

PSICOMETRÍA APLICADA

Morales, María Luisa

**Editorial Trillas
México, 1990**

ÍNDICE GENERAL

Prólogo	4
1. Aplicación de la teoría psicométrica a los diversos campos de la psicología	6
La medición en psicología	6
Importancia de la medición psicológica.....	7
Instrumentos de medición psicológica.....	9
Estructura de los instrumentos de medición psicológica	11
Evaluación de las pruebas psicológicas	12
Clasificación de las pruebas psicológicas	16
Aplicación de los tests a las diversas áreas de la psicología	22
2. Unidades de medición psicológica y normas de calificación	31
Unidades de medición psicológica.....	31
Unidades de calificación	33
Comparación de un estándar absoluto o dificultad de contenido.....	34
Comparación interindividual.....	34
Comparación intraindividual (de dos medidas del individuo).....	34
Normas de calificación.....	49
Articulación de las normas.....	51
3. Aplicación de las pruebas psicológicas a la medición de la inteligencia	52
Prueba de Stanford-Binet.....	57
Modernas aproximaciones al concepto de la inteligencia.....	64
Clasificación de la inteligencia	76
Constitución de las escalas de Wechsler.....	79
Descripción de la escala de inteligencia Wechsler	80
Descripción de los subtests de la escala de Wechsler	82
Escala de ejecución	85
Información adicional acerca de las escalas de Wechsler.....	88
Estandarización de la escala de inteligencia Wechsler para adultos y para niños, en la ciudad de México.....	90
4. Medición de habilidades múltiples	92
Concepto de habilidad.....	92
Medición de multihabilidades.....	95
Tests y baterías de tests.....	95
Medición de las habilidades especiales.....	99
Inventarios de intereses, valores y preferencias.....	102
Comentarios acerca del empleo de los tests de habilidades múltiples y de habilidades específicas en México	104
Medición de las habilidades especiales en el ámbito escolar.....	105

5. La medición psicológica aplicada al desarrollo del niño en sus diversas etapas: infante, preescolar y escolar.....	106
Desde el nacimiento hasta los quince años	106
El desarrollo físico	110
Medición y evaluación de la evolución de la conducta	112
Medición del desarrollo psicológico del infante	113
Medición del desarrollo psicológico del preescolar.....	117
Consideraciones generales	125
Utilidad del dibujo en la medición del desarrollo psicológico del niño.....	141
Medición del desarrollo psicológico en la etapa escolar.....	144
Medición de la inteligencia en la etapa escolar.....	145
6. La medición en el campo de la psicología clínica	157
Los tests aplicados al campo de la psicología clínica	160
Métodos para medir los rasgos de la personalidad	163
Tests para medir rasgos de la personalidad mas utilizados en México.....	164
Inventarios de personalidad	170
Consideraciones finales acerca de la medición y de la psicología clínica	181

PRÓLOGO

Lanzarnos a la aventura de escribir este libro de texto ha sido el resultado de las experiencias vividas, primero como alumna y luego como profesora de la materia ahora denominada psicometría, que forma parte del currículo académico que todo estudiante de la carrera de psicología deberá estudiar en algún momento de su preparación universitaria.

Nuestro propósito es ofrecer a los estudiantes un texto útil, sencillo, escrito de manera directa en nuestra lengua y que presenta la recopilación de la información, el aprendizaje práctico en sus diversas aplicaciones y la experiencia obtenida sobre todo en el campo de la docencia. Puede decirse que desde la fundación de esta carrera en México, el estudiante de psicología se ha visto en la necesidad no sólo de leer, sino de estudiar y entender los conceptos básicos en un idioma extranjero, generalmente el inglés, tanto en psicología como en todas las materias que comprende la carrera psicológica. Esto provoca que muchas veces, por la complejidad de los temas, al estudiarse en otro idioma no sólo sean más difíciles, sino con frecuencia se distorsionan, ocasionando aprendizajes equivocados.

Por otra parte, en México, las traducciones de los libros dedicados a la psicometría en especial son muy pocos, y su contenido resulta hasta cierto punto obsoleto. Exponen una gran cantidad de material que ha de ser aprendido por estudiantes y que en realidad no les será de utilidad cuando tengan que desenvolverse en el campo profesional. Presentan un gran número de tests no sólo desconocidos y de inútil aplicación en nuestro país, sino difíciles de adquirir con el fin de traducirlos, “adaptarlos y estandarizarlos” para el medio mexicano, actividad que, por otra parte, se encuentra bastante descuidada a pesar de ser una de las fuentes más importantes de trabajo en la psicología mexicana.

Este libro presenta exclusivamente la aplicación de la teoría de la medición psicológica a los diversos campos de aplicación de nuestra ciencia en México. Con esto queremos indicar que no estamos impugnando ni discutiendo los fundamentos teóricos de la medición psicológica en general, sino simplemente intentamos llevar estos principios a su máximo rendimiento y utilidad dentro del marco de referencia en el que ha de desarrollarse el trabajo del psicólogo mexicano.

El orden en que se presentan los capítulos de este libro está basado en nuestra propia experiencia en relación con lo que consideramos como fundamental jerárquicamente; es decir, estamos dando por sentado que el lector ha estudiado, al menos, el aspecto primario o curso anterior (es decir, la “teoría de la medición psicológica”) para que el contenido de este texto pueda comprenderse. El capítulo 1 está integrado por aquellos elementos que van a ser manejados en el uso de todo instrumento o test psicológico, tales como la aplicación de los tests psicológicos a los diversos campos de la psicología, las unidades psicométricas de medición, las tablas de normas de calificación y, esencialmente, la enumeración y explicación somera de los pasos que todo psicólogo debe observar al “evaluar” un test para sus objetivos específicos. Como sabemos, en cualquier tipo de test psicológico estos elementos son continuamente mencionados.

La distribución del empleo de los tests a lo largo del texto tiene el propósito de establecer cierto orden que vaya de aspectos generales a particulares; por ejemplo, se menciona en primer lugar la medición de la “inteligencia” en general, para hablar posteriormente de la medición de la inteligencia en las diversas etapas del desarrollo psicológico.

Además, cabe aclarar que los tests presentados y explicados aquí, son de tres formas: a) aquellos que, aún siendo de importación, se utilizan profusamente en nuestro país; b) aquellos que han sido real y seriamente estandarizados en México o que se encuentran en proceso, y c) aquellos contruidos especialmente para utilizarlos en algunos campos de la medición, tanto en la investigación como en situaciones particulares. Obviamente, si hemos omitido mencionar algunos tests o estudios relacionados con éstos, ha sido accidental, sobre todo a causa de la deficiente comunicación de estos trabajos.

Vaya pues este trabajo a todas aquellas personas interesadas en el campo de la psicometría, esperando que sea de utilidad y con la absoluta decisión de realizar las correcciones pertinentes cada vez que éstas lleguen a nuestras manos.

María Luisa Morales.

1

APLICACIÓN DE LA TEORÍA PSICOMÉTRICA A LOS DIVERSOS CAMPOS DE LA PSICOLOGÍA

LA MEDICIÓN EN PSICOLOGÍA

La Psicometría tiene como finalidad llevar a cabo la medición de la conducta, tanto en el ser humano como en los animales, lo que constituye uno de los pilares fundamentales de la ciencia psicológica, ya que el hombre cada día se interesa más por comprender su propia naturaleza y, por ende, la de los demás; puesto que en el mundo en que vivimos, los problemas que se plantean son primordialmente de naturaleza humana e impiden a menudo una trayectoria más sencilla en el desarrollo de los seres humanos.

En el desenvolvimiento de la psicología como ciencia, ha sido necesario realizar estudios a fin de poder determinar si difieren los seres humanos entre sí y en qué grado se presentan estas diferencias, lo que ha originado, desde hace aproximadamente un siglo, el desarrollo de la instrumentación adecuada para evaluar cuantitativamente las diferencias y las semejanzas existentes entre los individuos. Aún cuando se ha expresado que los psicólogos deben estar más interesados en las diferencias individuales que en las constantes de las personas, es necesario, a veces tener en cuenta cuáles atributos aparecen en mayor o menor grado, consistentemente, en sujetos o grupos diferentes.

Desde luego, haremos notar que los atributos, características, habilidades o rasgos que generalmente intentamos medir no lo hacemos de manera directa, sino por medio de la expresión conductual externa y observable. No podemos medir directamente las aptitudes matemáticas de un individuo; sólo podremos hacerlo cuando le solicitemos que resuelva cierto número de problemas, mediante los cuales el psicólogo constructor de instrumentos de medición sabe que puede hacerlo.

La ciencia psicológica se define a través de teorías psicológicas, las cuales no son más que un conjunto de afirmaciones acerca de una serie de conceptos que nos permiten describir algunos aspectos, hasta cierto punto limitados, del universo conductual. Frecuentemente construimos modelos formales que estructuran los conceptos generales de la teoría, que explican parcialmente la misma y que están relacionados de manera sistemática con fenómenos observables. Estos modelos tienen como propósito fundamental realizar la deducción lógica de las relaciones específicas y generales que no han sido empíricamente demostrables, pero que pueden serlo.

La psicología tiene una tarea esencialmente explicativa, común a todas las ciencias, pero más arraigada en ésta debido a su etapa “un tanto inicial” de desarrollo.

Cuando consideramos una regla de correspondencia para relacionar algún concepto precientífico mediante datos observables, de hecho estamos intentando obtener una explicación del concepto original.

Una de las funciones de toda teoría de la medición consiste en emplear construcciones hipotéticas, y las desviaciones de la medida de esos valores hipotéticos constituyen la explicación de “eventos paradójicos”.

La medición en psicología, como en cualquier otra ciencia, se inicia con la identificación de elementos del mundo real con los elementos o construcciones de un sistema lógico abstracto, al cual llamamos modelo; por ejemplo, el concepto de estatura de una persona se define fácilmente, teniendo en cuenta tres situaciones: a) identificación del objeto que va a ser medido, es decir, la persona (o la unidad experimental); b) identificación de las propiedades o conducta que va a ser medida (en este caso, la distancia que hay de los pies a la cabeza en posición erecta), y c) identificación de las reglas mediante las cuales asignamos un número a esta propiedad que va a ser medida (el metro).

A menudo, la identificación de elementos del mundo real con los del sistema matemático puede hacerse fácilmente, como ocurre en el caso de los “volados” (lanzar una moneda al aire) en donde puede caer “águila o sol” o bien “cara o cruz”, descartando, desde luego, la posibilidad de que caiga de canto; asociamos estas dos posibilidades definiéndolas como A y B, las cuales son mutuamente exclusivas ya que sólo es factible la relación de una de estas posibilidades. Podemos decir entonces que estas “designaciones simbólicas” A y B constituyen una medida; de aquí se deriva un concepto amplio y generalizado de lo que entendemos por medición: *es el procedimiento mediante el cual asignamos números (calificaciones, medidas) a las propiedades, atributos o características de los objetos (o de unidades experimentales), estableciendo las reglas específicas sobre las que se fundamentan tales asignaciones*. El procedimiento de medición especifica una regla de correspondencia, un patrón a partir del cual a estos eventos les corresponderá un subconjunto de números reales.

Las construcciones de la teoría psicológica están expresadas por medio de palabras; no obstante, tienden a traducirse en significados connotativos. Los modelos matemáticos difieren de los modelos puramente verbales en que aquéllos emplean sistemas matemáticos exactos generalmente de alto orden (álgebra o cálculo) mediante los cuales las construcciones elementales pueden manipularse a fin de facilitar las deducciones del modelo. Estas construcciones, aún cuando son más explícitas, tienden a tener menor significado connotativo y probablemente menor riqueza conceptual; por tanto, puede decirse que su utilidad actualmente se encuentra limitada en sus aplicaciones al universo conductual.

Las construcciones teóricas se relacionan con el universo conductual mediante variables que se consideran como medidas o indicadores resultantes de observaciones introspectivas.

IMPORTANCIA DE LA MEDICIÓN PSICOLÓGICA

Ahora bien, si partimos de la base de que lo que existe, cualquiera que sea su naturaleza, lo es en cierta cantidad, debemos estar plenamente interesados en ello, es decir, en medir en qué cantidad los atributos que se han descrito como característicos de los seres humanos (hay que

recordar que no se mide al hombre, sino a sus atributos) existen en cada uno y cuál sería la norma de cantidad en los diversos grupos humanos que forman nuestro mundo.

Actualmente, los psicólogos, siguiendo el pensamiento kantiano, están de acuerdo en la extensión con que los fenómenos psicológicos están a disposición de la observación y la medición, ya que la conducta humana está conformada de acciones, juicios, palabras, intereses, valores personales y culturales, hábitos, costumbres, etcétera. La naturaleza propia de todos estos aspectos del ser humano hacen un tanto difícil su investigación científica, ya que mientras pueden ser observables, el problema es menor; pero en cuanto surgen problemas en los que se encuentran implicados sensaciones, pensamientos e imágenes, lo “científico” de la investigación se complica. El interés de los científicos se centra en *el mundo de los hechos*; es decir, en aquello que se pueda ver, tocar, oír o expresar de alguna otra forma que sea observable por medio de los sentidos.

De este modo, los psicólogos, como investigadores científicos, han dirigido su atención al estudio de aquellos eventos que pueden ser evidentes si se siguen las reglas fundamentales que la ciencia ha adoptado y que se han demostrado a lo largo de su desarrollo desde hace varios siglos.

La naturaleza de los datos psicológicos es peculiar y difícil de manejar porque sólo se cuenta con la expresión verbal o motora-verbal del individuo que nos informa acerca de su experiencia personal. De tal suerte, con dichos datos recopilados trabaja el investigador psicológico, mismos que se restringen al hacer o al decir del hombre. Sin embargo, es demasiado lo que se puede aprender del comportamiento del individuo y, al respecto, existen muchas formas en las que puede hacerse una predicción de la conducta humana.

Debemos establecer aquí claramente una cuestión que ha sido discutida frecuentemente por los estudiosos de la psicología: el hecho de que ésta, como ciencia, persiga entre otras finalidades la *predicción* no implica la limitación de las actividades del hombre, ni su libertad de “ser”.

Asimismo, la ciencia psicológica, como tal, no constituye una cuestión de fe, ni es una religión o un credo, sino que tiene como base el método científico y la experimentación, y parte de éste le permite llegar a conclusiones y, por tanto, a establecer leyes acerca del comportamiento, lo cual también le permite abordar la investigación acerca de la religión y las creencias del ser humano, así como sus problemas éticos y morales. Sin embargo, en este punto no puede determinar cuál creencia es la mejor o la peor, si es “buena” o “mala”.

Como en todo estudio científico, en el campo de la psicología debe haber diversas formas de medición; así, con excepción de la precisión que hay en las matemáticas, resulta difícil hacer afirmaciones absolutas, ya que toda ciencia se tiene que auxiliar de ellas para probar sus hipótesis y darlas por verdaderas.

Como sabemos, la asignación de números a los atributos de los objetos se deriva de lo que denominamos *niveles de medición*, mismos que se diferencian por su grado de complejidad de los sistemas matemáticos que utilizan y las transformaciones matemáticas que admiten los datos para ser manejados en la obtención de resultados comprensibles.

A manera de recordatorio, aquí simplemente indicaremos que esos niveles de medición, llamados con más propiedad *escalas de medición*, son elementos que se emplean frecuentemente, los

cuales se clasifican desde la medida más elemental hasta la más compleja, de la siguiente manera: a) escalas nominales (o de *categorización*); b) *escalas ordinales*; c) *escalas intervalares*, y d) *escalas de razón*.

De estas cuatro escalas, en el campo de la medición psicológica sólo se pueden utilizar las tres primeras, ya que la última (la escala de razón), por partir del cero absoluto, no es posible emplearla, puesto que sabemos que el universo que medimos en psicología no parte de la nada, sino que siempre se inicia en “algo”.

INSTRUMENTOS DE MEDICIÓN PSICOLÓGICA

Cuando hablamos de los instrumentos de medición en psicología, nos referimos a todos aquellos procedimientos u operaciones que permitan llegar a obtener, objetivamente y con la mayor certeza posible, información acerca de la expresión de los fenómenos que se suceden en esa unidad biológica, social, psicológica, etcétera, que es la conducta humana. Estos instrumentos, contruidos para medir aspectos específicos de la conducta humana (y animal según las necesidades que plantee la situación de que se trate: estudios clínicos, escolares, experimentales propiamente dichos, etc.), se conocen con el nombre de *pruebas psicológicas*.

La necesidad de obtener, cuantitativamente, los atributos o las características de un segmento de conducta, al ser observada por quienes se interesan en comprender y predecir el comportamiento de los seres humanos, surgió realmente durante la primera mitad de este siglo. Claro que si nos remontamos a la historia tendremos que dar el crédito correspondiente a Aristóteles y Platón, quienes, en realidad, son los pioneros de la medición psicológica. Sin embargo, lo cierto es que alrededor de 1830 se llevan a cabo los primeros estudios de este tipo en campos fundamentales de la psicología y las matemáticas destinados a lograr la medición mental, distinguiéndose tres corrientes de influencia preponderantes: la *francesa*, con Esquirol, Pineal, Itard, Seguin, Charcot y Ribot; la *alemana*, con Weber, Fechner y Wundt, y la que implantó Laplace en el campo de las matemáticas, quien, junto con Gauss, demostró la aplicabilidad y utilidad de la curva normal (ya enunciada en 1733 por Moivre) como un instrumento matemático aplicable a la medición de las características del ser humano.

Se ha discutido extensamente a quién puede atribuírsele el nombre de “padre de la medición mental”, lo cual, aún cuando no es de primordial importancia, sí es en cierta medida necesario para establecer una explicación justificada y apoyada en los hechos, del porqué los primeros instrumentos fueron contruidos de cierta manera y cómo ello ha influido hasta la actualidad, en la elaboración y utilización de dichos instrumentos. Parece ser que se hacen acreedores a tal designación tanto sir Francis Galton (1822-1911) como James McKeen Catell (1860-1944) y Alfredo Binet (1857-1911). Lo verdaderamente importante es que los tres personajes han contribuido considerablemente al desarrollo de este campo. Sin embargo, dado el momento en que históricamente aparecen los primeros intentos por lograr la medición de la conducta humana, el mérito corresponde a sir Francis Galton, quien, influido fundamentalmente por Darwin, publicó su famosa obra *El genio heredado* y poco tiempo después su libro *Investigación del desarrollo de las facultades humanas*, en el cual de inmediato podemos encontrar como contenido esencial lo que hemos llamado principios de los estudios científicos acerca de las diferencias individuales y sus primeros intentos por medirlas.

En 1886, Galton creó el primer laboratorio antropométrico destinado a la medición de las características físicas y sensoriomotoras de los seres humanos; y a través de todos sus trabajos estableció los principios del empleo del método estadístico como un auxiliar importante para obtener medidas más objetivas y precisas en sus investigaciones. Utilizó fundamentalmente el concepto de correlación, que posteriormente habría de ser desarrollado plenamente por Karl Pearson (1857-1936). Se inicia, de esta manera, el estudio científico de las diferencias individuales y la evaluación de las pruebas mentales, apoyado en la instrumentación matemática.

Por otra parte, Hugo Munsterberg, en 1891, diseñó pruebas específicas para niños, y fue el primer psicólogo que pensó en utilizar el registro del tiempo de ejecución de las tareas que se imponían con sus instrumentos, a sus sujetos de investigación. (No se informó del resultado de sus investigaciones).

Investigadores como Pearson, Charles Spearman, Cyril Burt, Goedffrey Thomson, en Inglaterra, y Kelley, Thurstone y Guilford y colaboradores, en América del Norte, estructuraron y fundamentaron el empleo de la inferencia estadística y de otros conceptos cuantitativos que han venido a integrarse al desarrollo y aplicación de la medición psicológica en la actualidad.

