955), basado en la teoría de los constructos personales del mismo autor, tiene como objetivo detectar las construcciones que los sujetos usan para definir, interpretar o calificar su entorno y a las personas importantes en él.

Por su parte, el Diferencial Semántico de OSGOOD (1957) evalúa las reacciones de los sujetos frente a palabras y conceptos estímulo en términos de estimaciones sobre escalas bipolares definidas por adjetivos. Ambos instrumentos presentan un gran valor no sólo clínico, sino, y tal vez ese sea el principal mérito, en la investigación psicológica de las más variadas vertientes.

En resumen, las técnicas subjetivas evalúan actividades psicológicas de los Sistemas Conductual y Microsocial, cuentan con un material de gran flexibilidad y son claramente voluntarias, no esmascaradas; se basan en estrategias racionales teóricas y no teóricas, y su marco referencial teórico es muy diverso, centrándose fundamentalmente en enfoques fenomenológicos y cognitivos.

4. Objetivas

El término «tests objetivos» ha sido utilizado por gran parte de autores (ANASTASI, 1973) para encuadrar los tests antes clasificados como psicométricos en virtud de su «objetividad» en las fases de corrección y puntuación y la tipificación de su material. Pervin, Cattell, Butcher y otros psicólogos prefieren calificar de «objetivos» aquellos instrumentos en los que el sujeto no puede manejar sus respuestas a voluntad. HUNDLEBY (1973) ha definido los tests objetivos como «los procedimientos... basados en las respuestas a estímulos específicos o grupos de estímulos, siendo su característica esencial que el sujeto no puede modificar sus respuestas voluntariamente» (p. 66). CRONBACH denomina a estos tests de «ejecución» y otros autores como EYSENCK (1959) prefieren el término «test de conducta». El apelativo de «objetivo» procede directamente de CATTELL (1950) habiéndose generalizado el término en los últimos años y siendo comúnmente empleado en la bibliografía actual.

Las características más importantes de estas técnicas son las siguientes: 1) El sujeto no percibe, necesariamente, la finalidad o la implicación de sus respuestas; es decir, son técnicas enmascaradas. 2) El sujeto no puede evitar responder, en la mayoría de los casos, ni modificar voluntariamente su comportamiento ante ellas; es decir, son involuntarias. 3) El material utilizado es estructurado. Los principios del aprendizaje y los procedentes de la Psicología cognitiva sirven de basamento teórico a estas técnicas que, a su vez, se apoyan fundamentalmente en una metodología experimental.

Se incluyen en este apartado aquellas técnicas propiamente objetivas que mantienen rigurosamente todas las características formuladas, así como otras dirigidas a la medición, a través de la observación directa y controlada, de comportamientos manifiestos.

Los instrumentos objetivos más utilizados son de condicionamiento, motores, perceptivos, de ejecución, así como de registro de las respuestas psicofisiológicas. La mayor parte de ellos exigen la utilización de precisos y costosos aparatos. No obstante, también existen otras técnicas objetivas que pueden ser empleadas fácilmente puesto que consisten en instrumentos de «lápiz y papel» (como por ejemplo el «Test de Figuras Enmascaradas» de Witkin).

Dos son los problemas fundamentales a la hora de considerar estas técnicas. Por una parte, su enorme diversidad impide un balance homogéneo

Cuadro 30
TESTS Y TÉCNICAS DE EVALUACION
(Resumen)

<i>Categoría</i>	<i>Características</i>	<i>Ejemplos</i>	<i>Clase de datos obtenidos</i>	<i>Enfoque teórico</i>	<i>Ejemplo teórico</i>	<i>Estrategias</i>
Psicométricas	Estructuradas Voluntarias	16 PF PMA	Rasgos Factores	Teóricos del rasgo	Cattell Thurstone	Empírica y factorial
Proyectivas	Semiestructuradas Enmascaradas	TAT Rorschach	Estructura inconsciente Conflictos	Psicodinámico	Murray	Racional teórica
Subjetivas	Semiestructuradas No-enmascaradas	Entrevista «Q-sort» REP	Percepciones individuales	Fenomenológicas Cognitivas	Rogers Kelly	Racional teórica No teórica
Objetivas	Estructuradas No-voluntarias	Registros Fisiológicos GEFT Códigos de Observación	Respuestas frente a estímulos	Aprendizaje Cognitivas	Witkin Skinner	Experimental

A partir de PERVIN (1973) (p. 181)

de todas ellas. De otra, el hecho de que sus puntuaciones se hayan efectuado inferencias utilizándolas como medidas de «constructos» con toda la problemática (repetidamente tratada) que esto encierra.

Por otra parte, incluimos en este apartado otra serie de instrumentos de observación que consisten en códigos de registro, a través de los cuales obtener medidas objetivas sobre las conductas manifiestas de los sujetos. Los psicólogos, como otros científicos, han desarrollado procedimientos de observación sistemática, tanto dirigidas a situaciones naturales como a situaciones artificiales o de laboratorio. Los mayores logros de la observación se han establecido, precisamente, en la estructuración de la información procedente de situaciones artificiales. En la actualidad existen un sinnúmero de test situacionales con distintos grados de estructuración para realizar la observación en contextos controlados (WEICK 1968, LAMBERT, COX y HARTSOUGH 1970, PAUL, TOBIAS y HOLLY 1972). También se han elaborado procedimientos de evaluación sistemática para situaciones naturales (BARKER y WRIGHT 1951, BERSOFF y MOYER 1973, LEWINSON y SHAFFER 1971, etc.), comprobándose altos índices de fiabilidad interjueces cuando se cuenta con precisas guías de observación y se controlan otras fuentes de error (FERNÁNDEZ BALLESTEROS 1979).

En resumen, las técnicas objetivas y conductuales se basan fundamentalmente en la Psicología del aprendizaje y en los hallazgos de los psicólogos cognitivistas, parten de estrategias experimentales, son técnicas altamente estructuradas y cuentan con diferentes grados de voluntariedad en la respuesta de los sujetos.

En el Cuadro 30 se presenta un resumen de la categorización propuesta, en la cual se han presentado las técnicas de evaluación psicológica. En él aparecen las categorías utilizadas junto con las características propias de las técnicas más frecuentemente empleadas, ejemplos de cada una de ellas, el tipo de datos obtenidos, los enfoques teóricos de referencia y las estrategias de construcción.

Ancaya Viéto, D. (2003). Diagnósticos en Educación. Diseño y uso de instrumentos. Sanz y Torres. Madrid

1. INSTRUMENTOS ESTANDARIZADOS Y PRUEBAS ESTANDARIZADAS

Por instrumento estandarizado se entiende todo aquel procedimiento de recogida de información que cuenta con un diseño altamente estructurado en el que las conductas objeto de observación quedan definidas de manera muy precisa como así, también, las condiciones de observación y las pautas para el registro, calificación e interpretación (en relación con un criterio o en relación con el rendimiento de un grupo normativo) de las observaciones, de modo que la información recogida con el mismo sobre distintos sujetos sea comparable.

La estandarización, como característica de un instrumento que asegura la comparabilidad de la información recogida con él referida a distintos sujetos, es una cualidad que puede estar presente en una amplia variedad de procedimientos de recogida de información. Así, la podemos encontrar en los tests cognitivos, en las pruebas de rendimiento académico, en los cuestionarios, inventarios y escalas que recogen información sobre muy diversas variables relacionadas con el desarrollo personal, en las entrevistas estructuradas y en las evaluaciones observacionales, igualmente, con alto nivel de estructuración, entre otros. Sólo se requiere que el instrumento en cuestión presente un diseño muy preciso que, de manera inequívoca, defina las conductas (o sus productos) objeto de observación y el modo como observarlas, registrarlas e interpretarlas.

Evidentemente, este diseño supone un trabajo muy elaborado que requiere el empleo de un tiempo y de unos recursos de los que el profesional del diagnósti-

co, a menudo, no dispone. Muchos de estos instrumentos se encuentran, por lo general, comercializados para que el profesional del diagnóstico pueda disponer de ellos. La elección entre éstos estará en función de los requerimientos de cada estudio diagnóstico concreto, teniendo siempre en cuenta las características, exigencias de uso y cualidades técnicas de estos instrumentos.

Los instrumentos estandarizados de corte comercial suelen contar con un manual en el que, entre otras cosas, además de la descripción del instrumento, se incluye información sobre sus fundamentos teóricos y empíricos, proceso de construcción, normas de utilización, cualidades métricas, y normas para la interpretación.

Las pruebas o tests constituyen un tipo particular de instrumentos estandarizados en los que las conductas (y/o productos de conductas) objeto de observación están relacionadas con la realización de una serie de tareas bien definidas que se proponen al sujeto en unas condiciones controladas de manera muy precisa.

En tanto que en otros instrumentos estandarizados la información a recoger se refiere a conductas expresadas en entornos naturales, independientemente de que su registro se haga en directo (evaluaciones observacionales altamente estructuradas, por ejemplo) o en diferido (cuestionarios, inventarios, escalas, entrevistas estructuradas) de manera auto o heteroinformada, en las pruebas estandarizadas o tests, la información está referida a conductas (o productos de conductas) exhibidas en la realización de unas tareas muy definidas que se presentan en situaciones artificiales. Las pruebas o tests cognitivos, las pruebas o tests de rendimiento académico y las pruebas o tests proyectivos son ejemplos de pruebas estandarizadas en las que las respuestas son provocadas mediante la exposición del sujeto a un material tipificado siguiendo unas instrucciones muy precisas dentro de un contexto artificial altamente controlado.

Como expresa Green (1981, p. 1001), “Una prueba estandarizada es una tarea o un conjunto de tareas dadas bajo condiciones estándar y que está diseñada para evaluar algún aspecto del conocimiento, habilidades o personalidad de un individuo... Una prueba proporciona una o más puntuaciones cuantitativas obtenidas de manera objetiva, de modo que, en tanto sea posible, a cada individuo se le evalúa del mismo modo. La intención consiste en proporcionar una comparación justa y equitativa entre las personas evaluadas”.