En 1895 un discípulo de Kraepelin, Oehr, utilizó en sus estudios el conteo de letras, el tachar cierta letra en algún párrafo, el buscar y encontrar errores y la memorización de dígitos y de sílabas sin sentido, siendo tal vez el primer psicólogo que supuso que, estableciendo correlaciones entre los resultados que obtenía con esas pruebas en los sujetos que investigaba (quizá pacientes psiquiátricos), podría obtener una mayor información y, por tanto, planear un mejor tratamiento terapéutico en ellos.

En 1897, Hermann Ebbinghaus (llamado el “padre de la psicología experimental del aprendizaje”), al ser designado para estudiar a niños en edad escolar en relación con el problema de la fatiga, utilizó tres pruebas, que más tarde habrían de servir como base de muchas otras, más elaboradas y mejor construidas, tales como la del cálculo rápido, la de memoria de dígitos y la de completamiento de oraciones.

Clark Wissler, en 1901, quien habría de responsabilizarse de la validación de las pruebas de Catell para la selección de estudiantes de nivel universitario, trató de determinar, mediante la aplicación de correlaciones, cuándo dos instrumentos medían las mismas (si eran positivas) o diferentes (si eran negativas) habilidades o cuál era su grado de correlación.

Por otra parte, Sharp, siguiendo el mismo método de Wissler e intentando descubrir la naturaleza del “pensamiento humano”, realizó una serie de estudios, elaborando tests complejos y tests simples, a fin de determinar si la complejidad o simplicidad de los instrumentos podría constituir un factor importante en los resultados. Sus conclusiones fueron en el sentido de que cada prueba medía diferentes funciones y, por tanto, dependía de las necesidades o de los objetivos de los estudios el empleo de un tipo u otro de tests.

Mientras Galton trabajaba en Inglaterra, Catell, en los Estados Unidos (utilizando la información experimental que Wundt obtuvo en su laboratorio de Leipzig, Alemania), desarrolló instrumentos de medición psicológica, a los cuales les agregó el clásico pragmatismo norteamericano. Catell siguió la escuela de Wundt y aprendió de él que es menester ejercer un

riguroso control al llevar a cabo las observaciones de la conducta; sin embargo, rechaza el concepto de que la meta fundamental en psicología sea la generalización de los hallazgos acerca de los fenómenos psicológicos ocurrientes en los seres humanos, y considera, sobre todo, las diferencias individuales en la ejecución de la conducta como variables principales en los estudios científicos.

Años después, Catell, en contacto con Galton, estableció laboratorios de psicología en las universidades de Pensilvania y Columbia, en donde se analizó exhaustivamente el desarrollo y depuración de los instrumentos de medición psicológica, a los cuales Catell designó con el nombre de *tests mentales* (1890). Estos instrumentos fueron diferentes de los que conocemos y utilizamos hoy día, ya que habían sido creados sólo para la medición de capacidades sensoriales y otras funciones simples, teniendo en cuenta los puntos de vista (epistemológicos) del empirismo inglés, que afirmaba que el conocimiento se obtenía mediante la experiencia; de tal suerte, las medidas relevantes fueron, fundamentalmente, de tipo sensoriomotriz: tiempo de reacción, capacidades auditivas y visuales, sensibilidad al dolor, al gusto, etcétera.

En 1896, Alfred Binet, en Francia como psicólogo experimental, se interesó más en el estudio de aquellos individuos que presentaban desviaciones conductuales, particularmente los niños. Después de minuciosos estudios en niños retrasados, normales y avanzados, Binet y Víctor Henri crearon una serie de instrumentos destinados a la medición de “procesos mentales muy complejos”. Posteriormente, analizaron una serie de pruebas a fin de determinar si era más adecuado pronosticar el potencial educativo de los niños mediante los juicios de los profesores cuando aquéllos tenían más edad, que cuando eran pequeños.

En 1904, comisionados por la Secretaría de Educación Pública del gobierno francés para estudiar la educación de los niños retrasados, Binet y Simon desarrollaron la primera investigación formal para llevar a cabo la medición de habilidades complejas, la cual, en contraste con la de Catell, fue más prometedora en cuanto a los datos que con ella podían obtenerse, al grado de que en la actualidad aún se sigue utilizando, naturalmente después de haber pasado por múltiples revisiones y estudios que la han clasificado como un instrumento útil, predictor del comportamiento intelectual.

ESTRUCTURA DE LOS INSTRUMENTOS DE MEDICIÓN PSICOLÓGICA

Antes de seguir adelante, es necesario explicar qué se entiende por instrumento de medición psicológica; sin embargo, esto constituye, hasta cierto punto, un problema, que a través de la historia del desarrollo de las pruebas psicológicas ha sido tema de innumerables polémicas por parte de los estudiosos de la psicología.

Como una primera observación que debe tenerse en cuenta al revisar los diversos conceptos que se han expuesto sobre la definición de instrumento de medición psicológica, consideramos que ésta depende, en la mayoría de los casos, del criterio que el autor o autores adopten para establecer dicho concepto. Otra observación consiste en que, dado que ciertos aspectos de la conducta humana han sido más accesibles a la observación de los investigadores (como la percepción, las sensaciones, los aspectos motores, y ese concepto que implica una infinidad de funciones u operaciones de la conducta humana que ha sido denominado *inteligencia* y que en capítulos subsecuentes intentaremos analizar con mayor extensión), el concepto de prueba psicológica varía aunque, analizando cada uno de esos aspectos, podemos encontrar que hay factores comunes que permiten sintetizar y establecer

una definición general que podría enunciarse de la siguiente manera: *un instrumento de medición psicológica es aquella técnica metodología producida artificialmente, que obedece a reglas explícitas y coloca al o a los individuos en condiciones experimentales, con el fin de extraer el segmento del comportamiento a estudiar y que permite la comparación estadística conductora a la clasificación cualitativa, tipológica a cuantitativa de la característica o características que se están evaluando.*

Simplificando esta definición, podríamos utilizar para fines prácticos la establecida por Pierre Pichot: *una prueba psicológica es una situación experimental y estandarizada que sirve de estímulo a un comportamiento y que, manipulada estadísticamente, permite clasificar a los sujetos tipológica o cuantitativamente.*

Es útil, también, mencionar la definición dada por Lee J. Cronbach: *una prueba psicológica es una técnica sistemática que compara la conducta de dos o más personas.* Esta definición puede ampliarse para incluir mediciones mediante aparatos, técnicas de laboratorio para observación de respuestas en el campo del comportamiento humano, cuestionarios de todo tipo, etcétera. Es necesario hacer énfasis en que, para nosotros, un instrumento de medición psicológica no se restringe sólo a aquellos casos que han sido considerados como “tests mentales” (psicológicos).

La mayoría de las pruebas psicológicas que se tratarán en este libro están constituidas por una serie de elementos llamados estímulos, reactivos (y por deformaciones del lenguaje, frecuentemente ítems), reunidos en una sola cédula, a la cual le llamamos *prueba* (test) si se utiliza como único instrumento de medida, o *subprueba* (subtest) si forma parte de lo que en medición psicológica denominamos batería o escalas de medición, porque están constituidos por más de dos pruebas que evalúan diversos aspectos o características del comportamiento en una mismas cédula.

La teoría de la medición que hemos estudiado abarca todos los tipos de pruebas psicológicas construidas para medir aquellas características que son de interés para el psicólogo. Cada reactivo, prueba o subprueba produce una medida que, dependiendo de cualquiera de los casos, se adiciona o sintetiza para obtener una medida total, *compuesta* de tantas partes como las mediciones que la *componen*; es decir, las características de una prueba compuesta están en función de las características de sus componentes, y el fin primordial de la teoría de una prueba es determinar el tipo y número de componentes que deben combinarse para construir “un compuesto de propiedades específicas”.

EVALUACIÓN DE LAS PRUEBAS PSICOLÓGICAS

Debemos hacer notar que al hablar de evaluación de las pruebas psicológicas nos estamos refiriendo a los criterios que el psicólogo, el psicómetra y el constructor de instrumentos de medición han de tener presentes para juzgar y determinar si el o los instrumentos que van a utilizar para sus fines específicos son apropiados para la situación particular; para esto, indicaremos algunos de esos criterios, los cuales deberán ser tomados en cuenta cuando se seleccione una prueba y se desee ser rigurosamente estricto, desde cualquier punto de vista, incluyendo el aspecto ético profesional.

En la práctica diaria, hemos podido observar lo inadecuado de las elecciones que a menudo hacen algunos psicólogos cuando se les plantea una situación específica para la cual tienen que dar

resultados válidos y confiables. Frecuentemente se seleccionan instrumentos impropios para solucionar un problema dado, sea individual o de grupo; entonces una de las principales consideraciones de quien utiliza los instrumentos psicológicos es que debe llevar en mente la elección de aquel o aquellos que reúnan los requisitos para el propósito que se persigue; es decir, si la función de una prueba psicológica es proporcionar la mayor y mejor información acerca de un problema, el primer paso a seguir será la especificación cuidadosa de las interrogantes que deberán ser contestadas y el tipo de personas que habrán de ser examinadas. Es necesario, también, considerar varias alternativas, para que en un momento dado, si la situación lo exige, puedan ser utilizadas y se pueden establecer así comparaciones pertinentes que permitan ratificar o rectificar los datos obtenidos, a fin de poder estar seguros de que las decisiones dadas son válidas.

Otra consideración importante consiste en determinar si la capacidad o potencialidad de la prueba es la adecuada para medir con el rigor necesario aquella característica o características que estamos investigando. La longitud de una prueba, en donde desempeña un papel relevante el factor tiempo, constituye otro de los criterios que deben tenerse en cuenta, ya que esto influye tanto en aspectos de esfuerzo como en aspectos económicos, puesto que el administrador de pruebas invertirá horas, estudio, trabajo, etcétera. Todo esto se vería alterado aún más si el examinado, a causa del cansancio, entre otras posibilidades modificara los resultados del examen.

La facilidad de administración y calificación de un instrumento, la disponibilidad de formas paralelas, la forma de adaptación del instrumento, etcétera, son aspectos que también deberán considerarse para su elección.

Es obvio afirmar que para poder hacer una buena elección de la instrumentación, útil en un momento dado, debe poseer conocimientos técnicos quien maneje las pruebas, a fin de establecer si existen los requisitos mínimos para considerar que se ha seleccionado el mejor o los mejores instrumentos. Estos conocimientos se refieren, cuando menos en forma elemental, al grado de confiabilidad y validez, a la objetividad y sensibilidad del instrumento, así como al hecho fundamental de si ha sido estandarizado, o cuando menos adaptado, y en qué grupo o grupos de sujetos se han hecho tales estudios.

No entraremos aquí en la definición y/o explicación de los conceptos precedentes, puesto que en un curso anterior fue visto con amplitud; sin embargo, sí haremos hincapié en una serie de aspectos de vital importancia y que se refieren a aquellos posibles errores de medición que se derivan de la falta de conocimientos e información técnica, y sobre todo metodológica, en el campo de la medición psicológica.

Uno de los principales errores en la evaluación de las pruebas es aquel que se origina en problemas de estandarización. Como se sabe, las medidas psicológicas carecen de un cero absoluto, lo cual garantiza relativamente la igualdad de las unidades a través del rango de la escala. Así, las calificaciones de un individuo no tienen significado sino en función de la ejecución de un grupo de individuos contra el cual se comparaba el individuo mismo (aunque posteriormente veremos que existe una excepción: aquella que se refiere a la medición intraindividual). La calificación de una persona en una prueba sólo tiene significado si se compara con la de otras personas del grupo; es el *error de interpretación* el cual puede deberse a dos situaciones: a) al grupo contra el que ha sido comparado el sujeto en cuestión, y b) a la forma en que se ha expresado la comparación entre el sujeto y el grupo. Por ejemplo, si intentamos comparar las calificaciones de un estudiante de primer

año de preparatoria obtenidas en una prueba psicológica o de rendimiento escolar, con las obtenidas en la misma prueba por alumnos de tercer año, la predicción y la acción que se establezcan serán absolutamente equivocadas, puesto que habrá diferencias significativas en ambas situaciones. Por esta razón, los procesos de adaptación y estandarización que culminan con el establecimiento de tablas de normas o grados de calificación para grupos específicos, deben tenerse en cuenta en la elección y evaluación de la instrumentación en el trabajo psicológico. Existen dos medios para establecer comparaciones individuo-grupo: a) comparar al sujeto con su grupo correspondiente, o b) compararlo con un grupo al cual aspira a formar parte. En el primer caso, el medio es tan sencillo que no podemos clasificar a un individuo, sino sólo cuando tenemos un estándar de su propio grupo. En el segundo caso, si lo que buscamos es predecir el grado de éxito de un estudiante de medicina (por ejemplo, cuando sea médico), habrá que compararlo con los individuos que forman el grupo de estudiantes brillantes, en vez de hacerlo con el grupo de sus compañeros de segundo año; además, existe otro tipo de errores que pueden incluirse en estandarización, la cual trataremos posteriormente: las relaciones entre las calificaciones de un grupo normativo y las del individuo cuando se trata de la conversión de calificaciones crudas o derivadas, como edad mental, cocientes intelectuales, rangos percentiles, calificaciones estándar, etc.

Otro tipo de errores que alteran considerablemente los resultados que se buscan, es el de los *errores variables*, o sea aquellos que provienen de “accidentes” que suceden en el momento de la aplicación de las pruebas; por ejemplo, durante la aplicación de una prueba de velocidad supóngase que se rompe la punta del lápiz; que surgen ruidos que impiden la atención del sujeto; que el examinados da instrucciones incorrectas, etcétera; esto trae como consecuencia que los examinados obtengan calificaciones ligeramente diferentes de quienes realizan la prueba en condiciones apropiadas y, por tanto, distintas. Estos errores se denominan *errores variables*, debido a que se producen por la variación de ejecución de una persona a otra, o porque las calificaciones de una persona son diferentes cada vez que es medida. La relativa libertad para cometer errores en psicometría es conocida como *grado de confiabilidad* de una prueba.

Otro tipo de errores son los llamados *errores personales*, los cuales se refieren a las diferentes formas de comprender, manejar e interpretar las respuestas obtenidas a través de una prueba psicológica, a partir de la posición teórica o metodológica en que se encuentre colocado el psicólogo que aplica las pruebas. En diversos estudios se ha demostrado que una persona puede rectificar sus respuestas si revisa o lee varias veces una prueba; asimismo, diferentes personas que leen una misma prueba, obtienen distintas conclusiones. Estos errores constituyen un tipo de error variable, debido a que quien califica las respuestas es susceptible de variar su criterio de tiempo en tiempo y de trabajo en trabajo; sin embargo, dichas respuestas son de suficiente magnitud para dedicarles especial atención al medir las características de la conducta. Tales características deben ser observadas, medidas e interpretadas a partir de un máximo grado de objetividad. Esta se halla íntimamente relacionada con la estructura del instrumento, su validez y su confiabilidad y, especialmente, con los criterios de calificación, los cuales deben ser establecidos con tal claridad y exactitud que al manejarlos dos o más personas, o al repetir una misma persona las calificaciones en varias ocasiones, no se altere el resultado, es decir, se tenga un alto coeficiente de confiabilidad por parte de los calificadores.

Uno de los errores más comunes y ampliamente desarrollado por los psicólogos es el *error constante*. Este tipo de error es debido a que las medidas de los atributos del individuo son indirectas, es decir, se mide la expresión del fenómeno y no el fenómeno en sí. No podemos abrir el

cráneo de una persona para mirar hacia adentro o colocar en una balanza un poco de materia gris y decir que el peso de ello es la “cantidad” de habilidad mecánica que posee. Lo que hacemos es obtener una medida de “eso” mediante un instrumento, el test psicológico, es decir, la medición de lo que observamos se expresa por medio de la presentación de los estímulos adecuados para extraer la conducta que se desea. Esta situación es extensiva a todo tipo de medición de cualquier fenómeno que nos interese conocer, independientemente de la ciencia que nos ocupe. Así, cuando utilizamos el termómetro, la altura de la barra de mercurio no es lo importante, sino lo que significa esa altura en tanto que muestra la relación con la cantidad de calor del fenómeno que se investiga. La diferencia entre este problema y el de la medición de la conducta es que aquí, el índice indirecto (calificaciones de la prueba) y los rasgos subyacentes en estudio no son, en principio, tan exactos como las cantidades físicas.

Un ejemplo de este problema lo constituye la ejecución de una prueba de lápiz-papel (que indirectamente mide habilidades mecánicas), la cual dependerá muy poco de la perfección que se tenga para leer. En tanto que los resultados reflejen diferencias que se tenga para leer. En tanto que los resultados reflejen diferencias en la habilidad de la lectura, más que en la habilidad mecánica, mayor grado de error habrá en la medida. A diferencia del error variable, mayor grado de error se denomina error constante, debido a que la cantidad de error, en el primer caso, será la misma para todas las personas a quienes se les aplique la prueba, y será el mismo error para una misma persona, cada vez que se repita la medida. Es decir, si la prueba mide el factor numérico en cierto grado, en ese grado medirá el factor numérico para todos los que reciban la prueba cada vez que les sea aplicada. El problema del error constante está íntimamente relacionado con la validez. Es necesario comprobar que la extensión con que se realiza la medición indirecta es realmente un indicador válido del rasgo que se investiga.

Todos los tipos de error mencionados anteriormente, como puede observarse, están relacionados con diferentes características de la prueba. Cada una de ellas representa un criterio independiente para evaluarla. Para decidir si un instrumento psicológico constituye un buen diseño de medición, se necesita examinar detenidamente los siguientes “criterios”:

- a) *Estandarización* que permite determinar el grado de errores de interpretación que han sido controlados, a fin de evitar caer en ellos.
- b) *Confiabilidad* que indica la libertad relativa de error variable.
- c) *Objetividad* que denota el grado en que los errores personales han sido evitados.
- d) *Validez* que indica el grado en que la prueba mide aquellos fenómenos para los cuales ha sido construida; por tanto, no está influida por el error constante.

Es necesario hacer notar que una prueba no es “buena” solamente por estar exenta de estos cuatro tipos de error. Puede ocurrir que una prueba sea poco confiable a pesar de estar bien estandarizada y, por lo contrario, puede estar mal estandarizada y ser altamente confiable, y ser poco útil por falta de validez. La falta de confiabilidad reflejará falta de objetividad, fundamentalmente cuando ésta se refiere a la calificación. La validez requiere un mínimo de objetividad, confiabilidad y estandarización. Una prueba falta de validez es absolutamente inútil, a pesar de que sea objetiva, confiable y esté bien estandarizada.

Como se indicó, la calificación cruda, ya sea que represente el número de reactivos correctos, el tiempo utilizado para realizar una tarea determinada, etcétera, será de poca significación

mientras no sea comparada con las calificaciones obtenidas por el mismo grupo de estandarización. Estas calificaciones crudas deben ser transformadas en calificaciones derivadas. Para comparar la calificación de un individuo con las obtenidas por los individuos de un grupo estandarizado, es necesario primero sintetizar la ejecución del grupo. Generalmente esto se efectúa elaborando una distribución de frecuencias de las calificaciones crudas mediante los procedimientos estudiados en la estadística descriptiva elemental. En ocasiones, la distribución, encontrada generalmente en la medida de las características humanas, se halla cerca de lo expresado por la curva normal de distribución de frecuencias, o sea la curva de Gauss. Existen tres excepciones a esta regla: a) que la prueba sea fácil o difícil de ejecutar para el grupo en cuestión; b) que parte del grupo tenga entrenamiento previo o haya sido seleccionado en relación al rasgo en cuestión, a diferencia del resto, y c) que el rasgo considerado desempeñe un importante papel en la estructura social, lo cual puede ser determinante para los resultados.

En el primer caso podemos encontrar una distribución oblicua; si la prueba es fácil, las calificaciones se acumularán en la parte alta de la escala, y si es difícil, se encontrarán en la parte baja. Suponiendo que las habilidades primordiales o los rasgos estén normalmente distribuidos, se puede *normalizar la distribución* cuando se obtienen calificaciones derivadas para una interpretación más comprensible y fácil de los resultados. Sin embargo, es necesario revisar los resultados que se obtienen mediante el análisis de reactivos, a fin de establecer el adecuado nivel de dificultad que debe contener la prueba.

En la segunda hipótesis, el resultado puede ser una distribución bimodal, en cuyo caso es necesario considerar el por qué de esta situación, no olvidando la combinación de sujetos en la estandarización. En este caso lo mejor es separar cada uno de los grupos y trazar su curva de distribución por separado, así como establecer sus normas también independientemente.

En la tercera situación, la curva que se obtendrá puede ser altamente oblicua o en forma de *J*, debido a que las observaciones del fenómeno que se está midiendo serán variables, dado que se trata de situaciones de presión social; por ejemplo, la señal de luz roja en la circulación de tránsito (alto) indica que las personas deben detenerse totalmente; la mayoría ejecutará esta conducta; de esta manera, las calificaciones se acumularán en el punto cero de la escala; algunas personas disminuirán la velocidad, pero se pasarán el alto; otras, no sólo no disminuirán la velocidad, sino que seguirán adelante e incluso acelerarán para pasar más rápidamente. Al vaciar estos datos en una distribución de frecuencias, aparecerá una curva, como lo indicamos anteriormente. Algo similar sucede en la medida de actitudes o rasgos de conducta, como la honradez, la sinceridad, etcétera.

CLASIFICACIÓN DE LAS PRUEBAS PSICOLÓGICAS

Es importante hacer una distinción de los términos que se emplean en el lenguaje común. Así, por ejemplo, cuando hablamos de instrumentos psicológicos, nos referimos a todo tipo de implemento que, dentro de las diferentes áreas y formas de investigación psicológica, utilizamos, los cuales se rigen por los fundamentos metodológicos de la medición científica y tienen características tanto formales como de contenido, que nos permiten emplearlos como instrumentos de medición.

Desde luego, la diferencia entre prueba y test no es más que distintiva del “para qué” se está utilizando el instrumento en un área específica, que es la medición de ciertos aspectos del individuo y que, ante la imposibilidad de emplear otro término, a pesar de la riqueza de nuestra lengua, nos

permite una comunicación exacta de lo que estamos haciendo y con qué fin, mediante ese tipo de instrumento. De tal suerte, por lo que respecta al término *test*, aún cuando sea un anglicismo, los psicólogos latinoamericanos han aceptado su uso, a fin de lograr una mejor comprensión.

Por otra parte, en un intento por establecer la distinción anteriormente explicada, se ha hecho común el uso del término *medida* para sustituirlo por *test* o hacerlo sinónimo de éste, cuando menos en español. Sin embargo, existe un desacuerdo en el sentido de que los términos *prueba* y *medida* son diferentes, ya que aquélla es una forma de medir aquel tipo de eventos psicológicos que en un momento dado son motivo de investigación y que, a menudo, se necesitan “probar” más de una vez, a fin de tener los datos que se requieren para hacer una afirmación. (La palabra *test* participa, en cierto modo, de este significado y es también una medida).

Por otro lado, los tests psicológicos pueden clasificarse desde diversos puntos de vista que les son característicos.

Así tenemos que, desde el punto de vista de su *objetivo*, los tests pueden ser: *de potencia o de ejecución máxima, o de rasgos o peculiaridades (ejecución típica o cualitativa)* llamados más comúnmente *pruebas proyectivas*.

Pruebas de potencia o de ejecución máxima

Este tipo de tests exigen del examinado su máximo rendimiento en la tarea o tareas que se le piden que ejecute, tanto en lo que se refiere a lo que “puede” hacer, como a lo que “rinde” en el momento del examen.

Bajo este rubro se incluye a las pruebas de inteligencia, de habilidades o aptitudes múltiples y específicas, de rendimiento (intelectual o de cualquier otro tipo que se esté investigando).

Al momento de administrar estos tests, los psicólogos deben presuponer que los sujetos están altamente motivados para ejecutar la tarea que se les ha asignado, a fin de dar su máximo rendimiento.

Existen, cuando menos, tres factores determinantes en la calificación de cualquier prueba de potencia: *a) habilidad innata*, es decir, la capacidad potencial para ejecutar cualquier tarea que se les presente a los sujetos; *b) habilidad adquirida*, es decir, el ejercicio y entrenamiento mediante el aprendizaje, que a través de su interacción con el ambiente va obteniendo el individuo, y *c) motivación*, o sea, el grado de disponibilidad en el que se encuentra el sujeto para ejecutar la tarea. Resulta difícil determinar qué porcentaje de la calificación de una persona se debe a cada uno de estos factores; más bien, debe pensarse que los tres influyen en la calificación que se obtiene; es decir, que la calificación de una persona puede deberse, en parte, a potencialidades innatas y a la forma en que éstas han sido modificadas por sus experiencias vitales (ambiente, aprendizaje, entrenamiento, etcétera), además de la motivación que la persona tenga al ser sometida al examen.

Como indicamos anteriormente, entre las pruebas de potencia o de ejecución máxima, se encuentran primero las llamadas “pruebas de inteligencia”. El concepto de inteligencia es demasiado amplio y abstracto, lo que ocasiona poca precisión y un desacuerdo general en cuanto a su significado. Así como cada quien puede tener su propia definición, cada autor de una prueba de inteligencia tiene su propia definición. No obstante, si observamos el criterio que siguen para llevar a cabo la medición y qué operaciones exigen en la tarea cada uno de ellos, podremos notar que existe

un acuerdo tácito entre ellos para que haya uniformidad, y lo que ocurre es que sólo consideran aspectos de aquello que comúnmente se conoce como “inteligencia”. Por ejemplo, algunos incluyen solamente reactivos verbales; otros, material visual o vasomotor (material no verbal); algunos enfatizan la solución de problemas, así como otros hacen hincapié en la memoria. Algunas pruebas de inteligencia producen una sola calificación (por ejemplo, C I), en tanto que otras arrojan varias y se evalúan independientemente.

Esta variedad de situaciones nos lleva a considerar que cuando una persona es examinada con distintas pruebas, seguramente producirá diferentes C I; por tanto, no podemos establecer similitud entre los resultados, a menos que busquemos la correlación que existe entre ellos. En ciertos casos, debería sorprender el hecho de que diferentes pruebas de inteligencia muestren resultados análogos.

El término inteligencia tiene muchos sinónimos que, si los analizamos, ratifican lo expresado anteriormente; es decir, miden aspectos diversos de ese comportamiento unitario al cual hemos dado en llamar “inteligencia”. Maduración mental, factor general de inteligencia, actitud escolar, habilidad general, habilidades mentales primarias, etcétera, son términos que tienen una misma significación acerca del concepto de inteligencia, pero que pueden diferir en cuanto al énfasis o aplicación. Para determinados propósitos, como predicción de logro escolar o programa de entrenamiento, las pruebas de inteligencia –pruebas de aptitud general o aptitud escolar – son formas distintas, de donde el P. de V. de su uso. Las pruebas de aptitud indican predicción, es decir, nivel futuro de ejecución de cierto tipo de tareas, y se emplean en la selección de personal, para admisión en programas de entrenamiento, para clasificaciones individuales, etcétera.

Existen diferentes tipos de pruebas de ejecución máxima, además de las de inteligencia. Las pruebas de rendimiento, empleadas para medir el nivel real de conocimiento, habilidades o capacidades, etcétera, constituyen una forma de lo que llamamos en teoría psicométrica una medida “transversal” y son de suma importancia para, por ejemplo, analizar el desarrollo psicológico o, como en el caso de las pruebas de rendimiento escolar, para medir lo que los estudiantes han podido aprender según el contenido de la materia que se está examinando.

La base principal de diferenciación entre las pruebas de aptitudes y las de rendimiento se refiere a sus aplicaciones. La misma prueba puede ser utilizada para medir cualquiera de las dos situaciones, es decir, ya sea para obtener el nivel de rendimiento presente o para predecir el futuro nivel de ejecución de los sujetos.

A menudo las pruebas de inteligencia son consideradas como pruebas de aptitudes; sin embargo, pocas veces, si no es que nunca, pueden utilizarse como indicadores de rendimiento académico. Una prueba de rendimiento mide, obviamente, la “inteligencia” considerada en su amplia extensión.

Las pruebas de ejecución máximas o de potencia no presentan dificultades para comprender lo que intentamos medir. En resumen, con las pruebas de aptitudes tratamos de predecir qué tan bien ejecutará el sujeto ciertas tareas específicas; con las pruebas de rendimiento medimos su logro actual; con las pruebas de inteligencia, aún cuando nos encontremos en desacuerdo en cuanto a las definiciones, estamos evaluando el nivel de funcionamiento o capacidad intelectual; más o menos consistente, sobre el cual se proyecta un individuo.

Pruebas de ejecución de rasgos (ejecución típica)

Como su nombre entre paréntesis lo indica, sería más conveniente llamarlas *pruebas de clasificación tipológica o cualitativa*, aún cuando se sabe que históricamente han sido denominadas *pruebas proyectivas*. Sin embargo, designarlas como *cualitativas o de rasgos* es más adecuado, ya que este tipo de pruebas representan una construcción de utilidad, aunque con frecuencia tienen problemas que, en algunos casos, parece que se han solucionado al aplicarles drásticamente el método científico en las formas que se han encontrado más adecuadas para hacerlo.

Entre este tipo de pruebas tenemos los siguientes tests: de ajuste, temperamento, intereses, valores, actitudes, preferencias, escalas de personalidad, inventarios de personalidad, índices, técnicas proyectivas, pruebas situacionales, etcétera.

La calificación que se lleva a cabo en pruebas de este tipo es definitivamente vaga y poco confiable –tal como se ha venido manejando hasta la fecha- por su carácter subjetivo en la mayoría de los casos.

Así, lo que para un autor significa “lo social” es diferente para otro. Entonces, la teoría subyacente y la explicación de una prueba de clasificación cualitativa es más compleja y menos obvia que en las pruebas de potencia. De este modo, mientras que las habilidades de una persona son más o menos estables, ciertos aspectos –sino es que todos – de carácter emocional, motivacional, etcétera, pueden cambiar en cortos periodos. Con estos instrumentos se intenta saber cómo reacciona, siente, opina o se comporta un sujeto frente a ciertos estímulos, a fin de poder entender si la conducta observada está determinada por los “supuestos elementos” que se hipotetizan en los tests y que son los productores de determinado comportamiento.

Es necesario aclarar aquí que las pruebas de actitudes, las cuales hasta hace algún tiempo habían sido incluidas en este tipo de tests, han salido de la problemática que plantean aquéllas, gracias a que ha sido posible manejarlas de otra manera, aplicando sobre todo las técnicas de construcción de escalas.

Como las actitudes son producto del proceso de socialización que es inherente al desarrollo del individuo, la influencia en su conducta es fundamental tanto para la forma de responder a su ambiente social particular, como hacia otros grupos y hacia sí mismo. No nos extenderemos aquí para referirnos a un tema que, por sí solo, es materia de un texto especial y muy amplio. Simplemente debe hacerse mención de que, desde el punto de vista de la clasificación de los tests, es menester colocarlas en su lugar apropiado. Así, este tipo de escalas depende de sus propias características, como cualquier otro tipo de prueba; por ejemplo: debe ser confiable y válida, además de considerar, en lo posible, la igualdad de unidades, la unidimensionalidad y el punto cero.

Resumiendo, bajo este rubro se incluyen todos los tests que se han elaborado para medir los tradicionalmente llamados rasgos de personalidad, exceptuando las escalas de actitudes. Este grupo de tests, aunque en la actualidad se ha podido adelantar considerablemente en su construcción y se puede hablar ya de la métrica de la personalidad –es decir, de tests más objetivos -, es materia de otro curso y de otros textos especiales.

La diferencia entre los tests de ejecución máxima y los de rasgos de personalidad radica en la objetividad de los datos, ya que ese tipo de pruebas está sujeto estrictamente a la metodología psicométrica. De tal suerte, podemos estar seguros de que, al pedir al sujeto su máxima ejecución en la tarea que se está explorando, obtenemos no sólo la más alta que él puede desarrollar, sino también la que expresará su capacidad real.

Uno no puede saber más álgebra o ser más hábil para la ejecución de algunas tareas que se nos exigen, que lo que “realmente” podemos realizar de acuerdo con nuestra capacidad o con nuestro aprendizaje previo. En cambio, en las pruebas de ejecución típica o de rasgos de personalidad el sujeto puede falsear las respuestas en uno u otro sentido, ya que en éstas no se le pide su máximo esfuerzo, sino que sea lo más espontáneo y honesto posible, según el criterio del examinador. Por otra parte, en este tipo de pruebas los criterios de “objetividad” son ambiguos, lo cual ocasiona que se pase por alto una de las principales reglas de la metodología de construcción de pruebas, que radica en la posibilidad de ser calificada por cualquier psicólogo que conozca los criterios de calificación.

Asimismo, estas pruebas presentan dos alternativas: o son calificadas por quien las administra y aplica su propio criterio o, cuando tienen que ser calificadas por diversos calificadores, cada uno de ellos da un juicio y un resultado diferente, lo cual anula la confiabilidad y validez de esos instrumentos de medición psicológica.

Por la *forma o tipo de respuesta que exigen*, las pruebas se clasifican en: a) pruebas objetivas, y b) pruebas subjetivas. Como ejemplo de las primeras tenemos a las pruebas de inteligencia; en el segundo caso están las pruebas de personalidad, especialmente los tests proyectivos como el TAT (pruebas de apercepción temática), el Rorschach, que es una prueba típicamente de “estructura de personalidad”, etcétera.

Por la *forma de administración de los tests*, éstos se clasifican en: a) tests de aplicación individual; b) tests de aplicación colectiva, y c) tests autoadministrados.

Los tests de aplicación individual, como su nombre lo indica, son administrados por el examinador al sujeto en condiciones “privadas”, o sea, en un lugar o cubículo especial, en donde sólo se encuentran ambos (examinador y examinado) en condiciones óptimas para ejecutar las tareas que, dependiendo de los objetivos del psicólogo, tendrá que realizar el sujeto.

Por otro parte, los tests de aplicación colectiva deben ser cuidadosamente manejados. Lo colectivo no significa que estos tests pueden ser administrados a una multitud. Al respecto, hay reglas, como en todo, para la cantidad de sujetos que pueden ser probados en un examen de este tipo. Así por ejemplo, si se examinan grupos muy numerosos disminuye la confiabilidad del test y, consecuentemente, la validez, pues no es fácil controlar a tantos sujetos para evitar que de alguna forma existan variables interferentes que distorsionen los resultados. Lo aconsejable, en este caso, es que los grupos no excedan de quince sujetos y que el número de “vigilantes” sea proporcional al de examinados, de tal manera que se eviten las interferencias hasta donde sea posible; es decir, debe haber tantos vigilantes como número de sujetos pueda atender cada uno. Si en algún caso resulta absolutamente necesario examinar grupos mayores, entonces habrá que aumentar el número de vigilantes en la forma antes explicada.

Por último, la forma autoadministrada de tests consiste en someter al sujeto a las mismas condiciones que en la forma de administración individual, excepto en que aquí el examinador solamente da las instrucciones acerca de lo que ha de ejecutar el sujeto, pudiendo ausentarse del cubículo y dejar a éste trabajar tranquilamente.

Sin embargo, en estos tres tipos de pruebas hay diferencias en cuanto a su manipulación. Así pues, muchas de ellas pueden ser administradas en las tres distintas formas, pero teniendo siempre en cuenta que cada forma debe tener sus propias normas de calificación, aunque se trate de la misma prueba, ya que habrá ciertos cambios en los resultados de una forma a otra de aplicación.

Sin embargo, no debe olvidarse que un gran número de tests sólo pueden ser aplicados individualmente por múltiples razones, de las cuales las más importantes son: a) que deba registrarse el tiempo de ejecución en forma estricta; b) esto puede ocurrir, entre otras cosas, porque se trate de una escala de inteligencia, por ejemplo, que por su estructura no permite la aplicación colectiva, por una parte, y por otra el tiempo no se puede registrar tan rigurosamente en grupo como individualmente; c) que se trate de una escala compuesta o mixta; es decir, que contenga tipos de tareas que exijan, de parte del examinado, la combinación de operaciones de tipo verbal y ejecutivo (motor), y d) que dependerá de los fines de la administración el que el propio psicólogo elija, a partir de las restricciones que la misma evaluación del instrumento, le plantee, la forma de administración.

En función de la *libertad de ejecución*, las pruebas pueden clasificarse en pruebas *de poder* y pruebas *de velocidad*.

Las primeras son aquellas en las que al sujeto examinado se le pide que emplee toda su capacidad en el examen, o sea, que realice lo más que pueda, sin escatimar esfuerzos en cuanto a sus habilidades y conocimientos. (En este tipo de tests o en algún test en especial se puede utilizar el registro del tiempo, pero ya estandarizado).

Por otra parte, las pruebas de velocidad exigen del examinado la ejecución más rápida que pueda realizar para terminar completamente la tarea que se le ha asignado.

Obviamente, estos dos tipos de tests tienen sus objetivos específicos y sus aplicaciones particulares mismos que dependen de las finalidades del examen. Por ejemplo, es frecuente emplear los tests de velocidad en la selección de personal, en lo educacional, etcétera, es decir, en aquellas situaciones en las cuales se debe probar *cómo* realiza una tarea una persona bajo tensión o presión. En los tests de poder solamente interesa conocer *qué tanto sabe* el sujeto o si puede realizar la tarea.

Por la *forma de dar las instrucciones*, las pruebas pueden ser orales o escritas, aunque a veces, en casos especiales, hay que darlas ejemplificando el examinador lo que quiere que el sujeto realice en cada tarea, como en el caso de las personas que carecen de la audición o de la palabra.

El material que se da para examen puede ser de varios tipos: a) de lápiz-papel; b) verbal completamente; c) de ejecución (material, manual, vasomotor); d) aparatos especiales para realizar determinada tarea, o e) una combinación de los tres primeros, sobre todo cuando se exploran en un mismo instrumento diversas funciones o habilidades. En este caso, generalmente se trata de escalas mixtas.

Esto nos obliga a referirnos ahora a las diversas formas de calificación, aún cuando sólo sea de modo general, ya que en el siguiente capítulo se expondrán con amplitud y detalle. Así pues, los tests psicológicos pueden ser calificados manualmente, según el tipo de unidad de calificación de que se trate; o bien, contando puntos, aplicando matrices de calificación o mediante computación (lo cual ahorra mucho tiempo, dinero y energía). Por supuesto, hay que saber qué tipo de tests lo permiten. Además, en algunos casos pueden combinarse las calificaciones manuales y por computadora.

Para cada caso, los “protocolos” u hojas de calificación serán especiales. Un protocolo es aquella forma en la cual el sujeto da sus respuestas. Muchos de estos protocolos contienen un lugar específico para anotar la calificación. En caso contrario, se debe tener una hoja especial para vaciar las calificaciones. A veces estas hojas están elaboradas especialmente para entrar de modo directo a la computadora.

Los tests psicológicos tienen *límites* de aplicación en lo que se refiere a la edad de los sujetos para quienes se han diseñado. Desde este punto de vista, los tests se clasifican en: a) tests para infantes o bebés (baby tests); b) tests para niños preescolares; c) tests para niños escolares; d) tests para adolescentes, y e) tests para adultos.

En muchos tests su longitud llega a contener dos o más de estas etapas de desarrollo, y algunos de ellos, tal como el de Binet-Simon, incluyen todas las edades, desde el nacimiento hasta la etapa adulta. En nuestra muy particular opinión, este tipo de instrumentos tienen menos confiabilidad y menos validez que los destinados específicamente a ciertos límites de edad, pues es obvio que el contenido para explorar cada función o factor deseado se reduce al máximo, para dar cabida a todo el rango de edades y, como recordaremos, la longitud de un test es un factor importante para recopilar mayor información acerca de lo que se intenta medir.