El presente Capítulo se va a centrar en la utilización de pruebas o tests estandarizados. En primer lugar, se hace referencia a los criterios para la selección de estos instrumentos para ser utilizados como recursos para la recogida de información en estudios diagnósticos concretos. En segundo lugar, se exponen aquellas estrategias que son comunes a la aplicación de cualquier prueba estandarizada y cuya observación, por parte del examinador, es esencial para su eficaz funcionamiento. En sendos apartados posteriores se trata de otras dos cuestiones de gran importancia en relación con la aplicación de pruebas estandarizadas: La observación de la conducta del sujeto durante la situación de examen, y la evaluación de la actuación del examinador en la aplicación de la prueba. Finalmente, como Anexo al Capítulo, se hace alusión, a modo de recordatorio, a las puntuaciones empleadas en pruebas estandarizadas con referencia a la norma.

2. CRITERIOS PARA LA SELECCIÓN DE PRUEBAS

Las pruebas estandarizadas a utilizar en un estudio diagnóstico concreto estarán en función de que se estime que pueden aportar información de calidad respecto de las variables de interés en ese caso concreto. Para realizar una valoración de esto, se han de tener en cuenta, al menos, los siguientes aspectos:

Conductas que muestrea la prueba

Algunos manuales contienen una descripción útil de las conductas muestreadas pero, en la mayoría de las ocasiones, sólo se nombran los dominios que se muestrean. Una manera eficiente de proceder en estos casos consiste en examinar las instrucciones de prueba, en especial aquéllas acerca de cómo calificar las respuestas del estudiante y el protocolo u hoja de respuestas, dado que, por lo general, estos materiales proporcionan una idea bastante buena acerca de cuáles son las conductas que en realidad se miden y, a partir de ahí, se puede realizar una descripción de las conductas en términos directos u operativos.

Sujetos a los que es aplicable la prueba

Aquí se deben examinar las edades y grados académicos de los estudiantes con los cuales se establecieron las normas para la prueba y toda aquella informa-

ción que describa a estos estudiantes, que deberá ser lo más detallada posible. Así, información socioeconómica, como ocupaciones de los padres, logros educativos de los mismos, ingresos familiares; distribución por sexo; distribución por grupos étnicos o raciales; distribución geográfica, etc. Esto es, información acerca de la medida en la que este grupo es representativo de la población a la que pertenece el sujeto estudiado.

Tipos de puntuaciones disponibles

Esto se puede obtener a partir de las instrucciones para calificar las respuestas y para combinar las puntuaciones crudas o directas en puntuaciones derivadas, así como las relacionadas con la obtención de puntuaciones totales desde la combinación o de puntuaciones parciales o de subpruebas. También, a partir de la consulta de los cuadros normativos, donde deben figurar las puntuaciones derivadas, de subpruebas y totales, correspondientes a las puntuaciones directas.

Interpretaciones de las puntuaciones obtenidas

Se han de examinar, en este sentido, las instrucciones proporcionadas para la interpretación (con referencia a la norma y/o con referencia al criterio), así como la claridad con la que son descritas las normas y/o los criterios que sirven de base para la interpretación.

Datos sobre fiabilidad de la prueba

Aquí se deberá prestar especial atención a clarificar cómo han entendido los autores del test la fiabilidad, qué datos ofrecen y con qué procedimientos han obtenido los resultados. Así, si los datos se refieren a consistencia interna, a equivalencia entre pruebas paralelas o a estabilidad; qué muestras se han utilizado para obtener estos datos; cuáles han sido los procedimientos estadísticos utilizados; cuál ha sido el intervalo temporal tenido en cuenta en la fiabilidad test-retest (estabilidad); si los resultados están referidos sólo a las puntuaciones totales o, también, a las puntuaciones parciales o de subpruebas; si están referidos sólo a la muestra total o se han calculado para cada grupo de edad o nivel académico a tener en cuenta dentro de aquélla; etc.

Datos sobre validez

Al igual que en el caso de la fiabilidad, la atención deberá fijarse en clarificar cómo han entendido los autores del test la validez, qué datos ofrecen y cómo han obtenido estos datos. Aunque actualmente la validez se entiende como un concepto unitario (ver en Anaya, 2002), puede resultar útil, por la sistematicidad que suele aportar, afrontar este examen de la validez desde los distintos tipos o clases tradicionales de validez. De este modo, cabe examinar qué datos se ofrecen sobre validez de contenido, dentro de los cuales se ha de prestar una especial atención a las explicaciones que se dan acerca de cómo se eligieron los reactivos de la prueba; qué datos se ofrecen sobre validez de criterio, tanto concurrente como predictiva, en relación a los cuales se han de examinar, con especial dedicación, los criterios utilizados, las muestras que soportan los resultados y los procedimientos con los que éstos han sido obtenidos; qué datos se ofrecen sobre validez de constructo, insistiendo, también en este caso, en las muestras que soportan los hallazgos y en los procedimientos con los que éstos han sido obtenidos.

En el Cuadro 3.1 se ofrece una guía que puede resultar de utilidad a la hora de seleccionar una prueba estandarizada. Por otra parte, una referencia obligada son los *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), que constituyen la más importante fuente acerca de normas técnicas y profesionales para la construcción y uso de pruebas.

Cuadro 3.1

Pautas para valorar una prueba

Información acerca de la prueba

1. ¿Cuál es el nombre de la prueba?
2. ¿Quiénes son sus autores?
3. ¿Quién la publicó?
4. ¿Cuándo se publicó?
5. ¿Existe disponible una forma alternativa?
6. ¿Cuánto cuesta?
7. ¿Cuánto tiempo requiere su aplicación?
8. ¿Existe un manual de prueba?

9. ¿Qué tan recientemente se revisó la prueba?
10. ¿Cuál fue el grupo de estandarización?

Ayudas para interpretar los resultados de la prueba

1. ¿El manual proporciona una exposición clara de (a) los propósitos y aplicaciones para los cuales sirve la prueba y (b) la

(continúa)

Cuadro 3.1
(Continuación)

<p>capacitación necesaria para la aplicación e interpretación de la prueba de manera adecuada?</p> <p>2. ¿La prueba, el manual, las formas de registro y los materiales acompañantes guían al usuario hacia interpretaciones fidedignas y correctas de los resultados de prueba?</p> <p>3. ¿Las afirmaciones que expresan relaciones se presentan en el manual en términos cuantitativos, de manera que el lector puede decir cuánta precisión o confianza puede asignarles?</p>	<p>9. ¿La prueba es adecuada para aplicación individual o grupal?</p>
<p>Confiabilidad y validez</p> <p>1. ¿Qué tan confiable es la prueba y sus partes componentes?</p> <p>2. ¿Qué tan válida es la prueba para los propósitos que se exponen?</p>	
<p>Aplicación y calificación</p> <p>1. ¿Las instrucciones para la aplicación están completas y son claras?</p> <p>2. ¿Son claros los procedimientos de calificación?</p>	
<p>Escalas y normas</p> <p>1. ¿Se describen de manera clara y cuidadosa las escalas utilizadas para informar de puntuaciones?</p> <p>2. ¿Las normas se expresan de manera apropiada, en general en puntuaciones estándar o rangos percentiles?</p> <p>3. ¿Se definen y describen de manera clara las poblaciones a las cuales se refieren las normas?</p> <p>4. Si está disponible más de una forma, ¿existen cuadros disponibles que muestren las puntuaciones equivalentes de las formas diferentes?</p> <p>5. ¿El manual discute el valor posible de las normas locales y proporciona cualquier tipo de ayuda para la preparación de normas locales?</p>	
<p>Consideraciones acerca de la persona evaluada</p> <p>1. ¿Qué habilidades de prerequisite necesita la persona evaluada para responder la prueba?</p> <p>2. ¿En qué lenguajes o modos de comunicación se puede aplicar la prueba?</p> <p>3. ¿El nivel de vocabulario de las instrucciones de prueba es el apropiado para la persona evaluada?</p> <p>4. ¿Cómo se presentan los reactivos de prueba?</p> <p>5. ¿Cómo se responde a los reactivos de prueba?</p> <p>6. ¿Qué adaptaciones formuladas y no formuladas pueden realizarse a los modos de presentación y respuesta?</p> <p>7. ¿La prueba está libre de prejuicio sexual y étnico?</p> <p>8. ¿Los materiales de prueba son interesantes para la persona evaluada?</p>	

3. ESTRATEGIAS PARA LA APLICACIÓN DE PRUEBAS ESTANDARIZADAS

Se exponen, a continuación, aquellas estrategias, de carácter general, que son comunes a la aplicación de cualquier prueba estandarizada y cuya observación, por parte del examinador, es esencial para su eficaz funcionamiento.

Establecimiento del *rapport*

La aplicación adecuada de cualquier prueba requiere de la colaboración del sujeto examinado. El evaluador deberá crear la atmósfera propicia para conseguir el interés, la actitud positiva y la participación activa del sujeto en la realización de las diferentes tareas que conforman la prueba. Para ello, es fundamental que el examinador muestre empatía, autenticidad, calidez y respeto por el evaluado y que ayude a éste a mantener un buen nivel de autoaceptación y autoestima.

Al inicio de la sesión de evaluación, antes de comenzar la aplicación formal de la prueba, es muy importante que, tras las presentaciones oportunas, se informe brevemente al evaluado del propósito de la evaluación y del interés para que haga las tareas lo mejor que pueda. Igualmente, mediante la comunicación oportuna, se debe alentar en el sujeto la sensación de confianza y seguridad en sí mismo suficientes para afrontar con decisión y optimismo la realización de la prueba.

En el transcurso de la aplicación debe mantenerse el *rapport*, de modo que permanezcan vivos el interés, la actitud positiva y la participación activa del sujeto en la realización de las diferentes tareas que conforman la prueba. A ello contribuirá de manera decisiva la utilización de los refuerzos convenientes, cuestión a la que haremos referencia algo más adelante.

Evidentemente, la manera efectiva de establecer y mantener el *rapport* estará en función de las características singulares del evaluado. Así, por ejemplo, cuestiones como la edad, la presencia de deficiencia mental, de problemas comportamentales, temperamento y expectativas del sujeto deberán ser tenidas en cuenta entre otros muchos aspectos. Más adelante, se insistirá en las características del evaluado.

Adecuación de las condiciones físicas

La aplicación de la prueba debe realizarse en unas condiciones físicas que aseguren que el rendimiento del sujeto no se vea afectado por aquéllas. Esto inclu-

ye la disposición de un local con ventilación, luminosidad y temperatura adecuadas, y, al mismo tiempo, acogedor, confortable, funcional respecto de las tareas a realizar, y al recaudo de elementos distractores tanto externos (ruidos, interrupciones por entradas y salidas de personas...) como internos (mobiliario y adornos llamativos, timbres o teléfonos que puedan sonar...). En todo momento, el examinador y los materiales de prueba deben ser los estímulos más sobresalientes (por ejemplo, se debe evitar situar al sujeto frente a una ventana) y el evaluado ha de contar con el mobiliario (mesa, silla, etc.) y espacio adecuados para realizar con comodidad las tareas demandas por la prueba.

Ajuste a las instrucciones de aplicación

Como se dijo anteriormente, una de las características esenciales de las pruebas estandarizadas es que cuentan con unas normas muy precisas acerca de cómo deben ser aplicadas. Cada prueba, por consiguiente, ha de ser aplicada de acuerdo con las instrucciones de aplicación contenidas en el manual correspondiente. Esto es muy importante, dado que las desviaciones respecto de las normas establecidas durante la aplicación de la prueba pueden cambiar el significado de las puntuaciones.

Sin embargo, habrá ocasiones en las que interese recoger información adicional acerca de las habilidades del evaluado mediante la modificación de las condiciones de aplicación. En estos casos, la aplicación modificada de alguna parte o de la totalidad de la prueba sólo deberá realizarse después de que la prueba completa se aplicó siguiendo las instrucciones estándar.

Por lo general, las modificaciones suelen consistir en la eliminación de los límites de tiempo en aquellas tareas que lo exigen; en el empleo de materiales alternativos; en la indagación de los procedimientos utilizados por el evaluado para resolver las tareas (por ejemplo, mediante preguntas al sujeto sobre qué y cómo lo está haciendo), y en la facilitación de ayudas en aquellas tareas que el evaluado no logra resolver con éxito por sí solo. Evidentemente, la información recogida en condiciones de aplicación modificadas no puede ser interpretada de acuerdo con las normas de interpretación de la prueba, que suponen un procedimiento de aplicación estándar. Pero ésta es una información verdaderamente útil sobre aspectos del rendimiento del sujeto muy importantes en el ámbito del diagnóstico en educación, cuales son el rendimiento en condiciones sin presión de tiempo, el rendimiento diferencial ante alternativos materiales o formas de presen-

tación de los reactivos, las estrategias utilizadas por el sujeto para afrontar las tareas, y el grado en el que el evaluado es capaz de aprender a partir del empleo de determinadas estrategias de enseñanza-aprendizaje.

Como es obvio, la información recogida en condiciones modificadas ha de hacer referencia, además de a las realizaciones del sujeto, a las correspondientes situaciones modificadas en las que se producen, que han de ser descritas con el máximo detalle.

Utilización adecuada de los refuerzos

El mantenimiento de la atención y la cooperación del evaluado durante la aplicación de la prueba puede exigir la utilización de refuerzos cuyo estilo y asiduidad estarán en función de las características del sujeto. Como normas generales, el reforzamiento deberá ser siempre positivo y estar dirigido a reforzar el esfuerzo, independientemente de que éste se acompañe o no del éxito. Los refuerzos pueden expresarse en forma de asentimiento, sonrisas o sonidos de apreciación, que, por lo general, suelen ser más efectivos que las palabras. El empleo de reforzadores tangibles, como golosinas, dulces, etc, deberá restringirse a sujetos con características muy especiales que, como niños autistas o con deficiencia mental profunda, pueden requerir este tipo de reforzadores primarios.

Aparte del reforzamiento positivo, otro motivador importante es la experimentación del éxito. Esto se puede conseguir disponiendo las tareas de tal modo que el sujeto experimente éxitos de manera intermitente. Pero jamás, y esto muy importante, se debe proporcionar al examinado retroalimentación acerca de si sus respuestas son correctas o incorrectas. Sólo debe reforzarse, repetimos, el esfuerzo.

Atención a las condiciones del evaluado

En el establecimiento y el mantenimiento del *rappor*t y, en general, en todo el proceso de aplicación, deben tenerse en cuenta las características singulares del evaluado, pero sin que ello afecte al procedimiento estandar de aplicación de la prueba. Existen, sin embargo, ciertos sujetos cuyas características sí que requieren cambios sustanciales en el procedimiento estándar de aplicación. Este es el caso de sujetos con alteraciones visuales o auditivas, con discapacidades físicas (que les pueden impedir, por ejemplo, escribir o manipular objetos), con daño cerebral,

con retraso mental grave o profundo, autismo, trastornos emocionales, o sujetos de minorías étnicas o emigrantes con pobre conocimiento del idioma.

Aparte de las escasas pruebas diseñadas expresamente para estos colectivos, existen algunas otras que cuentan con normas de aplicación específicas para algunos de ellos. Sin embargo, en la mayoría de las ocasiones esto no es así. En estos casos, las modificaciones realizadas sobre el procedimiento estándar de aplicación deben ser descritas con el nivel de detalle suficiente como para poder interpretar de manera adecuada las realizaciones del examinado.

Precisión en el registro de las respuestas

Las respuestas deben registrarse de manera tan precisa como sea posible. A ello ayudará el registro inmediato de las respuestas en los espacios apropiados que se proporcionan en el protocolo de registro de la prueba y, también, que el examinador haya estudiado, con antelación, los criterios de calificación de la prueba, de modo que ante cada respuesta, sepa cómo calificarla de una manera ágil. Las dudas que, no obstante, puedan surgir deberán resolverse acudiendo al manual en el que, por lo general, se suelen proporcionar pautas para discernir el tipo de respuesta que requiere cada reactivo o, si esto no es así, se deberán examinar los ejemplos que, aunque no pretenden representar todas las respuestas, pueden servir de guía para la calificación.

El registro preciso y ágil de las respuestas y, en general, la soltura y destreza en toda la aplicación de la prueba, requieren que, previamente, el evaluador se haya familiarizado con el material de la prueba y con las normas de aplicación y calificación. Este entrenamiento previo es muy importante porque, durante el examen, el evaluador debe tener el suficiente dominio de la situación como para que la aplicación fluya automáticamente y, de esta forma, pueda centrarse en la atención y observación del evaluado. La calificación y el registro de las respuestas jamás deben interferir la relación con el examinado y, por consiguiente, las consultas al manual para resolver dudas han de restringirse al máximo o, mejor, obviarse por completo.

Autocontrol del evaluador

Durante la aplicación de la prueba, el evaluador debe ser consciente de todo lo que hace y dice y de cómo esto puede afectar a las realizaciones del examinado.

La conciencia acerca de la propia conducta y de sus posibles efectos sobre el evaluado es esencial para el autocontrol de las propias actuaciones con vistas a evitar la contaminación sobre los resultados de prueba.

Pero para la minimización de las interferencias sobre los resultados no basta con el autocontrol de las actuaciones sino que se requiere, además, el autocontrol de las expectativas.

Existe una abrumadora base empírica que apoya el efecto de las expectativas del evaluador o de las impresiones personales acerca del evaluado —reacción conocida como *efecto halo*— sobre los resultados de prueba. La información preprueba que se tiene del examinado se ha comprobado repetidas veces que influye en las calificaciones que los examinadores otorgan a las respuestas y, así, por ejemplo, los sujetos que previamente habían sido descritos como más inteligentes tienen mayor probabilidad que los descritos como menos inteligentes de recibir calificaciones mejores por respuestas similares (Grossman, 1978; Ficus, 1975; Babad, Marin y Mar-Hayim, 1975).

Pero las expectativas o impresiones generales que el evaluador se forma sobre el evaluado se ha comprobado que no sólo influyen en la calificación de las respuestas sino, además, en las respuestas que obtienen de los evaluados y, en general, en toda la interacción evaluador-evaluado durante el proceso de aplicación (Ryan, Prifitera y Powers, 1983; Cuenot y Darbes, 1982; Flanklin, Stillman, Burpeau y Sabers, 1982).

Aunque parece que es prácticamente imposible eliminar por completo los efectos de halo en la aplicación, calificación y valoración de pruebas (Sattler y Squire, 1982), es necesario hacer todo lo posible por reducir al mínimo la influencia de estas reacciones en la aplicación de la prueba y en la calificación de las respuestas del evaluado y, más en general, por reducir cualquier fuente de subjetividad relacionada con el evaluador que pueda afectar la validez de las puntuaciones, dada la reiteradamente comprobada, también, variabilidad entre evaluadores en la calificación de unas mismas respuestas de prueba (Bradley, Hanna y Lucas, 1980; Ryan et al., 1983; Stillman et al., 1982).

Como cierre de este apartado, se ofrecen en el Cuadro 3.2 algunas pautas que pueden resultar de utilidad en la aplicación de pruebas estandarizadas.