APLICACIÓN DE LOS TESTS A LAS DIVERSAS ÁREAS DE LA PSICOLOGÍA

En esta sección es importante referirnos, aunque sea brevemente, al cómo, por qué y para qué se utilizan los test psicológicos en las áreas, al menos las más comunes de la ciencia psicológica en México, ya que a menudo se cometen errores que conducen no sólo al desprestigio de quien emplea los instrumentos e ignora alguna información al respecto, sino de la profesión misma, ocasionando con ellos, críticas de los mismos psicólogos, que suponen –por ausencia del conocimiento del campo de la medición psicológica, en su mayor extensión – que los tests no representan ninguna utilidad. Al respecto, puede plantearse la siguiente pregunta, ¿acaso ya se ha encontrado alguna forma de obtener información acerca del comportamiento humano que sea mejor y más rápida, aunque no absoluta, porque en el campo de la ciencia lo único absoluto son las matemáticas?

En principio, actualmente suele hablarse de cuatro áreas de trabajo, desarrollo e investigación psicológicas: a) la psicología industrial; b) la psicología clínica; c) la psicología educacional, y d) la psicología social.

En cada una de estas áreas, los tests son diferentes tanto en su contenido como en su utilización. A veces algunos de ellos pueden servirnos en todas o en algunas de las áreas para obtener información.

Psicología Industrial

Desafortunadamente, el desarrollo, tanto de los métodos como de personal capacitado para el trabajo psicológico en el campo industrial, se encuentra censurado hasta cierto punto por aspectos de tipo social y ético profesional.

Desde luego, en forma general, la psicología industrial aplicada tiene en su contenido la selección y colocación de personal de trabajadores, el entrenamiento en la o las tareas que el trabajo específico implique y los problemas de relaciones humanas.

El mayor problema que se presenta en este campo es el de los factores humanos. Al respecto, el trabajador desempeña un importante papel en esta área, y su intervención ha sido, en los últimos años, motivo de un sinnúmero de investigaciones, que han conducido al reconocimiento de la complejidad que existe entre la relación del trabajo u oficio del empleado y muchos aspectos de su vida. Así pues, tanto su vida social como personal están supeditadas a su actividad específica en muchas formas, y viceversa. Su necesidad de trabajar, el gusto que encuentre en el desarrollo de sus tareas, su capacidad creativa para lograr desarrollar más allá de lo que le piden, la satisfacción por sus propios logros, en suma, sus sentimientos personales hacia sus objetivos sociales y personales forman parte del papel que desempeña en la industria.

La industria está organizada para incrementar el trabajo y llegar a la producción de sus objetivos finales. (Los factores humanos mencionados llevan a este incremento). Sin embargo, lo importante es reconocer que al ser aceptados los factores humanos como parte primordial de la industria, ésta tiene la obligación de comportarse recíprocamente con el trabajador. Así pues, está obligada a atender acuciosa y constantemente la adaptación de las necesidades y habilidades del trabajador a sus tareas.

A grandes rasgos, las implicaciones sociales concomitantes con el desarrollo de la industria son las siguientes: cambio en el nivel de vida del trabajador, modificación de la forma de organización del trabajo, y cambios en las condiciones de confort o comodidad social. Esto debe ser conocido, estudiado y entendido por aquellos psicólogos que se dediquen a este campo profesional.

Psicología científica

Este tipo de psicología tiene pocas soluciones específicas ya estructuradas para resolver la problemática humana en la industria, debido a la complejidad de la estructura conductual, que impide establecer generalizaciones, dada la variabilidad de los individuos en cuanto a su comportamiento y, sobre todo, ya situados en el ambiente del trabajo. Es ahí donde la capacidad y los conocimientos del psicólogo pueden y deben crear los implementos necesarios apropiados a la situación y establecer reglas que sean útiles para la solución de problemas conductuales en el trabajo. Además, en el lado opuesto, encontramos la situación de la fábrica, oficina o lugar donde el trabajador ha de desempeñar sus labores. Estas siempre son diferentes de las que pueden crearse artificialmente en un laboratorio para investigar el área industrial. (Creemos que, efectivamente, el

laboratorio, no sólo en la psicología industrial, sino también en la psicología general, no representa completamente las situaciones reales). Los laboratorios son de gran utilidad, siempre y cuando el psicólogo entienda que los hallazgos que ahí tienen lugar no pueden ser aplicados rígidamente, sino que debe haber una flexibilidad, la cual será dada por los conocimientos, la práctica, la experiencia, y sobre todo, la comprensión de lo que es su tarea.

Como este libro solamente se refiere a la aplicación de los tests al campo de la industria, nos concretamos a exponer esto que consideramos necesario para que el estudiante advierta las limitaciones que existen con el instrumental psicológico en la industria.

Así pues, en la ejecución de un trabajo existen factores que son determinantes para que aquél tenga una realización eficiente. Estos factores varían en cada situación particular y en cada individuo, lo cual obliga a los psicólogos industriales a una constante revisión de sus instrumentos, es decir, a una constante revisión de la confiabilidad y validez de estos últimos.

Evaluación de los instrumentos en el campo de la Psicología industrial

Cuando no se tiene una visión clara de los requisitos necesarios para la ejecución de las tareas que un determinado trabajo exige, no se está en condiciones de hacer una elección adecuada del personal que habrá de desempeñarlo.

Uno de los primeros pasos para realizar la selección y clasificación del personal de trabajo es el *análisis del trabajo*, especializado o no, y cualquiera que sea el nivel o rango de que se trate. Esto incluye tanto la descripción como la especificación del trabajo u ocupación, las cuales tienen relación con las condiciones físicas y administrativas, así como con los métodos y procedimientos del trabajo.

Para poder predecir la efectividad y el éxito en una tarea, deben tenerse en cuenta todos los elementos que implica el trabajo, es decir, debe evaluarse la tarea. Así pues, hay que considerar los métodos y procedimientos de trabajo, el instrumental y equipo y los programas de entrenamiento. Sin embargo, como para hacer un análisis efectivo es necesario observar tanto al trabajo como al trabajador, habrá que apreciar la forma natural en que este último se desempeña, a fin de determinar las actividades u operaciones que realiza para obtener su objetivo de trabajo.

Una de las primeras observaciones a establecer es la capacidad de aprender que tenga el trabajador, así como su habilidad para ejecutar las operaciones necesarias; descripciones que deben darse en términos cuantitativos, a fin de obtener la norma de ejecución individual a partir de la comparación con un cierto número de individuos.

La medición del aprendizaje y de las habilidades de los sujetos sirve para los siguientes propósitos: a) para pronosticar el éxito o fracaso de los trabajadores de nuevo ingreso; b) para tener datos objetivos con miras a la promoción, cambio o retiro del trabajador; c) como medio para evaluar la propia organización por secciones y en su totalidad; d) para evaluar la capacidad, calidad y productividad de los equipos, instrumentos y condiciones de trabajo; e) para observar la extensión con que los instrumentos de medición psicológica estén siendo útiles a los propósitos del psicólogo seleccionador y clasificador de personal; f) para evaluar los programas de entrenamiento que se estén

realizando, y g) como base para la elaboración de nuevos instrumentos, programas de entrenamiento, etcétera. Otro aspecto muy importante es el poder advertir, mediante los instrumentos adecuados, la situación de las relaciones interpersonales a diversos niveles y las posibles razones que estén determinando ciertas situaciones (negativas o positivas).

Los instrumentos psicológicos de mayor utilidad en este campo, guardando siempre las limitaciones indicadas anteriormente, son:

- a) La entrevista psicológica, la cual sirve para obtener una serie de datos personales, indispensables para el trabajo. Dicha entrevista debe ser preparada de antemano, en función del análisis previo de las características que se requieren para el desempeño del trabajo.
- b) Grupos de tests que midan la capacidad de aprendizaje del examinado, ya que este aspecto es fundamental en todo tipo de actividad conductual humana.
- c) Escalas, tests, etcétera, de habilidades múltiples y especiales.
- d) Tests de rendimiento intelectual. Aún cuando este tipo de instrumento no es fundamental, sí se requiere un mínimo de información acerca de ciertos factores que miden estos instrumentos, como la comprensión, atención, capacidades de anticipación, planeación de las tareas, etcétera, con el fin de poder ratificar los instrumentos diseñados para probar la capacidad de aprendizaje.

Consideramos que estos son los instrumentos más importantes para la selección y clasificación de personal, ya que tienen la ventaja de dar resultados que caracterizan a los tests que se emplean en la industria: la predicción de la eficiencia, tanto del trabajador como de la organización.

En nuestra opinión particular, realmente debe hacerse una clasificación de personal, más que una “selección” del mismo. Esto nos permitirá ser más justos y en la designación de las tareas que deberán realizar los aspirantes. La selección es realmente “eliminación”. La “clasificación” conduce a la “colocación” de las personas en aquellas ocupaciones o puestos en donde puedan rendir mejor, permitiendo así la posibilidad de elaborar programas de entrenamiento que ayuden a las personas, en vez de eliminarlas, impidiéndoles realizar las tareas para las que están capacitadas.

Utilización de los tests en psicología clínica

Uno de los problemas que más se han estudiado es el de la utilidad de los tests psicológicos en el diagnóstico de todos aquellos aspectos de la conducta que se ha dado en denominar *personalidad*. Haremos aquí un paréntesis para explicar la frase anterior: consideramos que no es posible, en el funcionamiento real de un individuo, separar su actividad intelectual, sus actitudes, intereses, valores, creencias y características peculiares de comportamiento (rasgos), sino solamente para su estudio. El hombre funciona como una unidad; cuando actúa expresa no sólo la forma en que ha recibido la estimulación del mundo exterior y la ha codificado, sino también cómo reacciona ante él, tanto en la forma en que lo ha percibido, como en la manera en que lo ha experimentado.

Desde luego, en el campo de la psicología clínica se ha desarrollado, desde un principio, una gran cantidad de tests para medir los rasgos de la personalidad. Aquí ocurre un fenómeno común siempre que se trata de medir algo: según el aspecto que se mida de ese algo, que es una

unidad, el constructor del instrumento de medición dará su propia definición de acuerdo a su punto de vista, muchas veces teniendo en cuenta que está midiendo el o los elementos sobresalientes del fenómeno. Así ha ocurrido en las definiciones de conceptos tan complejos, como son los de personalidad e inteligencia.

Por cuanto toca a la personalidad, encontramos que existen diversas definiciones y que cada uno corresponde, en la mayoría de los casos, a las diferentes aproximaciones teóricas que llevan consigo un instrumento de medición para probar experimentalmente que los datos corresponden a la realidad. Sin embargo, todas las definiciones dadas incluyen todos los rasgos de la habilidad humana que, para su estudio, se separan.

El estudio de la personalidad tiene una amplia relación con lo que las personas son o parecen ser en un momento dado y en la forma en que adquirieron esa manera de comportamiento. Desde luego, la forma tradicional de medición en psicología clínica, en relación con la personalidad, tiene que ver con la primera parte de la afirmación anterior. No obstante, para explicarnos cómo ha desarrollado el individuo un conjunto de rasgos, tendríamos que recurrir a mecanismos genéticos, los cuales hasta ahora poco pueden aportar a la teoría del aprendizaje, sobre todo al aprendizaje social, que influye definitivamente en la forma específica de ser (rasgos) de cada persona.

La psicología clínica no sólo tiene que ver con los problemas estrictamente de enfermedad mental, sino con todo aquello que se considere alteraciones conductuales.

Por otra parte, para llevar a cabo estudios clínicos no es indispensable que el instrumento sea estrictamente de construcción para medir problemas psicológico-clínicos, sino que cualquier instrumento de medición –en este caso, los tests psicológicos en general- puede darnos datos importantes acerca de comportamientos considerados como fuera de la norma.

Todas las pruebas objetivas que tienen un uso determinado, por ejemplo, tests de inteligencia, escalas de inteligencia, tests de aptitudes especiales y múltiples, etcétera, fundamentalmente nos dan información acerca del comportamiento general del sujeto, de sus eficiencias y deficiencias en su manifestación conductual. De ahí que algunas escalas mixtas de medición –por ejemplo, las de la inteligencia de Wechsler – puedan utilizarse en el diagnóstico clínico, es decir, en la detección de problemas que, independientemente de la medida de la inteligencia, se presentan en el comportamiento frente a las tareas que cada uno de los subtests le imponen al sujeto examinado.

Evaluación de los tests en psicología clínica

Uno de los problemas más comunes en el campo de la psicología clínica es la generalización que hacen algunos psicólogos respecto al uso de los tests. Durante mucho tiempo, al menos en México, los psicólogos utilizaron los tests de medición de la personalidad para cualquier tipo de estudio psicológico, aunque no fuera estrictamente clínico. Afortunadamente, parece ser que ya se empieza a efectuar una discriminación más clara acerca del empleo que debe darse a cada uno de los diversos tipos de tests existentes.

Al igual que en todos los campos y especialmente en éste, en lo referente a tener en consideración las exigencias psicométricas inherentes a todo test para poder ser utilizado con

confianza y con la debida ética profesional, la evaluación dependerá de los objetivos que el psicólogo clínico pretenda lograr. Estos objetivos pueden ser los siguientes:

- a) Diagnosticar el deterioro mental, a partir de instrumentos como los tests de formación de conceptos, de desarrollo de la percepción, de alteraciones de la atención, de la comprensión, de la memoria, etcétera.
- b) Diagnosticar otros aspectos del comportamiento humano, como las denominadas conductas anormales y patológicas o, simplemente, ciertas formas especiales de comportamiento que no llegan a tener las dimensiones de las anteriores.

Ahora bien, el diagnóstico, en sí, tiene también un objetivo: pronosticar, con el fin de intentar prevenir o reforzar, o intentar, suprimir la conducta que se está detectando, siempre con miras a una meta que, en principio, puede ser la salud mental.

Existe cierto tipo de pruebas en el campo de la psicología clínica que definitivamente no deberían utilizarse, porque carecen de los principios fundamentales de la medición psicológica. Así pues, hay tests que intentan medir la “estructura de la personalidad”, la “proyección de la personalidad”, los “rasgos de la personalidad”, etcétera. Algunos de estos tests están constituidos por material no estructurado, láminas con escenas o dibujos, frases o temas para completar no estructurado, láminas con escenas o dibujos, frases o temas para completar o inventarios de personalidad. Todos ellos serían valiosos si se ajustaran a la metodología psicométrica.

Afortunadamente, con los inventarios de personalidad se ha iniciado, y ahora es más patente, el desarrollo de lo que se llama *métrica de la personalidad*. Si bien es cierto que ésta aún tiene grandes limitaciones, también resulta cierto que es más completa que la forma antigua de manejo de estas técnicas de mediación. Sin embargo, la investigación en este sentido se está incrementando considerablemente, a fin de llegar a obtener datos cada vez más precisos acerca de los problemas que se le plantean a la psicología clínica.

Aplicación de los tests a la psicología educativa

La educación es uno de los campos de la psicología que revisten mayor importancia. Consecuentemente, si al paso de la historia el hombre ha desarrollado sistemas de medición para todos aquellos fenómenos que ha podido observar en su ocurrencia y los ha refinado y hecho más confiables, también tenía que observar el fenómeno educativo aplicándole algún sistema apropiado de medición.

En un principio, los maestros tenían la necesidad y la responsabilidad de juzgar y observar los efectos de la enseñanza. Subjetivamente podían decidir quién aprendía más rápida o más lentamente; quién tenía mayor capacidad de retención y quién olvidaba rápidamente; en fin, esta tarea, bastante difícil, no sólo tenía que referirse a los alumnos, sino también a los maestros mismos.

Sin embargo, recientemente se han desarrollado instrumentos de medición cada vez más refinados que, influidos por métodos utilizados en otros campos de la ciencia, han venido a tomar lugar en la educación. Estos métodos han tenido un gran impacto en el aprendizaje y en la enseñanza, así como en todo aquello que tenga relación con el ambiente escolar.

Uno de los usos más comunes de los tests psicológicos en la educación es el que le dan las escuelas, considerándolos como elementos de gran valor para la enseñanza individual. Esto se debe a que, si partimos de la base de que existen diferencias individuales, muchos aspectos de la educación deben ajustarse a tales diferencias. El maestro no solamente utiliza los tests para advertir el grado de aprendizaje que ha obtenido el alumno en las materias que forman el currículo del nivel de que se trate, sino también, y como consecuencia de la variabilidad humana, se ha visto en la necesidad de intentar conocer cada vez mejor las características del rendimiento escolar del alumno en el presente, pretendiendo establecer hasta dónde puede llegar en lo futuro.

Independientemente de que hoy día sabemos que existen técnicas de aprendizaje capaces de desarrollar las habilidades humanas con una mayor extensión que como lo hacían las técnicas antiguas, los tests psicológicos son necesarios para medir esos cambios que se operan mediante el entrenamiento en la ejecución de las tareas.

Así pues, sabemos que al ingresar el niño a la escuela por vez primera, trae consigo un cúmulo de experiencias personales, es decir, su historia personal hasta ese momento. Esto determina una forma particular de comportamiento en todos sus diversos niveles de funcionamiento, que, al enfrentarse con las exigencias del medio escolar, se expresan en forma integral, produciendo conductas específicas. De este modo, observamos niños que pueden aprender a leer antes de lo que la norma ha establecido; a otros se les dificulta esta actividad y lo hacen más tarde; otros más se desempeñan mejor en ciertas áreas del conocimiento que se trasmite a través de las diversas técnicas de aprendizaje existentes; por ejemplo, unos lo hacen mejor en las matemáticas que en otras áreas; a otros les resulta más fácil aprender geografía o historia; algunos aprenden utilizando sus habilidades en formas diferentes a las de otros (es decir, tienen estilos particulares de manejar sus habilidades); otros más aprenden con más lentitud; otros con mayor rapidez, etcétera.

Entonces, es necesario que el maestro discrimine lo mejor posible aquellas situaciones que en múltiples casos pueden dar lugar a serias alteraciones, las cuales en un momento dado impidan que la eficiencia de la enseñanza y del aprendizaje sea similar para todos los miembros del grupo escolar.

Utilizando los tests psicológicos, el maestro puede ir recopilando la historia escolar de cada uno de los alumnos bajo su responsabilidad. Desde luego, esto ocurre en aquellos países cuyo desarrollo educacional se ha visto favorecido. En México parece que ya empieza a haber interés por llevar a cabo ese tipo de registros en las escuelas, asignando psicólogos para que auxilien a los maestros en el desempeño de esas tareas.

Se supone que el rendimiento académico será más efectivo si la enseñanza que se imparte recae en sujetos que participen de una capacidad de aprendizaje análoga o, al menos, que no exista una dispersión significativa para alterar el funcionamiento normal del grupo.

Como podemos ver, los tests, al aplicarlos a la educación, no sólo sirven para que el maestro evalúe el rendimiento escolar, sino que son necesarios, utilizados por especialistas, tanto para determinar el tipo de población escolar con que se cuenta, como para desechar, modificar o crear técnicas de enseñanza y aprendizaje más operantes y cada vez más refinadas.

A menudo se oye hablar de programas de medición escolar a gran escala e incluso a nivel nacional. Al respecto podemos decir que, si bien es cierto que los maestros no están familiarizados con los tests y las técnicas de medición, también resulta cierto que la mayoría de ellos no desconocen los problemas educativos y comprenden perfectamente hasta dónde la medición psicológica puede resultar útil.

Los psicólogos educativos, los consejeros escolares y los directores de investigación psicológica especializados son los abocados para elaborar los programas de medición, los cuales obedecen a los siguientes propósitos:

- a) Clasificación del alumno (que en ocasiones sirve para promoverlo de nivel escolar);
- b) Homogeneidad en los grupos;
- c) Orientación y consejo escolar;
- d) Búsqueda de motivaciones para producir un mejor rendimiento en los alumnos;
- e) Prevención de posibles problemas de los educandos;
- f) Detección y estudio de niños “excepcionales”;
- g) Mejoramiento del profesorado;
- h) Educación escolar;
- i) Estructuración de las escuelas en función de la comunidad; etcétera.

Esta utilización de los tests en el campo educacional debe ser consistente y sistemática, a fin de poder recopilar poco a poco la información de los adelantos y posibilidades de los alumnos.