Cuadro 3.2

Pautas para la aplicación de pruebas estandarizadas

I. Precauciones generales de evaluación.

A. Lea, aprenda y vuelva a leer las instrucciones.

B. Siempre adhiérase a los procedimientos estandarizados.

1. Utilice el fraseo correcto.
2. Mantenga un manejo adecuado del tiempo.
3. Presente los materiales en el modo prescrito
4. Siga las instrucciones de calificación de manera rígida.
5. No dependa sólo de la lectura de instrucciones escritas sino téngalas disponibles para una referencia rápida.

C. Sea objetivo.

1. No dé indicación de lo correcto o incorrecto de las respuestas del evaluado.
2. No dé claves acerca de las respuestas que se esperan; vigile la entonación verbal; recuerde que se está evaluando, no enseñando.

D. Sea natural.

1. Sea cálido, pero impersonal.
2. Aprenda a utilizar el fraseo estandarizado de un modo natural e informal.
3. Logre el *rapport* y un «toma y daca» verbal antes de que comience la prueba; tome una actitud de escuchar.

E. Prepare el ambiente:

1. Evite distracciones.
 - a. Visuales: Tenga al evaluado de espaldas a las puertas y ventanas donde el movimiento y la actividad ocurren. Haga que le

dé la espalda a espacios abiertos donde existen, por ejemplo, pinturas, colores o juguetes distractores. Evite obstrucciones.

- b. Auditivas: Evite áreas ruidosas.
- c. Emocionales: Evite la evaluación cuando el evaluado tiene prisa, está molesto o enfermo.

2. Proporcione condiciones óptimas para la buena ejecución.

- a. Vea que el evaluado se encuentre en una postura cómoda y que tenga una clara visión de los materiales.
- b. Proporcione una habitación bien iluminada, con ventilación adecuada y temperatura cómoda.
- c. Evite luces deslumbrantes, reflejos de las páginas y otras distracciones.
- d. Hable con voz clara y audible, a una velocidad moderada.
- e. Mantenga el interés por medio del entusiasmo, la atención al evaluado y la presentación uniforme del material.
- f. Alabe y aliente la ejecución general, pero nunca en reactivos específicos.
- g. Haga que el evaluado sepa que usted desea ver qué tan bien puede hacer las cosas.

II. Aplicación y calificación de la prueba.

A. Eficiencia.

1. Proporcione un orden y método eficientes de manejo de materiales para:
 - a. Registro.

Cuadro 3.2

(Continuación)

b. Vea el manual sin que éste se vuelva una barrera entre usted y el evaluado.

c. Retire y saque los materiales.

d. Evite demoras y distracciones para el evaluado.

2. Realice una transición continua de una prueba a otra y de un reactivo a otro. Se debe conocer cuál será la siguiente presentación en cada punto en la prueba.

3. Conozca suficientemente bien los materiales y calificaciones, de modo que no se extienda la prueba de manera innecesaria. Evaluar en exceso puede crear fatiga y desinterés.

a. Conozca los estándares de puntuación.

b. Comience en el punto apropiado.

4. Aprenda a manejar la conducta externa.

a. Ignore o redirija los comentarios irrelevantes.

b. Disminuya al mínimo los movimientos externos mediante el desarrollo de interés, motivación y orientación hacia la tarea. Si los movimientos externos no interfieren con el funcionamiento del evaluado, ignórelas. De ser necesario, proporcione al individuo una salida positiva como permitir que se tome de las rodillas de la mesa o que doble las manos.

c. Prever fatiga y distracción.

B. Calificación.

1. Es esencial que sepa bien las

normas de calificación. Este requisito se aplica en particular a la comprensión de la intención de cada prueba.

2. Recuerde que las normas de puntuación son sólo aquello que el término implica: son "normas" para la calificación, más que respuestas exhaustivas, con respuestas correctas e incorrectas. Con frecuencia es necesario evaluar respuestas equivalentes a la luz de otras respuestas que se enumeran en las normas de calificación, dado que no todas las respuestas posibles están incluidas.

3. Verifique todas las respuestas con el manual, a fin de revisar cualquiera de esté dudosa.

4. Vuelva a verificar cada paso en el proceso de calificación.

5. Verifique dos veces todas las cifras y cálculos: la edad cronológica, número de reactivos correctos, adicionales y cálculos aritméticos.

C. Cuidado de los materiales.

1. Cuando cualquiera de los materiales que se presentan al evaluado queden marcados o mutilados de cualquier forma que pueda influir en la respuesta del evaluado, reemplácelos.

2. Si se debe señalar hacia los dibujos, asegúrese de no marcar la página. Utilice la parte trasera de su pluma o lápiz.

3. Si se pierde o daña cualquier material, reemplácelo con objetos idénticos al original.

(continúa)

4. OBSERVACIÓN DE LA CONDUCTA DEL SUJETO DURANTE LA SITUACIÓN DE EXAMEN

La observación de la conducta del evaluado durante la situación de examen proporciona información muy valiosa con vistas a la correcta interpretación posterior de los resultados de prueba, a la estimación de la validez de estos resultados, y a la obtención de datos sobre otras variables interesantes del sujeto no estrictamente evaluadas por la prueba.

La observación de la conducta no debe interferir con la aplicación de la prueba. Se requiere una gran habilidad para estar pendiente del mantenimiento del *rapport*, del manejo de los materiales, del control de las tareas, de la calificación y registro de las respuestas, del propio autocontrol y, al mismo tiempo, atender a la observación de la conducta del evaluado. Esta habilidad no surge de manera automática sino que se adquiere a base de entrenamiento y práctica.

Las conductas, cuya observación y registro caben tener interés en relación a los fines antes mencionados, pueden ser muy diversas y variar, además, en cada caso concreto de evaluación. En el Cuadro 3.3 aparece un listado de manifestaciones comportamentales cuya observación puede resultar interesante. Con la experiencia, cada evaluador va depurando una serie de indicadores conductuales que considera de utilidad al servicio de los anteriores fines en los casos que, por lo general, él atiende. En este supuesto, es recomendable la confección de una lista de verificación conductual (o las listas alternativas que se precisen en función de la variedad de los casos que, habitualmente, se atienden) debido a las facilidades que proporcionan al examinador a la hora de la observación y el registro de los comportamientos de interés. En el Cuadro 3.4 se ofrece un ejemplo de lista de verificación conductual para ser utilizada en la aplicación de pruebas.

5. EVALUACIÓN DE LA ACTUACIÓN DEL EXAMINADOR EN LA APLICACIÓN DE PRUEBAS

La mejora constante que el profesional del diagnóstico debe procurar en su competencia para la aplicación de pruebas pasa, aparte del entrenamiento y experiencia convenientes, por la evaluación y reflexión sobre sus propias actuaciones

Cuadro 3.3

Conductas a observar durante la aplicación de pruebas

Rasgos actitudinales

Actitud hacia usted, el examinador

1. ¿Cómo se relaciona el evaluado con usted (y de qué manera se relaciona usted con él)?
2. ¿El evaluado es tímido, temeroso, agresivo o amistoso?
3. ¿Es negativo, obediente de manera normal o está demasiado ansioso por complacer?
4. ¿La actitud del evaluado hacia usted cambia durante la prueba?
5. ¿El evaluado trata de inducirlo a usted a que le dé respuestas a las preguntas?
6. ¿Lo vigila usted de manera cercana para descubrir si sus respuestas son correctas?

Actitud hacia la situación de prueba

1. ¿El evaluado se encuentra relajado y cómodo, tenso e inhibido o inquieto?
2. ¿Se encuentra interesado o no se involucra en la situación?
3. ¿Parece confiado en su propia capacidad?
4. ¿Está deseoso o renuente de participar?
5. ¿Ve las tareas como juegos, oportunidades para superarse o fuentes amenazantes de fracaso?
6. ¿Qué tan bien pone atención el evaluado en la prueba?
7. ¿Es necesario repetir las instrucciones o preguntas? Si es así, esta necesidad de repetición sugiere un problema auditivo, comprensión limitada del problema, problemas de atención, comprensión deficiente o un esfuerzo por parte del evaluado por obtener mayor tiempo para pensar en la pregunta. Pedirle al

sujeto que repita la pregunta puede proporcionar claves acerca de cuáles son los factores que explican su conducta (Zimmerman y Woo-Sam, 1985).

8. ¿Es fácil o difícil recuperar la atención del evaluado una vez que se ha perdido?
9. ¿El evaluado parece hacer su mejor esfuerzo?
10. ¿Hace el intento sólo cuando usted se lo exige?
11. ¿Se rinde con facilidad o insiste en continuar trabajando con reactivos difíciles?
12. ¿El interés del evaluado varía durante la evaluación?
13. ¿Cómo reacciona el sujeto ante las preguntas de sondeo (por ejemplo, considera la respuesta, define la primera respuesta, dice con rapidez «No sé» o permanece en silencio)?

Actitud hacia sí mismo

1. ¿El evaluado tiene aplomo y confianza?
2. ¿El evaluado realiza frecuentes afirmaciones autodespreciativas o jactancias o es bastante objetivo acerca de sus logros?
3. ¿Qué tan consciente está de lo adecuado de sus respuestas?

Hábitos de trabajo

1. ¿El ritmo de trabajo del evaluado es rápido o lento?
2. ¿Parece pensar en las respuestas y organizarlas o las da de manera impulsiva o descuidada?
3. ¿Reconsidera cualquiera de las respuestas?
4. ¿Piensa en voz alta o sólo da las respuestas finales?

(continúa)

Cuadro 3.3
(Continuación)

- 5. ¿Escribe las respuestas en la mesa con un dedo, pide continuamente que usted le haga aclaraciones o utiliza otros medios para resolver los problemas?