En el área de la psicología educativa, los tests deben seleccionarse de acuerdo con aquellas variables que son indispensables para medir el funcionamiento escolar del individuo.

Utilización de los tests en la psicología social

Una de las áreas más importantes, sobre todo en los últimos años, al igual que la educativa, es la de la psicología social. El individuo, como elemento perteneciente a una sociedad específica, participa de sus valores culturales, de la aceptación de la conducta de los otros y de las relaciones interpersonales satisfactorias.

El ambiente social sufre el fuerte impacto y la penetración de la cultura. Para poder entender el comportamiento humano, es necesario comprender las formas aprendidas y organizadas de su conducta particular que, en última instancia, constituyen la cultura. Puede decirse que un *patrón cultural* es un conjunto de formas de conducta compartidas por un grupo y que los hace diferentes unos de otros.

Para entender los factores que influyen en el desarrollo de la sociedad, debemos enfocar nuestra atención hacia la cultura, la cual, en última instancia, será la única que nos indique la generalidad o particularidad de la conducta humana. Cada grupo social tiene estilos peculiares para lograr sus objetivos, ya sean de tipo educativo, político, social, etcétera. Por tanto, el proceso de socialización depende del contexto cultural en que se dé.

En esta situación, una de las formas para llegar a la comprensión tantas veces mencionadas de los productos de la conducta humana social, es la utilización de una metodología específica.

Dentro de esa metodología quedan considerados los tests psicológicos aplicados al campo de lo social. Los tests más comúnmente utilizados son los de opinión, de valores, de intereses y, fundamentalmente, los de actitudes, ya que éstas constituyen un sistema casi permanente de reacciones afectivo-evaluativas que expresan el concepto de las características de los objetos sociales que son aprendidos en el proceso de socialización del individuo. Esas reacciones afectivo-evaluativas, que se traducen en actitudes hacia los objetos del ambiente, tienen su origen, fundamentalmente, en estructuras cognoscitivas, transmitidas algunas y experimentadas otras, por los componentes de un grupo social determinado.

Una de las formas más importantes de medición y que en la actualidad se está convirtiendo en un instrumento fundamental en el campo psicosocial son las escalas de actitudes, las cuales requieren un tratado o texto especial, tanto por sus características como por su metodología de construcción y eficiencia en los datos que producen, siempre y cuando se conozcan perfectamente y sean utilizados de manera pertinente.

Obviamente, este tipo de tests no es el único instrumento con que se cuenta en el campo de la psicología social, tanto aplicada como en la investigación. Ya hemos manifestado que existen otros elementos de trabajo, entre ellos los modelos matemáticos aplicados a esta área, mismos que, por el momento, salen de nuestro interés.

2

UNIDADES DE MEDICIÓN PSICOLÓGICA Y NORMAS DE CALIFICACIÓN

UNIDADES DE MEDICIÓN PSICOLÓGICA

En el capítulo anterior nos referimos brevemente a la forma como se lleva a cabo la medición en el campo de la psicología, así como a sus limitaciones, es decir, cómo asignamos números a los atributos de los “objetos psicológicos”: asignamos números para representar cantidades de uno o varios atributos de los objetos, estableciendo cuidadosamente las reglas o procedimientos para ese fin. Hay que recordar que estos procedimientos deben ser estandarizados, porque no sólo permiten la *objetividad* de los resultados, sino también establecer el tipo promedio de comportamiento –o atributo que se mide -, llamado también “norma”, para poder dar sentido al proceso de medición y al empleo de métodos eficaces de análisis matemático.

Debe tenerse presente que, mientras que la medición está relacionada directamente con el mundo real, con el de los objetos, los sistemas matemáticos son puramente abstractos; en consecuencia, sus reglas sirven para manejar símbolos, que pueden ser letras, las cuales generalmente no tienen relación con los números. De este modo, la medición, en tanto que su fin es la cuantificación, siempre maneja números. Su interrogante principal es: ¿qué tanto de un atributo existe en el o los objetos?; es decir, establece siempre como principio el concepto de “cantidad”.

La ciencia ha desarrollado reglas para cuantificar los atributos de los objetos reales, recurriendo a los sistemas matemáticos con el fin de poder establecer las relaciones internas entre los datos obtenidos con diferentes tipos de mediciones.

Mucho se ha inquirido acerca del por qué del empleo de las matemáticas en el campo de la psicología. Indiscutiblemente, no sólo la psicología, sino también la economía, la sociología y la biología lo han hecho. Es indudable que la única área de las matemáticas utilizada por estas ciencias del comportamiento ha sido la de la *teoría de la probabilidad*, porque proporciona los fundamentos teóricos de la estadística inferencial, desde los problemas de delimitación de los fenómenos y muestreo, hasta los diseños experimentales; además, constituye un lenguaje con sintaxis propia, a través de la cual pueden explicarse múltiples fenómenos conductuales.

Así, la teoría de la probabilidad establece, por una parte, la asignación de números a los eventos o atributos de los objetos bajo tres condiciones principales (es decir, tres reglas básicas): a) que los números *siempre* deben ser positivos –*condición de no negatividad*; b) *condición de normalización*, que establece que la probabilidad de que un evento ocurra es una, y c) *condición de aditividad*, que indica que si dos eventos son mutuamente exclusivos, la ocurrencia de uno elimina la del otro, de tal suerte que la ocurrencia de uno u otro de ellos es la suma de las probabilidades

individuales de ambos eventos. De estas condiciones se derivan todas las leyes de la teoría de la probabilidad.

Por otra parte, y ha sido problema de frecuente discusión, la teoría de la probabilidad, considerada en sí misma como una construcción matemática, no establece los procedimientos para la construcción del conjunto de eventos o para la asignación de números de ellos.

Ahora bien, como las reglas para calcular la probabilidad de ocurrencia de eventos complejos son idénticas a las utilizadas para el cálculo de superficies y volúmenes en geometría, los psicólogos han sustituido el término “conjunto” por *eventos o atributos* y el de “superficie” o “volumen” por *probabilidad*. El problema se reduce, entonces, a asignar superficies apropiadas para los conjuntos, lo cual pertenece al campo de la teoría de la medición.

Recientemente, los científicos psicólogos han venido desarrollando, igual que desde hace algunos siglos otras ciencias, tipos factibles de escalas de medición que puedan constituir la esencia de los diversos aspectos de la medición en psicología. De este modo, uno de los problemas más frecuentes entre los psicólogos que manejan instrumentos de medición es la interpretación que deben dar a los números que obtienen de las mediciones psicológicas. Esto parece derivarse de las conclusiones relativas a la legitimidad del uso específico de la clase de procedimientos matemáticos para la medición de atributos psicológicos. Así pues, resulta lamentable que muchos psicólogos no entiendan aún que en el desarrollo de la medición psicológica destaca fundamentalmente el *muestreo de contenido*, y no el muestreo de personas.

Sabemos que la física es la ciencia que mide con mayor exactitud su universo de población. Esta disciplina científica ha desarrollado un sistema de asignación de números a los fenómenos que investiga: las escalas de medición. De aquí, los psicólogos han tomado aquellas que les son útiles, en función de los objetos que mide la psicología. Al respecto, existe una clasificación convencional de esas escalas, conocida por todos nosotros.

El propósito fundamental de una escala es generar un continuo, en el cual queden o puedan quedar ubicadas las personas u objetos y, a partir del atributo que se intenta medir, ser localizados, ya sea para clasificar, ordenar o conocer la magnitud relativa de dicho atributo, a fin de establecer semejanzas o diferencias en la posesión que cada objeto o persona tiene de él.

En psicología sólo podemos utilizar tres de las escalas físicas: las nominales, las ordinales y las intervalares. Todas ellas nos conducen a la obtención de una serie de resultados expresados en números. A estos números se les denomina *calificaciones crudas o directas*.

Las calificaciones crudas o directas son el número real de respuestas dadas o de puntos obtenidos por un individuo o por un grupo. Estas calificaciones, en sí mismas, no tienen significado; así un individuo puede obtener 100 puntos, otro 200 y un tercero 300. ¿Esto nos dice algo? Sí, solamente que son puntos obtenidos por determinado sujeto; pero, ¿qué significan?, ¿qué significa que Juan obtenga 20 puntos en un test y 200 en otro?

Las calificaciones obtenidas por un sujeto o por un grupo sólo tienen significado en función de un promedio de calificaciones, es decir, de un estándar o parámetro de comparación.

En los ejemplos anteriores, los números obtenidos sólo adquieren significado si se obtiene la media de las calificaciones, o sea, el promedio. En el caso de Juan tendrá que procederse de manera semejante en cada test, a fin de poder traducir cada puntuación de 20 y darle su significado apropiado. Es probable que en cada uno de los tests que se apliquen a Juan, el número 20 signifique algo completamente opuesto, o al menos diferente. Lo mismo sucederá con el grado de variación de las calificaciones en un mismo sujeto y en diferentes momentos o en un mismo grupo de sujetos, tanto entre ellos mismos como en momentos distintos. Los cambios o diferencias que se presenten en las situaciones anteriores se llaman *variabilidad o dispersión*, y tendrán significado, también, en la desviación que muestren respecto a la medida o promedio, ya sea hacia arriba o hacia abajo.

Si se aplican tres tests diferentes a un mismo grupo, las calificaciones medias de cada test tienen lo que se llama “rango relativo”; es decir, los sujetos que obtienen cierta calificación en cada test, pueden caer hacia el promedio; otros, en la desviación inferior primera o segunda; otros más, en la desviación superior, etcétera. Entonces se dice que las calificaciones tienen una “significancia relativa”.

Todo esto permite establecer comparaciones del comportamiento obtenido por cada sujeto en cada test.

UNIDADES DE CALIFICACIÓN

Las unidades de calificación son las calificaciones derivadas, las cuales tienen dos objetivos fundamentales: a) convertir las calificaciones de diferentes pruebas en datos comparables expresándolos en un mismo tipo de escala, y b) hacer posible la mejor interpretación de las calificaciones.

Las calificaciones derivadas tienen sus ventajas y sus limitaciones. Cuando aplicamos una prueba, de hecho estamos tratando de medir tres elementos que tienen un papel importante en el desempeño de las tareas que se le piden al examinado, que son: sus habilidades, su motivación y sus conocimientos. Estos elementos pueden ser de tres formas para expresar las calificaciones de las pruebas: a) comparación con un estándar absoluto, b) comparación interindividual, y c) comparación intraindividual.

Existe una tendencia a utilizar solamente las calificaciones medias de los sujetos que se estudian y, como se afirmó anteriormente, esto no es importante en investigaciones de diferencias individuales y en la mayoría de los experimentos psicológicos. Es interesante hacer notar que la teoría del error de la medición haya sido desarrollada por los psicólogos especialmente y esto es debido a que los instrumentos psicológicos, por su propia naturaleza, asimilan una gran cantidad de errores de medición.

Entre muchos de los intentos que se han hecho para lograr una mayor exactitud y una mejor comprensión del significado de las mediciones en psicología, tenemos el desarrollo de la “instrumentación para la clasificación de los sujetos”, en una variedad de sistemas que en la teoría psicométrica se conocen con el nombre de *unidades de medición*. Las diversas unidades de medición existentes en psicometría se han elaborado para cubrir los casos específicos de medición de que se trate; es decir, qué se va a medir, a quiénes se va a medir, para qué y cuándo se va a efectuar la medición, qué fin persigue, y cómo se van a utilizar tales datos.

De esta manera, podemos hablar de tres amplios sistemas de unidades de calificación bajo cuyas normas se encuentran tipos específicos de unidades de calificación, dependiendo del manejo que se haga de ellos. Estas tres grandes áreas son, como se indicó anteriormente:

- I. Unidades de calificación de comparación con un estándar absoluto;
- II. Unidades de calificación de comparación interindividual, y
- III. Unidades de calificación de comparación intraindividual.

A continuación haremos una clasificación de estos sistemas.

I. COMPARACIÓN CON UN ESTÁNDAR ABSOLUTO O DIFICULTAD DE CONTENIDO

Porcentaje de calificaciones correctas.
Grados por letras.

II. COMPARACIÓN INTERINDIVIDUAL

A. *Considerando la media y la desviación estándar del grupo:*

1. Calificación Z
2. Calificación T
3. Calificación AGCT
4. Calificación CEEB
5. CI desviación: a) CI Wechsler, b) CI Stanford-Binet

B. *Considerando los rangos de los grupos :*

1. Rangos
2. Rangos percentiles y cadenas percentilares
3. Grados por letras.
4. Calificaciones estándar normalizadas (áreas de transformación): a) calificaciones T escaladas, b) calificaciones estandinas, c) calificaciones C escaladas, d) calificaciones estén, e) CI de desviación (subtest de Wechsler), y f) calificaciones ITED.
5. Rangos deciles.

C. *Considerando el rango de calificaciones en un grupo:*

Colocación en porcentajes.

D. Considerando la posición de aquellos que obtuvieron la misma calificación:

1. Calificación por edad: a) edad mental, y b) edad educacional.

III. COMPARACIÓN INTRAINDIVIDUAL (DE DOS MEDIDAS DEL INDIVIDUO)

- A. CI de razón.
- B. Índice de nivel intelectual.
- C. Cociente educacional.

Como es sabido (figura 1), en una distribución normal las calificaciones del tipo 2.1 y 2.2 están interrelacionadas. Podemos hacer transformaciones de un tipo de calificaciones a otro, si partimos de la hipótesis de que la distribución normal de calificaciones está basada en el mismo grupo de sujetos. Esto implica dos conceptos fundamentales: “normalidad” y “el mismo grupo”. Si se trata de grupos diferentes no podemos hacer comparaciones; si suponemos que las calificaciones no se distribuyen normalmente, podemos encontrar que algunas relaciones cambian, en tanto que otras permanece igual.

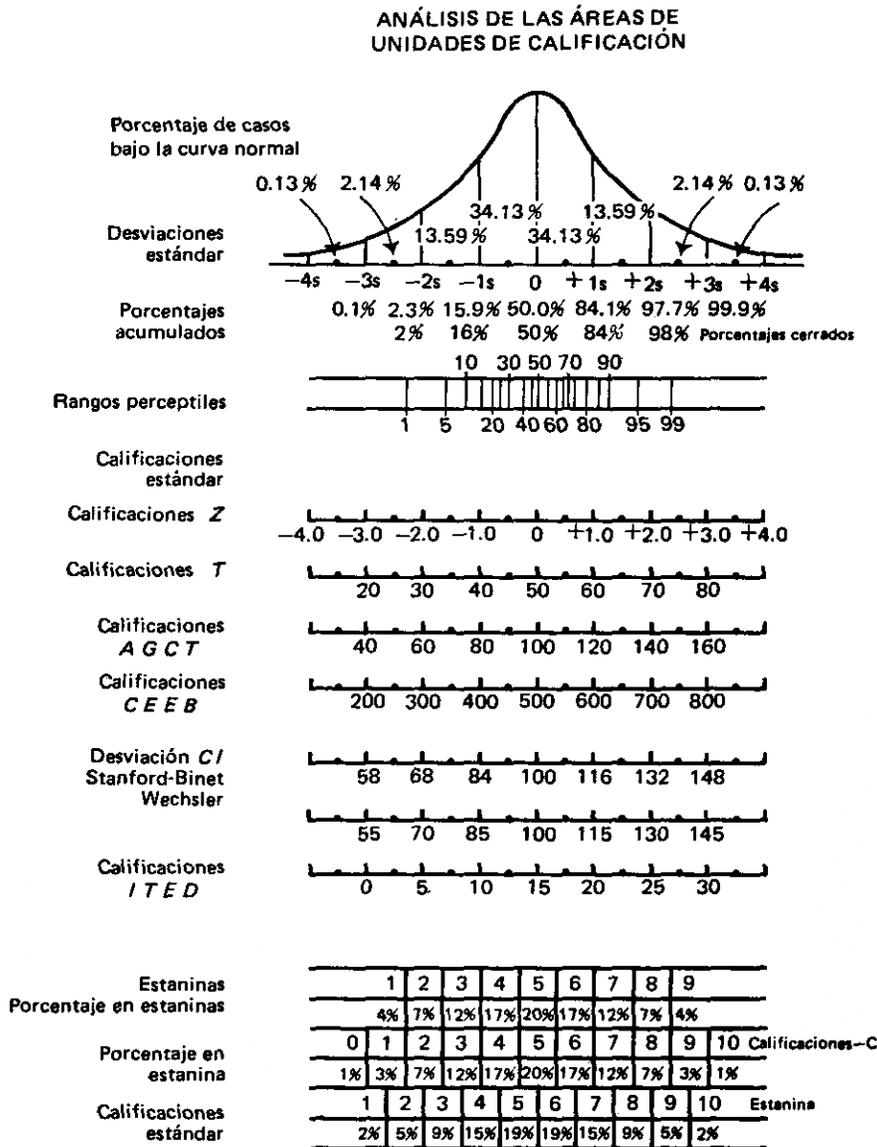


FIGURA 1. Curva normal de distribución y sus relaciones con diversas unidades de calificación psicométrica (tomada de Howard B. Lyman, *Tests Scores and What They Mean*).

En la figura 1 se muestra una serie de líneas de base trazadas a partir de la abscisa original, que nos permiten transformar un tipo de calificación a otra mediante el simple movimiento vertical, o sea, la superposición de las demás líneas a la abscisa original.

En la figura 2 podemos ver el mismo tipo de calificaciones en una distribución que tiende hacia el lado izquierdo. Puede apreciarse fácilmente el cambio de relación entre dos tipos de calificaciones. Para interpretar estas calificaciones se requiere un elemental conocimiento de la media, la mediana, la desviación estándar, el rango y la curva normal de probabilidad.

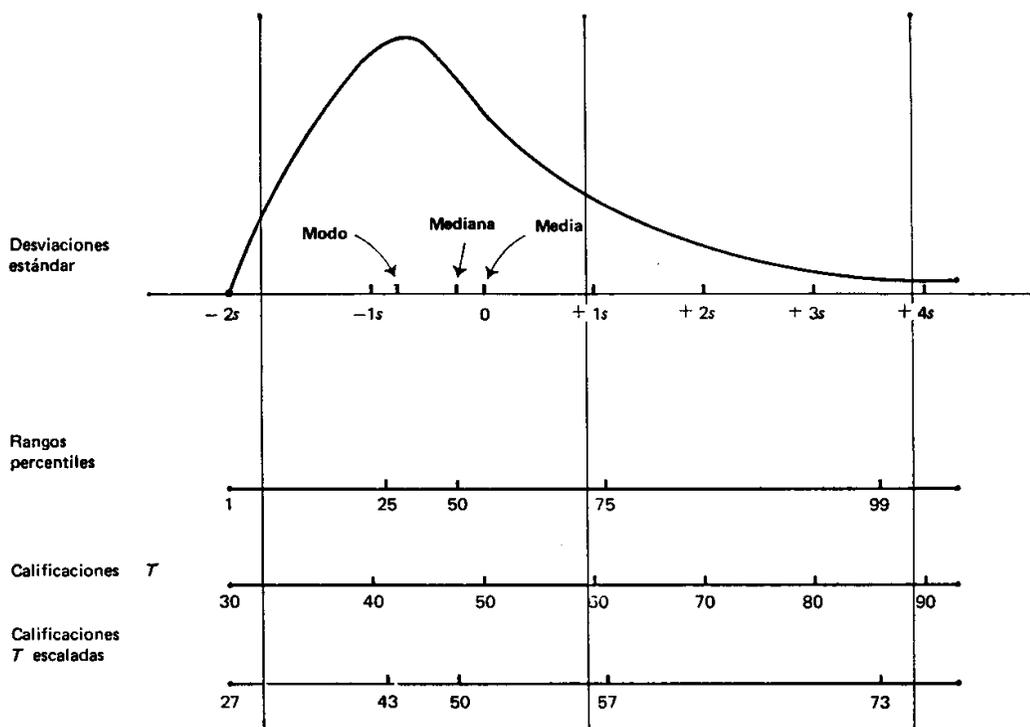


FIGURA 2. Relación entre algunas unidades de calificación psicométrica y una distribución positivamente coleada (tomada de H. B. Lyman).