Reacción ante los reactivos de prueba

- 1. ¿Qué tipo de reactivos de prueba producen reacciones como ansiedad, tartamudeo o sonrojo?
- 2. ¿Existe cualquier área de la prueba en la que el evaluado se sienta más o menos cómodo?
- 3. ¿El evaluado se encuentra más interesado en ciertos tipos de reactivos que en otros?
- 4. ¿El evaluado se bloquea en algunos reactivos ("lo sé, pero simplemente no puedo pensar") ¿Si ocurre el bloqueo es en los reactivos fáciles, difíciles, o en todos los reactivos?
- 5. ¿Es necesario que se le inste a responder? De ser así, ¿la exhortación conduce a la respuesta? ¿Esta indica que el evaluado tiene el conocimiento para responder de manera correcta y tan sólo deseaba que le persuadiera (Zimmerman y Woo-Sam, 1985)?

Reacción ante el fracaso

- 1. ¿Cómo reacciona el evaluado ante los reactivos difíciles? ¿Se retrae, se vuelve agresivo, trabaja más, intenta hacer trampa, se vuelve evasivo o admite abiertamente el fracaso?
- 2. Si él se vuelve agresivo ¿hacia quién o qué dirige su agresividad?
- 3. ¿De qué manera reacciona ante el fracaso?
- 4. ¿Se disculpa, racionaliza, medita, acepta con calma el fracaso o se siente humillado?

- 5. Si se siente humillado ¿Expresa impotencia o perplejidad, sugiriendo pérdida de capacidad (Zimmerman y Woo-Sam, 1985)?
- 6. ¿El evaluado puede aceptar que se le trate de dar seguridad?

Reacción ante la alabanza

- 1. ¿Cómo reacciona el evaluado ante la alabanza?
- 2. ¿La acepta con gracia o de modo desagradable?
- 3. ¿La alabanza motiva al evaluado a trabajar más?

Lenguaje

- 1. ¿El evaluado se expresa de manera clara? ¿Su discurso es fluido, vacilante, expresivo, inexacto o preciso?
- 2. ¿Qué tan precisamente se expresa? ¿Sus respuestas son directas y específicas, vagas y evasivas, por asociación libre, perseverativas o extrañas?
- 3. ¿Las respuestas reflejan preocupaciones personales o egocentrismo?
- 4. ¿Las respuestas son sumamente inmaduras?
- 5. Si el evaluado da información externa ¿ésta sugiere una necesidad compulsiva por cubrir todas las posibilidades o es completamente irrelevante (Zimmerman y Woo-Sam, 1985)?
- 6. ¿El evaluado conversa de manera espontánea o sólo en respuesta a preguntas?
- 7. ¿Su conversación parece derivarse de su manera amistosa de ser o de un deseo por evadir la situación de prueba?

(continúa)

Cuadro 3.3
(Continuación)

Visomotricidad

- 1. ¿Existe cualquier movimiento que el evaluado realice con sus manos, pies o rostro que valga la pena señalarse?
- 2. ¿Es diestro o zurdo?
- 3. ¿Su tiempo de reacción es rápido o lento?
- 4. ¿El evaluado procede de modo sistemático o en forma de ensayo y error?
- 5. ¿Es hábil o torpe?
- 6. ¿Ejecuta los movimientos bilaterales de manera hábil o torpe?
- 7. ¿El evaluado está consciente de los límites de tiempo en las tareas que lo requieren? De ser así ¿de qué manera afecta esta conciencia a su conducta?

(Sattler, 1992)

- 8. ¿El evaluado verbaliza mientras ejecuta las tareas? De ser así, ¿las verbalizaciones son congruentes con sus acciones?

Comparación de tareas verbales y no verbales

- 1. ¿Existen diferencias en la reacción del evaluado ante las tareas verbales y no verbales (por ejemplo, está más ansioso o cómodo con un tipo de tarea)?
- 2. ¿Comprende igualmente bien las instrucciones para las tareas verbales y no verbales?

Cuadro 3.4

Lista de verificación conductual para utilizar en la aplicación de pruebas

Nombre del examinado:

Edad:

Nivel académico:

Prueba(s) aplicada(s):

Examinador:

Instrucciones: Coloque una X en al línea apropiada para cada escala:

I. Actitud hacia el examinador y hacia la situación de prueba:

- | | | |
|------------------------|--|------------------------|
| 1. cooperativo | -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: | no cooperativo |
| 2. pasivo | -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: | agresivo |
| 3. tenso | -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: | relajado |
| 4. se rinde fácilmente | -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: -: | no se rinde fácilmente |

(continúa)

Cuadro 3.4
(Continuación)

II. Actitud hacia sí mismo:		
5. confiado	--:--:--:--:--:--	no confiado
6. crítico del propio trabajo	--:--:--:--:--:--	aceptante del propio trabajo
III. Hábitos de trabajo:		
7. rápido	--:--:--:--:--:--	lento
8. deliberado	--:~:~:~:~:~:~	impulsivo
9. piensa en voz alta	--:~:~:~:~:~:~	piensa en silencio
10. descuidado	--:~:~:~:~:~:~	cuidadoso
IV. Conducta:		
11. calmado	--:~:~:~:~:~:~	hiperactivo
V. Reacción ante el fracaso:		
12. consciente del fracaso	--:~:~:~:~:~:~	no consciente del fracaso
13. trabaja más después del fracaso	--:~:~:~:~:~:~	se rinde fácilmente después del fracaso
14. calmado después del fracaso	--:~:~:~:~:~:~	agitado después del fracaso
15. se disculpa después del fracaso	--:~:~:~:~:~:~	no se disculpa después del fracaso
VI. Reacción ante la alabanza:		
16. acepta la alabanza con gracia	--:~:~:~:~:~:~	acepta la alabanza de manera desagradable
17. trabaja más después de la alabanza	--:~:~:~:~:~:~	se retrae después de la alabanza
VII. Discurso y lenguaje:		
18. discurso deficiente	--:~:~:~:~:~:~	discurso correcto
19. lenguaje articulado	--:~:~:~:~:~:~	lenguaje desarticulado
20. respuesta directa	--:~:~:~:~:~:~	respuesta vaga
21. conversa espontáneamente	--:~:~:~:~:~:~	sólo habla cuando se le habla
22. lenguaje extraño	--:~:~:~:~:~:~	lenguaje orientado a la realidad
VIII. Visomotricidad:		
23. tiempo lento de reacción	--:~:~:~:~:~:~	tiempo rápido de reacción
24. ensayo y error	--:~:~:~:~:~:~	cuidadoso y sistemático
25. movimientos hábiles	--:~:~:~:~:~:~	movimientos torpes

(continúa)

Cuadro 3.4
(Continuación)

IX. Motricidad:		
26. coordinación motora defectuosa	--:~:~:~:~:~:~	coordinación motora adecuada
X. Resultados generales de la prueba:		
27. confiable	--:~:~:~:~:~:~	no confiable
28. válida	--:~:~:~:~:~:~	inválida

(Sattler, 1992)

evaluadoras. Se requiere una gran dosis de actitud crítica sobre sí mismo y de disposición permanente al cambio meliorativo si, realmente, se quiere ir alcanzando, de manera progresiva, un nivel competencial mayor en la correcta utilización de pruebas estandarizadas.

Aparte de las anteriores disposiciones de mejora, la eficacia de la evaluación sobre las actuaciones en las aplicaciones de pruebas llevada a cabo, bien por el propio examinador, bien por otros colegas con mayor experiencia y nivel competencial, requiere de la suficiente sistematicidad para que, por una parte, se asegure que son tenidos en cuenta todos aquellos aspectos de la actuación considerados relevantes y, por otra, que su observación y registro puedan realizarse con facilidad y rigor con vistas a su posterior análisis. En este sentido, es muy recomendable la elaboración de listas de verificación del comportamiento del examinador durante la aplicación de pruebas, que pueden ser cumplimentadas por el propio interesado, inmediatamente acabada la sesión de examen, o por otro colega que haya seguido como observador el desarrollo de la misma.

En el Cuadro 3.5 se presenta una lista de verificación para las prácticas generales de aplicación de pruebas, que puede ser utilizado como instrumento de auto-evaluación después de la sesión de examen o como instrumento de heteroevaluación por otro colega asistente a la misma como observador.

Cuadro 3.5

Lista de verificación para las prácticas generales de aplicación de pruebas

Nombre del examinado:

Edad:

Nivel académico:

Prueba(s) aplicada(s):

Fecha de aplicación:

Examinador:

Observador:

Rodee uno

- | | | |
|----|----|---|
| Sí | No | 1. Estableció <i>rapport</i> antes de comenzar la prueba |
| Sí | No | 2. Preparó al evaluado para la evaluación |
| Sí | No | 3. Arregló el ambiente de prueba para disminuir distracciones |
| Sí | No | 4. Evitó manierismo distractores |
| Sí | No | 5. Mostró interés en el evaluado |
| Sí | No | 6. Le dio amplio apoyo y aliento al evaluado |
| Sí | No | 7. Utilizó vestuario apropiado, no distractor |
| Sí | No | 8. Habló en un volumen apropiado para el entorno |
| Sí | No | 9. Pareció abierto y aceptante de los sentimientos del evaluado |
| Sí | No | 10. Pareció cómodo con él |
| Sí | No | 11. Mantuvo contacto visual frecuente con el evaluado |
| Sí | No | 12. Tuvo presentes todos los materiales necesarios |
| Sí | No | 13. Arregló los materiales de manera conveniente |
| Sí | No | 14. Colocó el manual para que el evaluado no pudiera leerlo |
| Sí | No | 15. Arregló los materiales de modo que el sujeto no pudiera ver otros reactivos de prueba que aquel (aquellos) en uso |
| Sí | No | 16. Manipuló los materiales con facilidad y confianza |
| Sí | No | 17. Leyó todas las instrucciones al pie de la letra |
| Sí | No | 18. Utilizó procedimientos precisos de manejo de tiempo |
| Sí | No | 19. Empleo procedimientos discretos de manejo de tiempo |
| Sí | No | 20. Registró las repuestas en el protocolo de registro |
| Sí | No | 21. Controló el ritmo de la evaluación para adecuarlo a la capacidad del evaluado |
| Sí | No | 22. Explicó de manera adecuada los procedimientos de prueba |
| Sí | No | 23. Utilizó el vocabulario adecuado para el evaluado |
| Sí | No | 24. Mostró conciencia de los signos de fatiga |