No consideramos necesario explicar, ni siquiera brevemente, algunas de los tipos y unidades de calificación derivadas, ya que no son de uso común en las pruebas psicológicas que se emplean en México, aún cuando ya hayan sido formuladas para usos específicos en otros países. Sin embargo, esto no quiere decir que si nosotros, en un momento dado, confrontamos una situación similar a otras que ya han sido experimentadas por investigadores fuera de México, no nos resulte útil, al menos, informarnos de los resultados y significado de esas unidades de medición. Esto es importante, ya que sabemos que cuando construimos un test psicológico, podemos emplear el tipo de calificación, siempre y cuando estemos seguros de que son las pertinentes, tanto por la naturaleza de nuestros datos, como por los fines que persigamos con la medición planeada.

CALIFICACIONES TIPO I. COMPARACIÓN CON UN ESTÁNDAR ABSOLUTO O DIFICULTAD DE CONTENIDO

Este tipo de calificaciones, en general, resultan poco útiles para llevar a cabo pruebas psicológicas, ya que se trata de calificaciones destinadas a evaluar la ejecución máxima, la cual está determinada tanto por el conocimiento como por las habilidades y la motivación, que presuponemos son comunes en el nivel de ejecución de cualquier individuo. Generalmente, se utilizan estas unidades de calificación en pruebas de logro escolar, a fin de calificar el rendimiento en el salón de clases; es decir, se emplean para calificar el aprendizaje obtenido del material específico mostrado.

Como ejemplo de este tipo de calificación podemos mencionar los exámenes del curso de psicometría: una vez que ha sido administrado, calificado y marcado un estándar criterio como mínimo para considerar el examen anterior de 70% de respuestas correctas para obtener S de calificación (*suficiente*, calificación que ejemplifica el tipo I-B, grados por letras), se toman cada una de las puntuaciones totales correctas obtenidas por los alumnos y, aplicando la fórmula que se indica abajo en el tipo I-A, se van obteniendo las calificaciones *derivadas*, que en este caso son NA, S, B y MB, las cuales ya nos pueden dar un significado acerca del resultado del examen para cada alumno.

Si no se utilizan letras en los exámenes, pueden asignarse, como había sido tradicional en México, números que equivalgan a las letras, que pueden ser más descriptivos. Para sustituir la fórmula (no es necesario que sean cien las preguntas), se toma el número total de preguntas y la puntuación total obtenida por cada alumno, y se multiplica el cociente por 100.

Porcentaje de calificaciones correctas

Este tipo de calificaciones se utilizan, como indicamos anteriormente, solamente para pruebas de rendimiento escolar. Consisten en la comparación de la máxima calificación posible con la calificación real obtenida por cada uno de los examinados.

Considerando el número total de reactivos como el 100% de respuestas esperadas, las calificaciones obtenidas se compararán con dicho 100%, aplicando para ello la siguiente fórmula:

$$X \% = 100 R/T$$

Donde $X\%_c$ = porcentaje de calificaciones correctas,

R = número de respuestas correctas (reactivos contestados correctamente),

T = número total de reactivos en la prueba.

Por ejemplo si Juan obtuvo 40 respuestas correctas de una prueba de 50, su porcentaje será:

$$(100 \times 40) / 50 = 80\%$$

Grados por letras

Lo mismo que las anteriores, este tipo se basa en los porcentajes de calificaciones correctas. En muchas instituciones de enseñanza son utilizadas de la siguiente manera: *A* para aquellas calificaciones que oscilan entre 90 y 100; *B*, entre 80 y 90, etcétera, basándose en el promedio de

calificaciones correctas en un examen de grupo. Las limitaciones de ambos tipos de calificaciones son similares, con la única distinción de que las calificaciones de grados por letras no reflejan las pequeñas diferencias de habilidad que pudieran presentarse; sin embargo, no difieren básicamente de la calificación real del sujeto.

Este tipo de calificaciones son incompletas, ya que dependen de la dificultad de la prueba, más que de la verdadera calidad de la ejecución.

CALIFICACIONES TIPO II: COMPARACIÓN INTERINDIVIDUAL

Este tipo de calificaciones se emplea, generalmente, en las pruebas estandarizadas; son relativamente independientes de la dificultad de contenido porque están basadas en la comparación de las calificaciones de un individuo con las obtenidas, en promedio, por el grupo normativo. La dificultad de contenido puede afectar indistintamente a cualesquiera de los elementos del grupo sobre el cual se obtiene la norma (promedio); por ejemplo, el nivel de dificultad afecta igualmente a cualquier miembro del grupo, pero ocurrirá del modo distinto en cada uno de ellos, lo cual hace que se utilice la misma prueba para individuos que difieren notoriamente de su nivel de habilidad.

La utilización de una prueba que contenga un nivel de dificultad del 50% permite una discriminación muy amplia. Además, las calificaciones de comparación interindividual se emplean también en las comparaciones entre grupos.

Comparación interindividual a partir de la media y la desviación estándar del grupo.

En este tipo de unidades, la comparación se expresa a partir del número de desviaciones estándar que se encuentran entre la calificación individual y el promedio. Cualquier cambio en el grupo de comparación puede cambiar el nivel de calificación. El tipo II A consiste en calificaciones estándar lineales. Se denominan *estándar* porque se basan en la desviación estándar. Además, son lineales porque pueden ser consideradas desde el punto de vista de la distancia respecto a la media, o pueden considerarse como calificaciones que sustituyen a la media y a la desviación estándar. Todas sus características hacen que las calificaciones de este tipo sean importantes, fundamentalmente en la investigación, porque: a) para toda prueba y para todo grupo proporcionan la misma media y la misma desviación estándar; b) estas calificaciones muestran la misma estructura de la distribución de las calificaciones crudas, difiriendo sólo en el “calibre” (fineza o explicitéz de los números); c) permiten la comparación intergrupo o interprueba, a diferencia de otras unidades de medida, que no lo hacen, y d) pueden tratarse matemáticamente, lo que otras calificaciones no permiten.

CALIFICACIONES z

La base de las calificaciones estándar es la calificación z . Las demás calificaciones estándar lineales se obtienen a partir de ella. Significan, sencillamente, la diferencia o distancia entre la media de un grupo y cualesquiera de las calificaciones individuales. Su fórmula de obtención es:

$$z = \frac{X - \bar{X}}{s}$$

s
donde X = calificación individual,

\bar{X} = media de calificación para algún grupo,

S = desviación estándar de ese grupo.

(Si la calificación z media es 0.00, la desviación estándar será 1.00.)

Las calificaciones z son de gran utilidad para la investigación, pero resultan un tanto complejas para aquellos que utilizan pruebas regularmente, excepto cuando las emplean para obtener un tipo distinto de calificaciones estándar lineales, debido a que, por su naturaleza, la mitad de las calificaciones z son negativas y la otra mitad son positivas y se expresan en puntos decimales. Las demás calificaciones estándar se han diseñado con el fin de eliminar los decimales y obtener un resultado más exacto y accesible. Esto se obtiene al multiplicar la calificación z original por un valor constante, a fin de eliminar los decimales; como las calificaciones z , a partir de la media, hacia la izquierda son negativas y hacia la derecha son positivas, si agregamos un valor constante a cada calificación z eliminamos el problema de los signos.

CALIFICACIONES T

La calificación T es una de las unidades estándar lineales más comunes. La diferencia entre las calificaciones T y la z consiste en que, en aquéllas, hay una media de 50 y una desviación estándar de 10.

La fórmula de este tipo de calificaciones es:

$$T = 10z + 50.$$

Donde $z = X - \bar{X}$ (para obtener primeramente la calificación z);

10 = constante multiplicadora (cada calificación z se multiplica por 10), y

50 = constante sumadora (50 se agrega a cada valor de $10z$).

Por ejemplo: la calificación z que obtuvo Pedro fue 1.5 (1.5 es el resultado de la resta de $X - \bar{X}$, dividida entre la desviación estándar, es decir, la obtención de z); entonces, la calificación T de Pedro es $= 10 (1.5) = 15 + 50 = 65$.

Si tenemos en cuenta la curva normal de distribución (véase figura 2), encontramos inmediatamente una dirección vertical hacia el lugar correspondiente y equivalente de $1.5z$ a $65T$.

La calificación T, para ciertos propósitos de investigación, tiene las mismas ventajas y limitaciones que las calificaciones z . Desafortunadamente, tiene la desventaja de producir cierta confusión, ya que utiliza dos dígitos, para indicar que en ocasiones se trata de percentiles o de calificaciones “T escalares”, si no se especifica claramente el tipo de unidad (véase figura 2).

CALIFICACIONES AGCT

La denominación de estas calificaciones se deriva de las siglas utilizadas en la prueba de clasificación general ARMY (*alfa o beta*) Army General Classification Test. Este tipo de

calificaciones es similar a la z y a la T , a diferencia de que tiene una media de 100 y una desviación estándar de 20. Su fórmula es:

$$AGCT = 20z + 100,$$

Donde 20 y 100 = las constantes multiplicadora y sumadora, respectivamente, y
 z = calificación z definida anteriormente.

Por ejemplo: la calificación z que obtuvo Juan fue 1.5 y su calificación T fue 65; entonces su calificación AGCT será:

$$20(1.5) + 100 = 30 + 100 = 130.$$

(Si observamos la figura 2 podremos ver que la z 1.5 se encuentra en posición vertical hacia abajo directamente, en la misma posición que la T 65 y la AGCT 130.)

Las calificaciones AGCT fueron originalmente utilizadas para la selección de soldados durante la Primera Guerra Mundial. Para esta situación, la media fue de 100 y la desviación estándar 20. Podría suponerse que en otros grupos se obtuvieron resultados diferentes; sin embargo, en la aplicación de esta prueba, la media y la desviación estándar obtenidas han sido las mismas, o cuando menos muy semejantes, lo que permite comparar los resultados. Una ventaja que presentan estas unidades AGCT es que permiten un análisis posicional más exacto que las dos anteriores.

CALIFICACIONES CEEB

Se llaman así estas calificaciones debido a que fueron utilizadas por primera vez en el servicio de exámenes educacionales para alumnos de nuevo ingreso por el College Entrance Examination Board; estas calificaciones son similares a las calificaciones estándar lineales, a diferencia de que aquéllas analizan detalladamente la posición de los sujetos dentro de la curva normal de distribución; debido al número de casos estudiados, se opera con una media de 500 y una desviación estándar de 100. Su fórmula es:

$$CEEB = 100z + 500,$$

Donde 100 y 500 = constantes multiplicadora y sumadora, respectivamente, y
 z = calificación z original.

Por ejemplo: si Juan obtuvo una z de 1.5, su calificación CEEB será de 650 (véase figura 2).

Las calificaciones CEEB se han utilizado año tras año en la selección estudiantil y, sin embargo, siempre han producido la misma media y la misma desviación estándar, lo cual permite la comparación de los resultados de un año a otro.

CALIFICACIONES CIs (estándar)

El CI (cociente intelectual), ideado hace aproximadamente 60 años por el psicólogo alemán William Stern, está basado en la relación existente entre la edad mental y la edad cronológica. Un gran avance en este aspecto ha sido el desarrollo de un cociente intelectual basado en la desviación estándar que contenga todas las edades mediante la prueba sobre la cual se determina.

El término *cociente intelectual de desviación estándar (CIs)* se emplea para describir tres tipos de calificación diferentes. En esta sección nos referimos sólo al primer tipo, es decir, a la calificación estándar lineal. Este tipo de calificación tiene las mismas ventajas y limitaciones que otras calificaciones estándar, a excepción de que el autor de la prueba puede, a partir de una media de 100, fijar su desviación estándar según las metas que se haya trazado. Así, tenemos, por ejemplo:

CI de Wechsler

Está constituido por las pruebas de inteligencia de Wechsler, que son: la escala de inteligencia de Wechsler para niños en edad preescolar y para niños de primer año de primaria (WPPSI), la escala de inteligencias de Wechsler para niños (WISC) y la escala de inteligencia de Wechsler para adultos (WAIS).

Todas estas escalas, aunque difieren en algunos aspectos, determinan análogamente sus calificaciones CIs estándar. Un ejemplo de lo anterior lo constituye la prueba de WAIS, la cual consta de seis subpruebas verbales y cinco de ejecución que, al combinarse, producen tanto un CIs verbal como un CIs de ejecución y, finalmente, un CIs total. Esta prueba abarca un rango de edad de 16 años o más y utiliza siete tablas de normas de grupos diferentes.

Se aplica una calificación cruda para cada una de las once subpruebas, las cuales son convertidas a calificaciones estándar normalizadas (véase tipo II B 4e) con una media de 10 y una desviación estándar de 3. Una vez obtenida la suma de las calificaciones estándar normalizadas, se convierte a una CIs de desviación con la ayuda de la tabla de calificaciones estándar que se obtuvo separadamente para cada uno de los siete grupos de edades.

Al construir la prueba se obtuvieron la media y la desviación estándar de cada una de las pruebas que, sumadas, dieron resultados separados para cada grupo de edades. El autor determinó que su media debería ser 100, con una desviación estándar de 15, mediante el empleo de la fórmula $CIs = 15z + 100$ obtenida independientemente para cada grupo. De esta manera, para utilizar la prueba actualmente, sólo se necesita consultar la tabla correspondiente, para encontrar el valor de CI por medio de la suma de las calificaciones estándar de las subpruebas. Los CI estándar verbal y de ejecución se obtienen del mismo modo, el primero sobre seis subpruebas y el segundo sobre cinco.

CI de Stanford- Binet

En la revisión de la prueba de Binet y Simon (1905) en el año de 1960, llamada Stanford-Binet (S-B), se utilizaron por primera vez *CIs* de desviación estándar, ya que los autores de la revisión consideraron que podría ser más constante de edad a edad.

Las tareas que exige la prueba S-B han sido dispuestas por niveles de edad desde los dos años hasta niveles de adultos superiores. Mediante el procedimiento conocido, el examinador debe encontrar la edad mental (EM) para cada una de las pruebas. Esta edad mental, a diferencia de la edad cronológica de los examinados, se anota en una tabla (ajustándola cuando se refiere a los adultos); de esta manera, se obtiene el CI.

Para cada grupo de edad cronológica de los examinados se obtuvieron la media y la desviación estándar de la edad mental. La edad mental media (100) fue igual para todas las personas de una misma edad cronológica. Los autores encontraron que la edad mental fue una desviación estándar inferior a la media, o sea un CIs de 84. A una desviación estándar superior a la media correspondió un CIs de 116, y así sucesivamente.

Así en general, en el caso del S-B tenemos una calificación lineal con una media de 100 y una desviación estándar de 16; por tanto la fórmula queda:

$$CI = 16z + 100$$

Como puede observarse, la desviación estándar es diferente de las pruebas de Wechsler, a pesar de que ambas utilizan la misma unidad de medida; la explicación de esto puede ser que, en las pruebas S-B, además de la consideración del error de medición, hay diferencia tanto en el contenido como en las tablas de las normas de los grupos.

Las calificaciones estándar lineales denotan la localización de una calificación cruda de un examinado, en relación con la media de un grupo específico y en términos de su desviación estándar. Ya sea que se trate de una distribución normal o no, podemos convertir las calificaciones crudas dadas a cualquier tipo de calificación estándar lineal, lo cual no alterará la forma original de la distribución.

Comparación interindividual a partir del rango de los grupos

A diferencia de las calificaciones del tipo II A, estas calificaciones se basan en el número de personas con calificaciones más altas o más bajas, en vez de utilizar un valor de calificación específico; no hay información acerca de la distancia que existe hacia la media, y producen el efecto de una distribución más aproximada a la normal que la distribución de calificaciones crudas en las cuales están basadas.

Rangos

Entendemos por rango la posición relativa de una calificación cruda a lo largo de la curva normal; es decir, la primera para la más alta o mejor, la segunda para la que sigue, etcétera. El problema que presenta este tipo de calificaciones es que nunca pueden utilizarse para proporcionar resultados generales de ninguna prueba.

RANGOS PERCENTILES Y CADENAS PERCENTILARES

Los rangos percentiles, en ocasiones llamados *centiles*, constituyen quizá una de las calificaciones más utilizadas en las tablas de normas de pruebas estandarizadas, porque permiten en general una mejor interpretación de los resultados, aún cuando tienen ciertas limitaciones, como se podrá notar posteriormente.

Un *percentil* es cualquiera de los 99 puntos de igual amplitud en que se divide una distribución de frecuencia dentro de un grupo de 100. Un *rango percentil* es la posición relativa de un individuo dentro de un grupo específico.

Para obtener percentiles y rangos percentilares, deben observarse las reglas siguientes:

- a) Ordenar todas las calificaciones crudas en orden de menor a mayor.
- b) Tabular la frecuencia con que aparece cada una de ellas.
- c) Encontrar la frecuencia acumulada a través de cada una de las frecuencias de las calificaciones, sumando la frecuencia de la calificación más baja con la frecuencia inmediata, y así sucesivamente, hasta llegar a la frecuencia de la calificación más alta.
- d) Hallar el punto medio de cada una de las frecuencias acumuladas, añadiendo un cero a la frecuencia acumulada más baja; el resultado de esta suma se divide entre dos, y este resultado es el punto medio. El siguiente punto medio se obtiene sumando la frecuencia acumulada más baja con la frecuencia acumulada y dividiendo este resultado entre dos, y así sucesivamente, hasta llegar a la frecuencia acumulada mayor.
- e) Convertir a porcentaje acumulado mediante la fórmula $Pc_{pm} = 100 (fc_{pm}) / N$,

Donde Pc_{pm} = porcentaje acumulado del punto medio,

Fc_{pm} = frecuencia acumulada del punto medio,

N = número de casos, y

100 = la constante que multiplica sucesivamente a los valores del fc_{pm} .

- f) Encontrar los rangos percentiles, completando los valores de los puntos medios a números enteros, a excepción del 1 para el 0 y 99 + para el 100.

Para obtener las calificaciones crudas equivalentes a los valores percentilares establecidos, deben desarrollarse los siguientes pasos:

- a) Disponer las columnas del 1 al 3, como se indicó anteriormente.
- b) Cambiar de percentiles a número de casos, multiplicando por $N/100$.
- c) Contar el número de casos encontrados en el paso b), suponiendo que los casos están distribuidos igualmente a través de cada una de las calificaciones.
- d) El valor de la calificación cruda correspondiente es el percentil deseado.

La ventaja principal de los percentiles es la facilidad de su interpretación. Aún cuando hay quienes afirman que los percentiles se encuentran distribuidos en espacios iguales (lo que podría ocurrir sólo si un mismo número de personas obtuvieran cada una de las calificaciones crudas), se considera que un percentil es la obtención del porcentaje de casos en un grupo específico, que caen hacia arriba o abajo del valor de una calificación dada. Por otra parte, es fácil observar las diferencias cerca de la mediana y las diferencias en los extremos.

Dado que las distancias interpercentilares no son iguales, no son promediabiles en forma directa, como en el tipo II A de calificaciones. Con percentiles podemos utilizar escalas comunes para la distribución de todos los valores en la prueba.

Si dividimos en cien subgrupos de igual medida la distribución de calificaciones ordenadas por rango (véase figura 1), observaremos que hay 99 puntos percentilares distribuidos a lo largo de los cien subgrupos, los cuales, al quedar clasificados por rangos, los separamos mediante una línea con el fin de indicar los límites reales de cada rango percentilar. Noventa y nueve de estas unidades tienen 0.5% hacia cada uno de los extremos de la distribución: los extremos 1 y 99+.

Finalmente, el empleo de los percentiles presenta la desventaja de que en ellos se utiliza una escala de números fácilmente confundibles si no se estipula claramente con qué tipo de unidades se trabaja: con las calificaciones T, con los CI o con porcentajes de respuestas correctas.

Una aplicación de los rangos percentiles es la llamada *cadena percentilar* que, como su nombre lo indica, es una agrupación de rangos percentilares. El límite superior de la cadena corresponde al rango percentil de una calificación sobre un error estándar de medida superior a la calificación obtenida, y al límite inferior de la cadena corresponde al rango percentil con un error estándar de medida inferior a la calificación obtenida. El psicólogo puede hacer sus propias cadenas percentilares teniendo en cuenta las indicaciones expresadas anteriormente acerca del error estándar de medida.

CALIFICACIONES ESTÁNDAR NORMALIZADAS (TRANSFORMACIONES POR ÁREA)

Estas son calificaciones derivadas, a las cuales se les asignan calificaciones estándar como valores, pero se constituyen más adecuadamente como rangos percentilares. Cuando se emplean calificaciones estándar, la forma de la distribución de las calificaciones crudas se reproduce con gran exactitud. Si la línea de base representara un polígono de frecuencia, veríamos que los valores de cualesquiera de las calificaciones estándar serían correspondientes si se trazara una línea hacia debajo de la distribución de las calificaciones crudas. Este tipo de calificaciones estándar normalizadas, como su nombre lo indica, tienen la característica de permitir que una distribución sea lo más aproximadamente posible a la distribución normal.

Para obtener calificaciones estándar normalizadas se utiliza el siguiente procedimiento:

- i) Ordenar las calificaciones crudas.
- ii) Encontrar la frecuencia de ocurrencia de cada calificación.
- iii) Hallar la frecuencia acumulada a través de cada calificación.
- iv) Encontrar el punto medio de la frecuencia acumulada de cada calificación.
- v) Convertir a porcentajes acumulados el punto medio de la frecuencia acumulada.
- vi) Sustituir el valor de la calificación estándar normalizada correspondiente por su porcentaje acumulado en una distribución normal de probabilidad. *Estos valores, diferentes para cada calificación estándar normalizada, pueden encontrarse en una tabla de transformación para calificaciones derivadas.*

Estas calificaciones se denominan *áreas de transformación*, debido a que están basadas en calificaciones estándar que podrían corresponder a porcentajes acumulativos en una distribución normal.

Calificaciones T escaladas

Tienen las mismas propiedades que las calificaciones T en una distribución normal e, incluso, la misma media de 50 y la desviación estándar de 10; por ello, se denominan calificaciones T, más bien que T escaladas; sin embargo, tienen la desventaja de que son fácilmente confundibles con las calificaciones T y con otras más.

Calificaciones estaninas

Son cadenas de valores (nueve unidades estándar), a cada una de las cuales corresponde la mitad de una desviación estándar, a excepción de las estaninas de los extremos, la 1 y la 9. La media se encuentra en el punto medio de la estanina 5, la cual ha sido utilizada como calificación estándar normalizada con una media de 5 y una desviación de 2, teniendo las estaninas 1 y 9 la mitad de una desviación estándar de amplitud.

Calificaciones C escaladas

Este tipo de calificaciones fueron elaboradas por Guilford para dar una unidad adicional en cada uno de los extremos de la escala de estaninas; es decir, tienen once unidades con valores de 0 a 10. Se construyen de modo análogo a las estaninas, pero los valores dados se distribuyen diferentemente.

Calificaciones estén

Son similares a las anteriores, es decir, son calificaciones estándar normalizadas, con diez unidades estándar. Tienen cinco unidades normalizadas a cada lado de la media y poseen media desviación estándar de amplitud, excepto en los extremos 1 y 10, que son abiertos. Este tipo de calificación es utilizado en algunas de las pruebas de Catell y su procedimiento de obtención es igual que los anteriores, con la única diferencia de que los valores asignados son distintos. (Véase la figura 1, a fin de localizar todas estas calificaciones.)

CI de desviación

Aún cuando son poco usuales para la normalización los CI de desviación estándar en la actualidad, anteriormente fueron muy utilizados. Presentan las mismas ventajas y limitaciones que las calificaciones T escaladas, a excepción de que su media deberá ser de 100 y la desviación estándar puede ser determinada por el autor de la transformación de las calificaciones.

Calificaciones ITED (Iowa Test of Educational Development)

Son calificaciones estándar normalizadas que tienen una media de 15 y una desviación estándar de 5. Este tipo de calificaciones fueron obtenidas en una muestra representativa de estudiantes de décimo y decimoprimer años en los Estados Unidos.

RANGOS DECILES

El *decil* se define como cualquiera de los nueve puntos que dividen la distribución de frecuencias en diez grupos de igual medida o amplitud. Algunas de las pruebas de Catell utilizan

esta unidad de medición en sus tablas de normas de calificación, pero en ellas se introducen algunas variaciones que modifican el significado de decil, incluyendo una cadena (rango) en el 10% de los casos, 5% para cada lado, en cada medida o amplitud del punto decil real. A fin de evitar una confusión entre los términos deciles y calificaciones deciles, es más adecuado llamarles rangos deciles.

Comparación interindividual a partir del rango de calificaciones en un grupo

Existe un solo tipo de calificación dentro de esta clasificación, al cual llamaremos *calificaciones jerárquicas porcentuales* que se basan en la comparación interindividual a partir del rango de calificaciones crudas.

Este tipo de calificación indica la posición de una persona en una escala de 101 puntos, en donde la calificación más alta asciende hasta 100 y la más baja desciende a 0. La fórmula es:

$$X \% = 100 \frac{(X - B)}{(A - B)}$$

donde X = cualquiera de las calificaciones crudas,

B = calificación cruda más baja,

A = calificación cruda más alta,

(jr) = calificación jerárquica (en esta fórmula califica el porcentaje de calificaciones crudas).

Comparación interindividual a partir de la posición de aquéllos que obtuvieron la misma calificación

Este tipo de calificaciones se emplea para expresar la ejecución en términos de promedios de grupos que difieren en situación, ya sea de edad cronológica o de calificaciones escolares jerarquizadas. En estas unidades de calificación, el examinado no es comparado con su propio grupo sino con diversos grupos que difieren en nivel, a fin de observar, supuestamente, a cuál se asemeja o a cuál podría pertenecer.

Este tipo de calificaciones es utilizado frecuentemente en la estandarización de pruebas de logro escolar y pruebas de inteligencia para niños en edad escolar. Parecen de fácil comprensión, pero tienen muchas limitaciones, difícilmente apreciables.

CALIFICACIONES POR EDAD

Este tipo de calificaciones resulta útil para cualquier característica humana que sufra una modificación con el desarrollo del individuo, especialmente en niños escolares. La calificación de este tipo empleada más frecuentemente es la *edad mental (EM)*, que como se sabe, es un concepto establecido por Binet alrededor de 1900, quien la utilizó en las primeras pruebas de inteligencia.

Una calificación de edad es la expresión de la ejecución de una tarea que se le pide realizar a un examinado y que está basada en el promedio del nivel de desarrollo que por su edad le corresponde al sujeto.

Debe tenerse cuidado en la interpretación de las edades mentales, ya que, dentro del rango de los 5 a los 15 años, su significado es perfectamente aproximado a la edad cronológica correspondiente. Así, sería impropio suponer que un adulto que muestra déficit mental corresponda a la edad cronológica promedio del grupo de sujetos de 0 a 6 años. Los adultos tienen habilidades motoras y hábitos diferentes de los que generalmente tiene un niño, aún cuando éste pueda ejecutar mejor que un adulto deficiente un gran número de tareas.

Una de las principales dificultades de interpretación de estas calificaciones es la de que las desviaciones estándar son diferentes de prueba a prueba e, incluso, de edad a edad, con la misma prueba. Por tanto no hay manera de generalizar los valores de las calificaciones por edad, cualquiera que sea la distancia establecida a partir de la media; por ejemplo, una edad mental de 13-3 para un niño de 12-3 no indica el mismo grado de superioridad que una edad mental de 6-3 para un niño de 5-3. Las calificaciones de edad mental son problemáticas en su interpretación porque parecen realmente significativas, y en muchas ocasiones no lo son. En la escuela, un maestro puede intentar evaluar las potencialidades de sus alumnos; puede notar que algunos de ellos tienen dificultades en el aprendizaje; al remitirlos al psicólogo escolar, éste probablemente tendrá que utilizar pruebas individuales que le ayuden a entender la problemática de los niños en lo individual y ratificar la validez de la indiferencias hechas a partir de las calificaciones de las pruebas del grupo. En realidad, este tipo de calificaciones no debe utilizarse en situaciones de selección de personal, ya sea industrial, escolar, etcétera.

Dentro de este grupo de calificaciones tenemos la llamada *edad educacional (EE)*, que indica la ejecución de una tarea hacia un nivel específico expresado como la edad promedio de los individuos que en ese nivel deben ejecutarla. Esta edad puede ser edad de rendimiento escolar, edad de lectura, de matemáticas, de biología, etc. Este tipo de calificación supone que los niños adquieren conocimientos y habilidades más o menos en forma uniforme a través del año escolar; es decir, que aprenden mes tras mes al mismo ritmo lo que el maestro está enseñando. Por otra parte, estas calificaciones suponen que la edad es relevante para determinar el nivel de ejecución de una tarea de un niño, más que su calificación escolar jerárquica. Estas dos suposiciones, desde nuestro punto de vista, son falsas ya que en la determinación de habilidades intelectuales y motoras existe una gran cantidad de factores ambientales (el hogar, los sistemas escolares, los maestros, etc.) que influyen considerablemente para modificar el desarrollo y la ejecución de las tareas que se imponen para determinar calificaciones intraindividuales, a fin de establecer cuáles habilidades se han desarrollado mejor que otras en un mismo individuo, pues existen otros tipos de calificaciones más exactas para evaluar los niveles mencionados.

CALIFICACIONES TIPO III: COMPARACIÓN INTRAINDIVIDUAL (DE DOS MEDIDAS DEL INDIVIDUO)

Cociente intelectual de razón

Debemos hacer hincapié en que este cociente intelectual (CI) es diferente del mencionado anteriormente, el cual se refiere al cociente intelectual de desviación estándar.

El cociente intelectual de razón (CI) fue propuesto por Stern y utilizado por Terman desde hace más de cincuenta años. La fórmula para obtener este cociente es:

$$CI = EM \times 100 / EC,$$

Donde EM = edad mental,
EC = edad cronológica,
100 = constante multiplicadora

Este cociente es muy común; sin embargo, sus limitaciones son desconocidas. La calificación depende de la suposición de unidades de edad mental de igual medida, lo cual puede no existir. Este tipo de CI, para su uso, se limita en forma razonable de los 5 a los 15 años; no obstante, para edades superiores puede ser muy útil.*

Índice de nivel intelectual

Este tipo de calificación es similar al anterior en cuanto a su obtención, pero difiere en cuanto a que tiene como base el supuesto de que la calificación del alumno en una prueba de inteligencia está determinada por su posición académica en la escuela, más que por su edad cronológica. Esta puede ser razonable, pero debe tenerse cuidado en la amplitud de la desviación estándar en los diferentes niveles de edad.

Cociente educacional

Éste se obtiene dividiendo la edad educacional entre la edad cronológica y multiplicándola por 100; generalmente se emplea en las pruebas de rendimiento escolar. Por sus grandes limitaciones, no se pueden hacer comparaciones entre una materia y otra, aún cuando las pruebas hayan sido estandarizadas con el mismo grupo.

Por último, debemos indicar que existen otros tipos de calificaciones, creadas o propuestas para situaciones específicas de quien las utiliza o de quien construye las pruebas.

No debemos olvidar que la calificación de una prueba psicológica, antes de ser interpretada, ha de ser comprendida, y que el empleo de cualquier unidad de calificación en psicología debe ser planeado, antes de estipularlo, en función de los objetivos que se persiguen con la medición. Una calificación no puede ser significativa si carece de validez, tiene baja confiabilidad o se trata de una prueba de poca calidad; así, aún cuando las calificaciones derivadas pueden tener más significado que las calificaciones crudas, también pueden ser menos exactas si no se tiene cuidado en utilizarlas adecuadamente.

* No confundir en su expresión el CI que deberá realmente ser CIS (de Stern) con el CIs (desviación estándar).

Entendemos por *norma* el promedio o calificación “típica” sobre una prueba dada, elaborada para una población determinada; por ejemplo, la media o promedio en niños de diez años en una prueba de inteligencia o de elementos de aritmética o de lectura, etc. Remitirnos a una tabla

de normas nos permite ordenar la ejecución relativa de un individuo dentro de su grupo o dentro de otros grupos por edad, sexo, grado, etc.

NORMAS DE CALIFICACIÓN

Las normas de calificación son de fundamental importancia para comprender los resultados de una prueba, ya que proporcionan los parámetros que permiten comparar la ejecución en dicha prueba. Las calificaciones crudas (es decir, el número real de respuestas o puntos obtenidos por un individuo), en sí mismas, no tienen ningún significado; por tanto, no pueden ser comparables con los puntos obtenidos en otras pruebas.

Obviamente, si las calificaciones obtenidas en cada una de las diversas pruebas necesitan ser comparadas, deben utilizarse índices que expresen el significado relativo de cualquier calificación dada, al cual se le ha denominado *rango relativo*. Así, un sujeto puede obtener en una prueba 65 puntos, en otra 105 y en una tercera 120; la primera prueba puede tener un rendimiento máximo de 150, en la segunda 200, y en la tercera 300. En este caso resulta difícil la comparación, ya que ninguna calificación de esas pruebas permite comparar las puntuaciones. Asimismo, es probable que la desviación de cada prueba sea absolutamente diferente, lo mismo que la desviación estándar, de tal manera que cada una puede tener una media y una desviación estándar diferentes, difícilmente comparables. Siguiendo el ejemplo anterior, si suponemos que las tres pruebas son aplicadas al mismo grupo, el promedio será de 90 para la primera, de 120 para la segunda y de 180 para la tercera; es decir, los individuos que obtienen estas calificaciones pueden estar colocados dentro del promedio en cada una de las pruebas. Igualmente, las calificaciones 75, 100 y 150 tienen el mismo “significado relativo” en las respectivas pruebas, y pueden dar una idea más clara de la posición que guardan los individuos, a partir del promedio.

Para facilitar la interpretación de calificaciones de pruebas psicológicas se han elaborado ciertas tablas, denominadas *tablas de normas de calificación*. Estas pueden ser de varios tipos y contener diversa información; por ejemplo, pueden constar de normas de edad, normas de calificaciones escolares, normas de rangos deciles, normas de rangos percentilares, normas de calificaciones estándar, etc. Al tener a nuestro alcance una tabla de normas, estamos en condiciones de ordenar la ejecución relativa de un individuo dentro de su propio grupo o dentro de otros grupos por edad o por grado.

Las normas de calificación de una prueba están representadas por una tabla que contiene valores correspondientes de calificaciones crudas y calificaciones derivadas. Estas últimas, como hemos visto, facilitan la interpretación de las pruebas dándoles el mayor significado y exactitud posible; lo cual las calificaciones crudas, por sí mismas, no pueden proporcionar. Una tabla de normas deberá contener las calificaciones crudas y sus correspondientes calificaciones derivadas, incluyendo la descripción del grupo sobre el cual se han elaborado. Puede presentar uno o varios tipos de calificación derivada de uno o varios grupos; de aquí que existan diversos tipos de tablas de normas de calificación, las cuales analizaremos a continuación.

Tablas de normas simples

Estas tablas constan de dos columnas: una contiene las calificaciones crudas y la otra las calificaciones derivadas; por ejemplo, cualquier prueba de rendimiento escolar o cualquier test en el que se muestren las calificaciones crudas de cada sujeto y su correspondiente calificación derivada.

Tablas de normas de grupos múltiples

Este tipo de tablas se ha elaborado para expresar los resultados de varios grupos diferentes en una sola tabla, lo cual permite la comparación de las calificaciones de una persona con tantos grupos como se desee. Podemos citar como ejemplo la prueba de matrices progresivas de Raven, que se estudia en las prácticas de psicometría. Pueden observarse los grupos por edad, en rangos de 5 en 5, a los cuales corresponde una calificación percentil dada también en rangos.

Tablas de normas de calificaciones múltiples

En este tipo de tablas se incluyen calificaciones de varias pruebas o subpruebas. Para situaciones prácticas y a fin de hacer una utilización adecuada de estas tablas, necesitamos conocer y entender la naturaleza de las unidades de calificación psicométrica. Un buen ejemplo de éstas son las escalas de Wechsler, cuyo entrenamiento para manejarlas se lleva a cabo en las prácticas de psicometría de este curso.

Tablas de normas abreviadas

Ocasionalmente, estas tablas contienen sólo valores de calificaciones crudas alternados. Esta situación obliga al examinado a interpolar aquellas calificaciones crudas que no aparecen en la tabla. Obviamente, este tipo de tablas, puesto que producen errores y problemas para quien las utiliza, tienen una función económica, aún cuando su costo sea bajo. Son poco útiles y, afortunadamente, ha sido generalizándose su aplicación. Otra vez nos sirven de ejemplos los baremos de Raven, en los que puede observarse la agrupación de calificaciones bien por cuarteles o por intervalos de cierta amplitud.

Tablas de normas condensadas

Este tipo de tablas son análogas a las abreviadas. Se utilizan fundamentalmente para propósitos de comparación cuando hay una gran cantidad de datos que deben incluirse en una sola tabla; un buen ejemplo vuelven a ser las escalas de Wechsler. Hemos dado los ejemplos anteriores intencionalmente por ser los tests que normalmente se estudian en las prácticas de psicometría. Es obvio que podríamos dar otros más, pero el aprendizaje de éstas pueden transferirse a otras.

Tablas de expectancia

Estas tablas difieren de las tablas de normas en que, en tanto que éstas establecen valores de calificaciones derivadas que corresponden a cada una de las calificaciones crudas, aquéllas muestran el criterio de ejecución por intervalo de las calificaciones crudas. No obstante, las tablas de expectancia también se consideran tablas de normas, ya que establecen resultados para cualquier

grupo específico. Cuando interpretemos las calificaciones de determinada persona en una tabla de expectancia o en una tabla de normas, debemos tener en cuenta si el grupo y la situación son comparables.

ARTICULACIÓN DE LAS NORMAS

La articulación de las normas se refiere a la variabilidad de una prueba específica en edición, forma y/o nivel.

La edición se refiere a la fecha en que una prueba ha sido puesta en uso; para ello, debe estar al corriente el contenido de la prueba; por tanto, es necesaria la edición de normas actualizadas.

La forma alude al hecho de que no existan variaciones en las diferentes versiones que se presenten, aún cuando se actualicen; es decir, aunque cambie la estructura formal de los reactivos, el contenido y el grado de dificultad deben ser conservados. Se sabe que existen formas paralelas de las pruebas para asegurar su coeficiente de equivalencia, a fin de disminuir la probabilidad de que los reactivos de la prueba sean omitidos por los examinados. Puede ocurrir que haya un mismo contenido, pero que las formas paralelas se califiquen de diferente manera; es decir, que una forma sea calificada manualmente y la otra por computadora, lo cual ocasionará posiblemente algunas diferencias en los resultados.

El nivel se refiere generalmente a la edad o a la jerarquización de calificaciones escolares en las que se encuentran los individuos que han sido examinados en una prueba específica. Puede ocurrir que sea necesario construir una prueba en la cual el contenido deba cubrir un rango de edades de tres hasta ocho niveles diferentes, dependiendo de la situación. En ese caso, la tabla deberá especificar claramente su contenido.

A menudo sucede que las pruebas tienen una sola edición, forma y nivel. Esto restringe la utilización de los instrumentos o pruebas. Por tanto, es menester hacer un uso adecuado de ellos, realizando sistemáticamente las versiones tanto en edición, como en forma y nivel; es decir, estandarizando adecuadamente y publicando tales estandarizaciones de manera que queden perfectamente especificadas tanto la amplitud con que pueden utilizarse los documentos, como las características de las muestras a las que corresponden las versiones emitidas; de otra manera, se cometerán una serie de errores de calificación, clasificación e interpretación sumamente graves en el trabajo de investigación.