(continúa)

Cuadro 3.5

(Continuación)

- | | | |
|----|----|---|
| Sí | No | 25. Manejó de manera apropiada la fatiga |
| Sí | No | 26. Mostró conciencia de las perturbaciones emocionales |
| Sí | No | 27. Manejó de manera adecuada las perturbaciones emocionales |
| Sí | No | 28. Tomó los descansos necesarios |
| Sí | No | 29. Manejó de modo adecuado las manipulaciones del evaluado |
| Sí | No | 30. Dio explicaciones y aclaraciones apropiadas |
| Sí | No | 31. Utilizó preguntas adicionales para aclarar, no para mejorar, las respuestas del sujeto |
| Sí | No | 32. Realizó el interrogatorio en un modo no amenazante |
| Sí | No | 33. Alabó en forma apropiada al evaluado (v. gr., no alabó las respuestas correctas, alabó el esfuerzo) |
| Sí | No | 34. Manejó de manera adecuada las interrupciones |
| Sí | No | 35. Respondió de manera honesta y positiva a las preguntas del sujeto, en formas consistentes con la evaluación |

(Sattler, 1992)

ANEXO

Puntuaciones empleadas en pruebas con referencia a la norma

Para facilitar la interpretación, las puntuaciones directas o brutas obtenidas en las pruebas estandarizadas suelen transformarse en puntuaciones derivadas. Las más utilizadas son las puntuaciones equivalentes por edad y por grado académico, los rangos percentiles y las puntuaciones típicas.

Puntuaciones equivalentes de edad y de grado académico

Las puntuaciones equivalentes de edad se corresponden con las puntuaciones promedio obtenidas en una prueba por los sujetos de cada una de las diferentes edades que conforman la muestra normativa con la que se tipificó dicha prueba. A partir de aquí, a un sujeto evaluado con la prueba se le asigna la edad de rendimiento en la misma cuyo correspondiente promedio esté más aproximado a la puntuación bruta obtenida por el sujeto. Con sujetos mayores, la unidad de edad más empleada es el año, pero con niños pequeños suele utilizarse con mayor frecuencia el mes e, incluso, la semana, en el caso de la evaluación de neonatos [ver en Anaya (2002), fundamentalmente en el Capítulo 3, ejemplos al respecto].

Cuadro 3.6

Ejemplo de puntuaciones equivalentes por edad

Grupo de edad	Puntuación media
6 años	8.50
7 años	12.26
8 años	15.81
9 años	17.28
10 años	19.23
11 años	22.05
12 años	25.11
13 años	27.73
14 años	29.58
15 años	32.15

Supongamos que las puntuaciones medias por edad del grupo normativo con el que se tipificó la prueba X sean las que figuran en el Cuadro 3.6. Si a un sujeto se le aplica la prueba X y obtiene una puntuación de 16 puntos, se puede decir que ha obtenido en la prueba un rendimiento equivalente a los niños de 8 años.

De manera similar, las puntuaciones equivalentes de grado académico se corresponden con las puntuaciones promedio obtenidas en una prueba por los sujetos de cada uno de los diferentes grados o niveles académicos abarcados por la muestra normativa que sirvió para la tipificación de dicha prueba. A partir de aquí, a un sujeto evaluado con la prueba se le asigna el grado o nivel académico cuyo correspondiente promedio esté más aproximado a la puntuación directa obtenida por el sujeto.

Supongamos que las puntuaciones medias por grado académico del grupo normativo con el que se tipificó la prueba X sean las que figuran en el Cuadro 3.7. Si a un alumno se le aplica la prueba X y obtiene una puntuación de 28 puntos, se puede decir que ha obtenido en la prueba un rendimiento equivalente a los estudiantes de 4.º de Primaria.

Cuadro 3.7

Ejemplo de puntuaciones equivalentes por grado académico

Grupo de grado académico	Puntuación media
1.º de Primaria	15.38
2.º de Primaria	20.81
3.º de Primaria	24.14
4.º de Primaria	27.82
5.º de Primaria	31.09
6.º de Primaria	34.75

Otra forma de entender las puntuaciones equivalentes de edad consiste en agrupar los ítems de la prueba por edades, de modo que cada grupo esté constituido por aquellos ítems que son satisfactoriamente resueltos por la mayoría de los sujetos de una edad determinada. Si, por ejemplo, la muestra normativa de una prueba cubre desde los 3 a los 16 años de edad y se quieren establecer los grupos de ítems por año, el primer grupo estaría constituido por aquellos ítems que son

resueltos con éxito por la mayoría de sujetos de 3 años; el segundo, por aquellos ítems que son resueltos con éxito por la mayoría de los sujetos de 4 años y así, sucesivamente, hasta el grupo correspondiente a los sujetos de 16 años. Evidentemente, cada grupo de ítems correspondiente a una edad determinada engloba los ítems propios de las edades precedentes.

Este procedimiento fue introducido en la edición de 1908 de la *Escala Métrica de la Inteligencia* de Binet y Simon [ver el Capítulo 2 de Anaya (2002)], en la que se estableció el concepto de *nivel* o *edad mental*. Este hacía referencia a la capacidad para resolver con éxito los ítems propios de una edad determinada. De este modo, a los sujetos examinados con la escala se les asignaba como nivel o edad mental la edad a la que correspondía el grupo de ítems que lograban resolver con éxito. Así, si un niño era capaz de realizar correctamente los ítems correspondientes a los chicos de 9 años de la muestra normativa, pero no ya los correspondientes a los chicos de 10 años, se le asignaba una edad mental de 9 años.

En la práctica, sin embargo, la asignación de una determinada edad mental no es tan sencillo porque, por lo general, los sujetos no resuelven correctamente la totalidad de ítems de una edad y, a partir de ahí, ninguno de los específicos de la edad siguiente, sino que pueden contestar correctamente a ítems de edades superiores y, al mismo tiempo, fracasar en ítems del último nivel de edad en el que el número de aciertos todavía es superior al número de errores e, incluso, en ítems de niveles anteriores. Es por ello que, en primer lugar, se suele asignar la *edad basal*, que es la edad más alta en la que fueron resueltos correctamente todos los ítems y, a partir de ahí, se van sumando meses en función del número de ítems de edades superiores que el sujeto contesta con acierto. La edad mental finalmente asignada será la que resulte de estas adiciones a la edad basal.

Con el fin de facilitar la interpretación de la edad mental en función de la edad cronológica, se introdujo el concepto de *cociente intelectual* o *cociente de inteligencia* (CI) en la Stanford-Binet de 1916 [ver el Capítulo 2 de Anaya (2002)]. Este se definió como la razón entre la edad mental (EM) y la edad cronológica (EC), que suele multiplicarse por 100 para evitar los decimales. Esto es, $CI = EM / EC \times 100$.

Pronto estos cocientes de razón fueron utilizados en otras pruebas estandarizadas al objeto de interpretar las puntuaciones equivalentes de edad en función de la edad cronológica y, aparte del CI, empleado en muchas pruebas de capacidad

intelectual, surgió el cociente de instrucción en pruebas o tests de rendimiento académico, o el cociente lingüístico en pruebas de capacidad lingüística, etc. Pero todos estos cocientes de razón tienen el grave inconveniente de hacer incomparables los cocientes para diferentes edades, dado que al no permanecer constante con la edad la desviación típica de su distribución, un mismo cociente tiene diferentes significados a diversas edades. Este inconveniente queda resuelto con el empleo del cociente de desviación o de dispersión, que es una puntuación típica que se verá más adelante.

Aparte de los inconvenientes de los cocientes de razón, las puntuaciones equivalentes de edad y de grado académico tienen, por sí mismas, varias limitaciones. Entre las más importantes cabe citar las siguientes (Bennett, 1982; Crocker y Algina, 1986):

- Las distintas puntuaciones equivalentes de edad o de grado académico no representan unidades constantes. Esto quiere decir que, por ejemplo, la diferencia entre las puntuaciones equivalentes a 3 y a 4 años, pongamos por caso, tiene un significado distinto a la diferencia entre las puntuaciones equivalentes a 9 y a 10 años, o que la diferencia entre las puntuaciones equivalente a 1.º de primaria y a 2.º de primaria tiene un significado distinto a la diferencia entre las puntuaciones equivalentes a 5.º y a 6.º de primaria.
- Las puntuaciones equivalentes de edad o de grado académico tienen significado distinto en distintas pruebas e, incluso, entre diversas subpruebas de una misma batería. Una puntuación equivalente a 6 años obtenida en una prueba, por ejemplo, no tiene el mismo significado que una puntuación equivalente a 6 años obtenida en otra prueba distinta, con lo que no pueden establecerse comparaciones para el mismo sujeto entre rendimientos obtenidos en pruebas diferentes.
- Las puntuaciones equivalentes de edad o de grado académico suponen que en el transcurso de las edades o de los grados académicos se produce un crecimiento o progreso continuo y constante; pero esta suposición no parece estar justificada porque el progreso no es el mismo entre edades o entre grados académicos diferentes. Así, por ejemplo, el progreso que puede darse en un individuo entre los 3 y los 4 años es distinto, cuantitativa y

cualitativamente hablando, que el que se pueda producir entre los 15 y los 16 años; e igual acontece si se compara el progreso que pueda haber entre 1.º y 2.º de educación primaria con el que se dé entre 5.º y 6.º

- Muchas de las puntuaciones equivalentes de edad o de grado académico se obtienen por interpolación o por extrapolación, de modo que, en realidad, no han sido obtenidas por los sujetos del grupo normativo.
- Las puntuaciones equivalentes de edad o de grado académico dan pie a interpretaciones erróneas. Con frecuencia, la obtención de una puntuación equivalente a una determinada edad o grado académico se interpreta como que el sujeto tiene un funcionamiento similar al de los chicos de esa edad o de ese grado académico; sin embargo, la interpretación debe hacerse sólo en relación con los sujetos de su misma edad o grado académico. Si, por ejemplo, un chico de 1.º de primaria obtiene en una prueba de lectura un rendimiento equivalente a 2.º de primaria, esto no debe interpretarse como que tiene un funcionamiento general de un chico de 2.º de primaria en habilidades lectoras, sino que su rendimiento en las tareas del test es superior a la media de los niños de su grado académico y es equivalente al rendimiento medio de los niños de 2.º de primaria, estrictamente, en la resolución de esas tareas, lo que no implica que tenga un desempeño a un nivel consistente con las expectativas curriculares de 2.º de primaria.

A pesar de estas limitaciones, que han de ser siempre tenidas en cuenta, las puntuaciones equivalentes de edad y de grado académico son útiles para reflejar el desarrollo de los evaluados, en términos de las realizaciones típicas de los sujetos de diferentes edades o grados académicos, fundamentalmente en edades en las que el progreso o el crecimiento se produce de manera rápida ligada a la progresión cronológica.

Rangos percentiles

Los rangos percentiles son puntuaciones derivadas que expresan el porcentaje de sujetos de la muestra normativa que obtienen puntuaciones inferiores a una puntuación dada. Así, por ejemplo, si a un sujeto se le asigna en una prueba un rango percentil de 58, ello quiere decir que el 58% de los integrantes del grupo normativo han obtenido puntuaciones inferiores a las obtenidas por dicho sujeto.

Los rangos percentiles, en definitiva, son puntuaciones ordinales que indican la posición relativa del examinado dentro de la muestra de estandarización.

El cálculo de los rangos percentiles es muy sencillo. Para datos no agrupados se utiliza la siguiente fórmula:

$$p_i = \frac{1/2 f_i + fa_i}{N} \times 100, \text{ donde}$$

p_i = rango percentil correspondiente a la puntuación i

f_i = frecuencia de la puntuación i

fa_i = frecuencia acumulada anterior a la puntuación i

N = número de sujetos del grupo normativo

Supongamos la distribución de frecuencias de las puntuaciones de un test en un grupo normativo de 166 sujetos que aparece en el Cuadro 3.8.

Cuadro 3.8

Ejemplo de distribución de frecuencias de las puntuaciones de un test

Puntuaciones directas	f	fa
24	3	166
23	7	163
22	9	156
21	12	147
20	14	135
19	15	121
18	17	116
17	23	99
16	20	76
15	18	56
14	13	38
13	10	25
12	8	15
10	5	7
9	2	2

Si, por ejemplo, queremos calcular el percentil correspondiente a la puntuación 19, tendríamos que

$$P_{19} = \frac{1/2 \cdot 15 + 116}{166} \times 100 = \frac{7.5 + 116}{166} \times 100 = 74$$

luego un individuo que obtuviera 19 puntos en dicha prueba habría obtenido un rendimiento que dejaría por debajo al 74% de los sujetos del grupo normativo.

Para datos agrupados se utiliza la siguiente fórmula:

$$p_i = \frac{\left(\frac{x_i - l_i}{a_i} \right) f_i + fa_i}{N} \times 100, \text{ donde}$$

- p_i = rango percentil correspondiente a la puntuación i
- x_i = puntuación directa i
- l_i = límite inferior del intervalo que contiene la puntuación i
- a_i = amplitud del intervalo que contiene la puntuación i
- f_i = frecuencia del intervalo que contiene la puntuación i
- fa_i = frecuencia acumulada anterior al intervalo que contiene la puntuación i
- N = número de sujetos del grupo normativo

Supongamos la distribución de frecuencias de las puntuaciones de un test en un grupo normativo de 260 sujetos que aparece en el Cuadro 3.9.

Si, por ejemplo, queremos calcular el percentil correspondiente a la puntuación 43, tendríamos que

$$p_{43} = \frac{\left(\frac{43 - 40.5}{5} \right) 18 + 223}{260} \times 100 = 89$$

luego un individuo que obtuviera 43 puntos en dicho test habría obtenido un rendimiento que dejaría por debajo al 89% de los sujetos del grupo normativo. [Nótese

Cuadro 3.9

Ejemplo de distribución de frecuencias de las puntuaciones agrupadas de un test

Puntuaciones directas	f	fa
60-56	3	260
55-51	7	257
50-46	9	250
45-41	18	241
40-36	34	223
35-31	47	189
30-26	51	142
25-21	41	91
20-16	29	50
15-11	18	31
10-6	9	13
5-1	4	4

que los límites reales de un intervalo se encuentran sumando 0.5 a la puntuación límite superior y restando 0.5 a la puntuación límite inferior. De ese modo, el límite superior del intervalo 41-45, que es el que contiene la puntuación 43, es 45.5; el límite inferior 40.5; y la amplitud del intervalo $45.5 - 40.5 = 5$].

Los rangos percentiles tienen varios inconvenientes. Entre los más destacados cabe citar los siguientes (Croker y Algina, 1986):

- Los distintos percentiles a lo largo de la distribución no representan unidades constantes. Esto supone que las diferencias entre percentiles, o el paso de un percentil a otro, a lo largo de la distribución de percentiles no tienen el mismo significado. Así, por ejemplo, el paso del percentil 50 al percentil 51 no tiene el mismo significado que el paso del percentil 96 al percentil 97, dado que la diferencia en puntuaciones directas que se requiere en este segundo caso es mucho mayor que en el primero. Igualmente, por el mismo motivo, no es comparable la distancia entre el percentil 40 y el 50 que entre los percentiles 80 y 90, pongamos por caso. La diferencia real de rendimiento en la prueba es mucho más pequeña entre un sujeto con percentil 40 y otro con percentil 50, que entre un sujeto con percentil 80 y otro con percentil 90.

- En relación con lo anterior, conforme nos aproximamos a la zona media de la distribución se necesita menor diferencia en puntuaciones directas para pasar de un percentil a otro, en tanto que esta distancia se acrecienta a medida que se avanza hacia los extremos.
- Los rangos percentiles son puntuaciones ordinales que sólo facilitan el ordenamiento de los sujetos de acuerdo con sus rendimientos en la prueba; por consiguiente, no permiten la realización de operaciones aritméticas con ellos.

No obstante las limitaciones anteriores, los rangos percentiles pueden resultar útiles, sobre todo, para facilitar los resultados de las pruebas estandarizadas a personas no expertas, como los propios sujetos evaluados, sus padres o sus profesores, dado que permiten una interpretación sencilla y directa en unos términos, como los porcentajes, cuyo uso es muy habitual en la vida diaria.

Puntuaciones típicas

Una puntuación típica expresa la distancia desde una puntuación directa a la media de la distribución en términos de la desviación típica de esa distribución. Esto facilita la interpretación de las puntuaciones o de los rendimientos de los sujetos en la prueba desde un patrón común y, por tanto, la comparación de los rendimientos obtenidos por los distintos sujetos e, incluso, de los rendimientos obtenidos en distintas pruebas, siempre que la forma de la distribución sea en todas ellas semejante. Además, al mantenerse como medidas de intervalo, al igual que las puntuaciones directas, conservan todas las ventajas matemáticas inherentes a este tipo de puntuaciones. Todas estas cualidades hacen que sean las puntuaciones típicas las puntuaciones derivadas preferidas.

Las puntuaciones directas se transforman en puntuaciones típicas mediante la siguiente fórmula:

$$z_i = \frac{x_i - \bar{x}}{s}, \text{ donde}$$

- z_i = puntuación típica correspondiente a la puntuación directa x_i
- x_i = puntuación directa i
- \bar{x} = media aritmética de la muestra
- s = desviación típica o desviación estándar de la muestra

Supongamos que un estudiante A obtiene una puntuación bruta de 35 puntos en el test X, siendo 42 la media y 12 la desviación típica o desviación estándar de la muestra de referencia. La puntuación típica correspondiente a dicha puntuación directa será:

$$z_{35} = \frac{35 - 42}{12} = -.58$$

Esto quiere decir que el rendimiento del estudiante en dicha prueba se aparta en .58 puntuaciones típicas por debajo de la media de rendimiento de su grupo de referencia. Si otro estudiante B obtuviera en la misma prueba una puntuación bruta de 50 puntos, le correspondería una puntuación típica de .66, que significa que su rendimiento se aparta en .66 puntuaciones típicas por encima de la media.

Dado que las puntuaciones típicas conservan las propiedades matemáticas de las medidas de intervalo, se pueden comparar las z de ambos estudiantes y decir que el estudiante B ha obtenido en el test X un rendimiento que es superior en 1.24 desviaciones típicas al logrado por el estudiante A. Por otra parte, si el estudiante A hubiese obtenido en otro test Y, con distribución semejante a la del test X, una z de .42, se podría decir que su rendimiento en el test Y es superior en 1.0 desviación típica al obtenido en el test X. Estas comparaciones son, como se vio anteriormente, imposibles de hacer utilizando rangos percentiles o puntuaciones equivalentes de edad o de grado.

Las puntuaciones z se distribuyen con media de 0 y desviación típica de 1. De acuerdo con esta distribución, en la mayoría de las pruebas estandarizadas, los rendimientos, expresados en términos de z , van desde -3.0 hasta +3.0. El empleo constante de los signos + y - puede resultar incómodo. Con el fin de eliminar este inconveniente, se han propuesto diversas puntuaciones típicas derivadas que se basan en distribuciones que tienen otras medias y desviaciones típicas. Entre las más utilizadas están las siguientes:

- *Puntuaciones T.* Estas puntuaciones se distribuyen con media de 50 y desviación típica de 10. Su obtención a partir de puntuaciones z se realiza del siguiente modo:

$$T = 10(z) + 50$$

Así, por ejemplo, la puntuación T correspondiente a una $z = -.58$ sería:

$$T = 10 (-.58) + 50 = 44.2$$

O la correspondiente a $z = .66$ sería:

$$T = 10 (.66) + 50 = 56.6$$

— *CI de desviación*. Estas puntuaciones se distribuyen con media de 100 y desviación típica de 15 (que es utilizada de manera más general) o de 16 (que es utilizada en algunas pruebas, como la Stanford-Binet). [Ver esto último en Anaya (2002)]. Su obtención a partir de puntuaciones z se realiza del siguiente modo:

$$CI = 15 (z) + 100$$

De esta manera, las anteriores dos z darían lugar a los siguientes CI:

$$CI = 15 (-.58) + 100 = 91$$

$$CI = 15 (.66) + 100 = 110$$

(se observará que se han redondeado los decimales)

— *Puntuaciones del College Entrance Examination Board* o puntuaciones CEEB. Se idearon con el fin de eliminar por completo los decimales. Estas puntuaciones se distribuyen con media de 500 y desviación típica de 100. Su obtención a partir de puntuaciones z se realiza del siguiente modo:

$$CEEB = 100 (z) + 500$$

De esta forma, las dos z anteriores darían lugar a las puntuaciones CEEB siguientes:

$$CEEB = 100 (-.58) + 500 = 442$$

$$CEEB = 100 (.66) + 500 = 566$$

De manera directa, se puede convertir cualquier puntuación típica en otro sistema mediante la siguiente fórmula general:

$$\text{Nueva puntuación típica} = \left(\frac{x_a - \bar{x}_a}{s_a} \right) s_n + \bar{x}_n, \text{ donde}$$

x_a = puntuación típica en el sistema anterior

\bar{x}_a = media del sistema anterior

s_a = desviación típica del sistema anterior

\bar{x}_n = media del sistema nuevo

s_n = puntuación típica del sistema nuevo

Así, las puntuaciones $T = 44.2$ y $T = 56.6$ convertidas a puntuaciones z , CI y CEEB serían:

— En el caso de T a z

$$z = \left(\frac{44.2 - 50}{10} \right) 1 + 0 = -.58 \quad (1)$$

$$z = \left(\frac{56.6 - 50}{10} \right) 1 + 0 = .66 \quad (2)$$

— En el caso de T a CI (con $S = 15$)

$$CI = \left(\frac{44.2 - 50}{10} \right) 15 + 100 = 91.3 \quad (1)$$

$$CI = \left(\frac{56.6 - 50}{10} \right) 15 + 100 = 109.9 \quad (2)$$

— En el caso de T a CI (con $S = 16$)

$$CI = \left(\frac{44.2 - 50}{10} \right) 16 + 100 = 90.72 \quad (1)$$

$$CI = \left(\frac{56.6 - 50}{10} \right) 16 + 100 = 110.56 \quad (2)$$

— En el caso de T a CEEB

$$CEEB = \left(\frac{44.2 - 50}{10} \right) 100 + 500 = 442 \quad (1)$$

$$\text{CEEB} = \left(\frac{56.6 - 50}{10} \right) 100 + 500 = 566 \quad (2)$$

Puntuaciones típicas normalizadas

La distribución de las puntuaciones directas o brutas, al igual que la distribución de las puntuaciones típicas, que, como acabamos de ver, se obtienen de manera lineal a partir de aquéllas, se ve afectada por determinadas características del test como puedan ser el nivel de dificultad de sus ítems o el grado de intercorrelación entre ellos. Para solucionar este inconveniente es aconsejable convertir estas puntuaciones en otro tipo de puntuaciones cuya distribución sea independiente de las características singulares del test, con lo cual se facilita la comparabilidad de las puntuaciones obtenidas en tests distintos. Por lo general, las puntuaciones más utilizadas en este sentido son las puntuaciones típicas normalizadas (z_n), que se distribuyen de acuerdo con el modelo de la curva normal. Ahora bien, esta transformación a puntuaciones z_n , que es no lineal, sólo se recomienda cuando la muestra de referencia se distribuye en forma próxima a la normal o cuando se presupone que la variable en cuestión se distribuye en forma próxima a la normal en la población. [En cualquier caso, siempre conviene comprobar el ajuste de la distribución empírica al modelo teórico de distribución normal, por ejemplo, mediante el estadístico chi-cuadrado]. Evidentemente, a mayor similitud entre la distribución empírica y la distribución normal, más parecidas serán las puntuaciones z y sus correspondientes z_n .

La conversión de una puntuación directa a una puntuación típica normalizada requiere, en primer lugar, su transformación a rango percentil y, en un segundo paso, la obtención de la puntuación típica que corresponde a dicho percentil en la distribución normal.

Supongamos la puntuación directa 43 del ejemplo del Cuadrado 3.9. Como se vio en dicho ejemplo, a la puntuación directa 43 le corresponde un percentil de 89, que, como allí se dijo, indica que la puntuación 43 deja tras de sí al 89% de los casos. Consultando en la tabla de áreas y ordenadas de la distribución normal en el Apéndice de este Anexo, en la columna 3 (área de la parte mayor) se busca el porcentaje más próximo al 89%. [Nótese que en la tabla, en lugar de en porcentajes, las áreas vienen expresadas en términos de proporción, con lo que habrá que

buscar la proporción más próxima a .8900]. En este caso, el porcentaje más aproximado es el 88.88%; esto es, la proporción .8888. Luego la puntuación típica normalizada (columna 1) que le corresponde es 1.22. Hasta la ordenada que corta a la curva normal a una distancia 1.22 desviaciones típicas a la derecha, a partir de la media, están contenidos el 88.88% de los casos en la distribución normal.

La z_n correspondiente a la puntuación directa de 43 es, por consiguiente, 1.22.

A partir de las puntuaciones típicas normalizadas (z_n) se pueden calcular puntuaciones típicas normalizadas derivadas de aquéllas. Entre las de mayor uso se encuentran las *puntuaciones equivalentes de curva normal* y las *estaninas*.

Las *equivalentes de curva normal* (ECN) se distribuyen con media de 50 y desviación típica de 21.06. Su obtención a partir de puntuaciones z_n se realiza del siguiente modo:

$$\text{ECN} = 21.06 (z_n) + 50$$

De esta forma, la anterior $z_n = 1.22$ daría lugar a la siguiente ECN:

$$\text{ECN} = 21.06 (1.22) + 50 = 75.69$$

Las *estaninas* (denominadas así a partir de la contracción de las palabras inglesas *standard* y *nine*), también llamadas *eneatipos*, son puntuaciones que se expresan mediante un solo número entero del 1 al 9 y se distribuyen con media de 5 y desviación típica de 2. Su obtención a partir de puntuaciones z_n se realiza del siguiente modo:

$$\text{Estanina} = 2 (z_n) + 5$$

Así, la anterior $z_n = 1.22$ se convertiría en la siguiente estanina:

$$\text{Estanina} = 2 (1.22) + 5 = 7.44, \text{ que se sitúa en el eneatipto 7.}$$

En el Cuadro 3.10 aparecen los límites en z y percentiles y el porcentaje de distribución normal correspondiente a cada estanina.

Equivalencia entre puntuaciones derivadas

Las diferentes puntuaciones derivadas son, sencillamente, diferentes formas de representar un mismo rendimiento de un sujeto en un test. En el Cuadro 3.11 se

Cuadro 3.10

Percentiles, z y porcentajes de distribución normal correspondientes a las estaninas

Estaninas	Percentiles	z	% Distribución normal
1	< 5	< -1.75	4%
2	5-11	-1.75 a -1.25	7%
3	12-22	-1.25 a -.75	12%
4	23-39	-.75 a -.25	17%
5	40-59	-.25 a .25	20%
6	60-76	.25 a .75	17%
7	77-88	.75 a 1.25	12%
8	89-95	1.25 a 1.75	7%
9	>95	> 1.75	4%

ofrece la equivalencia aproximada entre algunas de las puntuaciones derivadas, anteriormente revisadas, tomando como base la distribución normal. En concreto, aparecen las equivalencias entre rangos percentiles; puntuaciones z; puntuaciones T; CI de desviación con desviación típica de 15 –que fue introducido en las escalas Wechsler (ver en Anaya, 2002) y es el más utilizado–; CI de desviación con desviación típica de 16 –que es utilizado en algunas pruebas como la Stanford-Binet–; puntuaciones CEEB; y estaninas.

Como hemos visto, cualquier puntuación derivada puede convertirse en otro tipo de puntuación derivada. Sin embargo, las conversiones más utilizadas son entre puntuaciones típicas (fundamentalmente CI de desviación) y rangos percentiles, con lo que se consigue aunar las ventajas técnicas de las puntuaciones típicas con la fácil interpretación, sobre todo por parte de personas no expertas, propia de los percentiles. A este propósito, en el Cuadro 3.12 se ofrece la equivalencia entre CI de desviación, con desviación típica de 15 o de 16, y rangos percentiles. Este Cuadro se puede utilizar para convertir puntuaciones típicas de cualquier prueba con una media de 100 y una desviación típica de 15 o de 16 a rangos percentiles. Si, por ejemplo, un sujeto obtiene un CI de 121 en una prueba con media de 100 y desviación típica de 15 –WISC-III, pongamos por caso–, le corresponderá un rango percentil de 92. Si, por ejemplo, otro sujeto obtiene un CI de 96 en una prueba con media 100 y desviación típica de 16 –SB:FE, pongamos por caso–, le corresponderá un rango percentil de 40.

Cuadro 3.11

Equivalencia entre diversas puntuaciones típicas tomando como base la distribución normal