5

APLICACIÓN DE LAS PRUEBAS PSICOLÓGICAS A LA MEDICIÓN DE LA INTELIGENCIA

El término inteligencia ha sido utilizado con cierto recelo, debido a que, como veremos posteriormente, tiene diversas connotaciones. Lo empleamos sólo con el propósito de describir un aspecto conductual que, a pesar de ser infinitamente complejo, nos permita expresar “algo” que pueda ser comprendido con facilidad por aquellos que estudian el campo de la aplicación de la medición en psicología.

Para poder referirnos a la aplicación de las pruebas psicológicas a la medición de la inteligencia, es necesario primero remontarnos al estudio e investigación de la naturaleza de este fenómeno psicológico, a fin de observar cómo ha venido evolucionando y cómo los estudiosos de esta materia han intentado comprender la inteligencia y han establecido teorías acerca de su naturaleza, a la vez que han desarrollado las técnicas e instrumentaciones apropiadas, para llegar a comprobar sus hipótesis planteadas.

No podemos separar la investigación acerca de la naturaleza de la inteligencia, del desarrollo simultáneo de instrumentos de medición que puedan demostrar y comprobar que las hipótesis planteadas acerca de la misma ratifican el concepto que de ellas se ha establecido. Por tal razón, en este capítulo describiremos primero las diversas connotaciones, aparentemente diferentes de lo que se ha entendido por *inteligencia*.

Desde luego, podríamos iniciar la historia del desarrollo del concepto con pensadores tan notables como Platón y Aristóteles, quienes planteaban la necesidad de clasificar a los individuos a partir de su “facultades mentales”. Sin embargo, consideramos que el momento en que se inicia el estudio e investigación exhaustiva y que ha conducido a lo que en la actualidad se ha considerado fundamental en la comprensión y estudio tanto de la teoría acerca de la inteligencia como del manejo de la instrumentación para su evaluación, ocurrió durante la segunda mitad del siglo pasado.

Históricamente hablando, existen tres aproximaciones fundamentales para el estudio de este aspecto psicológico: la *psicología experimental*, las *investigaciones genéticas* y los *métodos multivariados*.

Los psicólogos experimentales han realizado sus investigaciones en los laboratorios, tratando de encontrar una solución a los problemas psicológicos básicos, haciendo estudios acerca de la percepción, cognición, aprendizaje y memoria, destinados a la comprensión de la inteligencia; pero poco entendidos por quienes emplean las pruebas psicológicas.

Desde el punto de vista genético, las aportaciones obtenidas conducen a la comprensión de cómo los seres humanos adquieren y emplean el conocimiento como punto fundamental para la comprensión de la inteligencia y las operaciones que ejecuta la mente humana. Piaget, al referirse a la inteligencia y como representante principal de este tipo de aproximación, distingue tres aspectos de ella: el *contenido*, la *función* y la *estructura*. Así, explica que el contenido es la expresión observable de la conducta; es la información con la cual los psicólogos llevan a cabo sus trabajos. La función se refiere a los principios de actividad intelectual, los cuales están generalmente relacionados con la edad o con el estado de desarrollo del individuo; de este modo, Piaget afirma que la función está constituida por conceptos y leyes que los científicos infieren del contenido observado. La estructura se refiere fundamentalmente al conocimiento, el cual varía con la edad y la experiencia y se desarrolla a través de la actividad. La estructura de la inteligencia es el enfoque principal de las investigaciones de Piaget, debido a su interés por el estudio de la genética. Siendo él, fundamentalmente, un teórico de la psicología cognoscitiva, muestra sumo interés en la actividad sensoriomotriz, especialmente en el infante. Considera que el desarrollo cognoscitivo del niño se inicia cuando llega al ambiente externo con una estructura biológica que ha sido conformada dentro de él, a partir de la herencia y solamente con pocos reflejos, como el chupeteo, el pataleo, el movimiento incontrolado de los brazos, movimientos de prensión, etc.

La aproximación multivariada sobre el estudio de la inteligencia, es decir, aquella que considera la existencia de un conjunto de variables que conforman el *comportamiento inteligente*, contrasta con el método experimental bivariado (de dos variables) que tradicionalmente se ha utilizado. Esta aproximación maneja, en el contexto de la inteligencia, los llamados factores a fin de explicarla a partir de habilidades intelectuales básicas y diferenciadas.

Los métodos multivariados más apropiados son las interrelaciones y el análisis factorial.

La razón de esta forma de plantear el análisis de la inteligencia se debe a que, puesto que existen diferencias individuales que originan variaciones significativas en la conducta, la única manera de entender el problema es tratándolo analíticamente, más que en forma generalizada; es decir, se considera que, más que una inteligencia general, existen habilidades específicas primarias y básicas, además de que existen tanto diferencias interindividuales como intraindividuales, dependiendo no sólo de las que potencialmente tiene el individuo al nacer, sino de la facilitación u obstaculización que el ambiente le proporcione para el desarrollo de todas y cada una de tales habilidades. Así, los métodos multivariados permiten, por una parte, describir a cada persona conforme a un perfil de funcionamiento intelectual, más que de acuerdo con un solo membrete, y por la otra, comparar sus expresiones conductuales intelectuales con las del grupo de individuos que sea necesario, según la finalidad que persigan las investigaciones.

Los estudios científicos relacionados con el hombre se iniciaron hace cerca de 2000 años con animales inferiores; sin embargo, las investigaciones sistemáticas acerca de las habilidades humanas fueron tardías y aún el campo es amplio para explorarlo. No obstante, con gran influencia para la psicología, los estudios de Charles Darwin acerca de la vida de los organismos inician un periodo de gran actividad en la investigación científica. Si se planteaba “la supervivencia del más apto entre las plantas y los animales inferiores”, debería existir también “la supervivencia del más apto entre los seres humanos”. La mayor parte de la actividad humana está dirigida a la adaptación o ajuste al ambiente. Si entre los animales de la misma especie existen diferencias en cuanto al tipo de supervivencia, las diferencias individuales entre los seres humanos son, evidentemente, similares. Darwin estableció que las características físicas de los animales se transmiten de generación a

generación y que su tipo de supervivencia ocurre a cambios lentos; asimismo, enunció que algunas diferencias individuales en las personas pueden verse influidas por la herencia.

Darwin mostró un gran interés acerca de sus teorías, lo cual motivó que muchos investigadores se abocaran a la tarea de estudiar estos problemas. Así, Francis Galton, en Inglaterra, se convenció de que no sólo se heredaban rasgos humanos de tipo físico, como la estatura, el color de la piel, de los ojos, etc., sino que también se heredaban habilidades y características de personalidad. Como científico que era, Galton consideró necesario establecer los principios mediante los cuales los rasgos son transmitidos de padres a hijos, para lo cual se requería llevar a cabo la medición de ellos. Debemos indicar que fue él quien implantó el tratamiento de la *eugenesis*, es decir, el propósito del mejoramiento de las razas humanas a través de un proceso selectivo, aún vigente en Inglaterra.

Galton estableció el término “prueba mental” para referirse a la medición de algunos de los atributos humanos, haciendo hincapié en la necesidad de estandarizar los métodos y los instrumentos de medición; es decir, los individuos sometidos a la solución de problemas deben estar bajo las mismas condiciones y las mismas instrucciones. Así, este investigador suponía que todo conocimiento necesariamente debería pasar a través de los sentidos, y que la persona con mayor agudeza sensorial puede ser la mejor dotada y la que tenga mayor facilidad para la adquisición del conocimiento. La mayoría de las pruebas construidas por Galton se relacionaron con la discriminación sensorial; por ejemplo, habilidad para discriminar niveles de sonidos, tonos de colores, agudeza visual y otras funciones sensoriales. Fundó en 1884 el primer laboratorio antropométrico, en el sur de Kensington Museum, en donde mediante un bajo precio se podían llevar a cabo mediciones a través de una serie de pruebas sensoriales y físicas que incluían la estatura, el peso, la capacidad respiratoria, la presión manual, la audición, las capacidades perceptivas, las capacidades visuales, etc. Para poder analizar los datos obtenidos, Galton utilizó por primera vez métodos estadísticos en este tipo de investigaciones, a fin de determinar las medias o promedios de ejecución y la dispersión de tales conductas; pero su principal interés consistió en establecer la semejanza entre las características de padres e hijos, para lo cual consideró que era necesaria una medida de asociación, dando así los primeros pasos para el desarrollo del análisis correlacional. Los trabajos de Galton surgieron en un momento relevante, ya que en ese tiempo aún no se había tratado el problema de las diferencias individuales que hasta la actualidad han sido trascendentales en el desarrollo de la psicología.

Fue Karl Pearson quien entusiasmado con el éxito que había obtenido Galton, desarrolló los métodos estadísticos para el estudio de las diferencias individuales. Así, desarrolló el coeficiente de correlación producto-momento (PM), la correlación múltiple y la correlación parcial; asimismo, estableció las bases del análisis factorial y otros métodos de análisis multivariado. En tanto que Galton puntualizó la necesidad de llevar a cabo operaciones matemáticas para el correcto manejo de la medición de las habilidades humanas, Pearson se encargó de desarrollarlas.

Por la misma época, en Francia, Alfred Binet, influido por los hallazgos de los científicos ingleses, se interesó en investigar también las diferencias individuales. Su interés surgió como consecuencia de sus trabajos con hipnosis, en los que supuso que existían diferencias individuales en las personas con posibilidades de ser hipnotizadas. Alentado por el desarrollo de la psiquiatría en Francia, con el que en un momento dado se encontró asociado, descubrió que la psiquiatría y la psicología, por su propia naturaleza, estaban relacionadas con las diferencias individuales.

Binet, siguiendo en cierto sentido las teorías de Galton, trabajó principalmente con simples variables de discriminación sensorial y con atributos físicos. Se dedicó al estudio de la quiromancia como una forma posible de medir las habilidades.

A principios de este siglo, los estudios de Binet sobre habilidades tomaron un sentido pragmático cuando fue comisionado por el gobierno francés para estudiar el problema de los niños de bajo rendimiento intelectual y, por consiguiente, escolar en las escuelas primarias, ya que era alarmante el número de niños incapaces de ejecutar con éxito las tareas que normalmente exigen los grados escolares. Había que clasificar a los niños en aquéllos capaces de seguir la secuencia que marcaban los programas de estudio y aquéllos que carecían de la capacidad de aprender simultáneamente al ritmo de los demás. Ante este problema práctico, Binet, como cualquier otro psicólogo habría hecho bajo las circunstancias que se le planteaban, trascendió de la investigación de los elementos que constituían las habilidades humanas, a la investigación en sí misma a fin de medir de alguna manera, en conjunto, tales habilidades humanas, es decir, la inteligencia.

Trabajando en colaboración con Theodore Simon, Binet construyó un instrumento que, de acuerdo a las exigencias imperantes hasta entonces, pudiera medir la inteligencia general, considerando que la inteligencia no se expresa en segmentos de conducta, sino más bien como una operación mental combinada, en la cual todo proceso funciona como un total unificado. Desde luego, sus primeros esfuerzos para la elaboración del instrumento estuvieron dirigidos a la psicología de las facultades mentales, prevalentes en su tiempo, pero a partir de una concepción “molar” de la inteligencia. Considerando que era imposible medir todas las habilidades simples elementales en la conducta inteligente, le pareció más factible estudiar los efectos del funcionamiento intelectual. Definió a la inteligencia como “la tendencia a tomar y mantener una dirección; la capacidad de adaptación destinada a obtener la o las metas deseadas y el poder de autocrítica”.

Esta concepción de Binet produjo una de las corrientes psicológicas más trascendentales acerca del concepto de inteligencia y a la cual se adhirieron muchos teóricos, investigadores y elaboradores de pruebas psicológicas de su tiempo. Si consideramos a la inteligencia como *la adaptabilidad mental general a nuevos problemas y nuevas situaciones de la vida*, es decir, *la capacidad para reorganizar nuestros patrones de conducta, a fin de conducirnos con más eficiencia y propiedad ante nuevas situaciones*, entonces la persona más inteligente sería aquella que con más eficiencia y amplitud fuera capaz de modificar su conducta de acuerdo con las demandas del ambiente, porque tendría un mayor número de posibles respuestas y una gran capacidad de reorganización creativa de la conducta. Así, podría abarcar un campo más amplio y controlar un gran número de situaciones, que le permitirían ampliar su área de actividad en mayor grado que aquellas personas que carecieran de estos atributos.

Otra corriente que surgió en torno a la anterior fue aquella que consideraba a la inteligencia como la *habilidad para aprender*. De acuerdo con esto, la inteligencia de una persona está en función del grado en que sea “educable”, es decir, que se encuentre en mejor disposición y capacidad de aprender; por lo que su rango de experiencia y actividad será mayor.

La tercera concepción acerca de la inteligencia es aquella que sostiene que ésta es la *habilidad para manejar el pensamiento abstracto*, es decir, el empleo efectivo de símbolos y conceptos en la solución de problemas, especialmente de aquellos que requieren la utilización de

símbolos numéricos y verbales. A decir verdad, los conceptos de Binet podrían situarse en esta definición de inteligencia.

Ciertamente, si analizamos las definiciones anteriores podremos apreciar que no son mutuamente exclusivas; la diferencia radica en el énfasis que se haga de los elementos que las forman. Es obvio que la capacidad de aprender suministra las bases para el ajuste y adaptación a nuevas o cambiantes condiciones a partir de las experiencias previas que cada quien ha tenido. Desde luego, hay diferencias individuales respecto a la capacidad de aprendizaje y a la habilidad de retener, interpretar, organizar y utilizar lo aprendido. Así, las experiencias previas, para distintas personas, tendrán significado y valor diferentes respecto al aprendizaje. No obstante, la capacidad de aprender, entendida como la adquisición de información y conocimiento, no es criterio suficiente para evaluar la conducta inteligente de una persona.

Los psicólogos están de acuerdo con categorizar y aplicar lo adquirido para interactuar en nuevas situaciones puede considerarse como una conducta “más inteligente”. Por otra parte, los seres humanos aprenden *abstracciones*, principalmente numéricas y verbales, mediante la experiencia con la percepción y el contacto con los objetos, cualidades, eventos o relaciones, para lo cual utilizan los símbolos; por ejemplo, la palabra “perro” adquiere significado para un niño, dentro de una variedad de objetos con los cuales está familiarizado. La palabra “azul” representa una cualidad que él ha percibido como atributo de una variedad de objetos, la palabra “caridad” tiene, para la persona suficientemente desarrollada, cierta connotación, en virtud de haber experimentado eventos denominados de “caridad”. El número siete tiene significado para una persona como resultado de experiencias con objetos concretos; así, esta persona aprende el término no sólo como una posición ordinal, sino también como una adición.

Si el sujeto ha aprendido a relacionarse eficientemente con símbolos (es decir, con abstracciones) debe comprender que la *palabra no es la cosa o cualidad* concreta a la que se refiere, sino que las palabras y los números son abstracciones que representan objetos, eventos, cualidades o relaciones, pero que mediante los símbolos pueden ser manipuladas como si fueran los objetos en sí. Esto constituye el resultado del aprendizaje; por tanto, es necesario establecer que, sin lenguaje y sin números, el aprendizaje se vería seriamente restringido. En consecuencia, la habilidad de manejar el pensamiento abstracto contribuye a la habilidad de las personas para adaptarse a situaciones nuevas o cambiantes, ya que a través de los símbolos nos es factible trabajar, sin gasto inútil de tiempo y esfuerzo acerca de un problema, solamente por ensayo y error. *Nuestro comportamiento está dirigido a ordenar, evaluar y manejar experiencias pasadas, a fin de proyectar nuestro pensamiento hacia el futuro.* En otras palabras, mediante el empleo de símbolos y abstracciones, el hombre incrementa su línea de conducta considerablemente, y es capaz de ajustarse, extender sus horizontes y trascender la situación inmediata concreta y específica.

Por razones prácticas, y volviendo específicamente al problema estudiado por Binet (planteado por el ministerio de educación del gobierno francés, que consistía en construir un instrumento que tuviera capacidad discriminativa suficiente para “detectar” a aquel sujeto hábil para aprender lo que se esperaba y al nivel en que se encontraba colocado en su situación escolar, de aquél otro que presentaba dificultades para los mismos objetivos), se inicia la construcción de un instrumento que había de ser trascendente, independientemente de las modificaciones pertinentes que en cada cultura deberían hacerse y de su actualización, que ha perdurado por su eficacia y, sobre todo, por su validez hasta nuestros días.

La escala de Binet-Simon proporciona una calificación de *edad mental* para cada niño. El hecho de que sólo se obtenga una calificación presupone que la inteligencia es general o que únicamente está considerándose un factor en los elementos de la prueba.

Consideramos necesario insertar aquí una breve descripción de la prueba de Binet y su historia hasta la actualidad, ya que es una escala de inteligencia de amplio uso en México, aunque no ha sido estandarizada todavía. Solamente ha sido adaptada para ciertos grupos en lo que es útil.

PRUEBA DE STANFORD-BINET

Binet sostenía que la inteligencia se expresa no en segmentos de conducta, sino como una operación mental combinada que en todo proceso involucrado opera como un total unificado.

La finalidad de la elaboración de su prueba parte del principio de que podemos identificar diferencias de mentalidad en grados de brillantez o subnormalidad, con diferencias en niveles de desarrollo, representados por la capacidad promedio de niños de varias edades.

La escala original, creada en 1905, comprende treinta pruebas-reactivos en orden creciente de dificultad, sin separarlos en grupos por edades. La “estandarización”, si así se le puede llamar, fue hecha en niños de edad escolar que cursaban la primaria considerados como normales, bajo el criterio de haber pasado satisfactoriamente sus años de estudio, y en una institución para deficientes mentales, tomando diez casos de cada una de las siguientes edades: 3, 5, 7, 9 y 11 años.

Reconociendo Binet los defectos de su primera escala (entre otros, la validez de las normas), su siguiente investigación dio como resultado la revisión de 1908, en la cual los reactivos fueron agrupados en símiles de edad de 3 a 13 años.

EJEMPLO PARA NIÑOS DE TRES AÑOS (cinco pruebas-reactivos):

- a) Señalar nariz, ojos y boca.
- b) Repetir oraciones de seis sílabas.
- c) Repetir dos dígitos.
- d) Enumerar objetos en una fotografía o tarjeta.
- e) Dar un nombre peculiar.

Algunas pruebas estaban constituidas por tres y hasta por ocho pruebas-reactivos, siendo en total 59.

Las diferencias entre la escala de 1905 y la revisión de 1908 fueron:

- a) Estandarización por edades y amplitud de la muestra (203 para 1908).
- b) Empleo del concepto de edad mental, concediendo un año por cada cinco pruebas-reactivos partiendo de una edad basal (se acreditaba al sujeto con el nivel de edad hacia el cual pasaba todas las pruebas).

Muchos países se interesaron en esta prueba, lo cual dio como resultado un gran número de aplicaciones y evaluaciones acompañadas de sugerencias.

Binet tuvo en cuenta las críticas y las sugerencias, entre otras: el que un gran número de reactivos tenían un error de colocación, resultando muy fáciles las del primer grupo (de edad inferior a los tres años) y muy difíciles las de nivel superior, calificando así muy alto al grupo inferior y muy bajo al superior. Entonces, elaboró la revisión de 1911, cuyos cambios principales fueron:

- a) Las pruebas para niños de 11 años de edad se ampliaron para los de 12 años.
- b) Las pruebas para niños de 12 años de edad se aplicaron a los de 15 años.
- c) Se formó un nivel adulto con las pruebas para adolescentes de 13 años de edad.
- d) Se omitieron algunos de los puntos de la revisión de 1908, por considerarlos muy dependientes de aprendizaje escolar o sobre información muy incidental.
- e) Se hicieron revisiones, traducciones, adaptaciones y estandarizaciones de esta prueba.

En 1908, Goddard publicó una traducción de la escala de 1905, para su aplicación en los Estados Unidos, y en 1911 revisó la de 1908, también para su aplicación en los Estados Unidos.

Yerkes, en 1915 y 1923, hizo revisiones formando agrupaciones por subpruebas en una escala de puntos. Horning publicó en 1922 una escala revisada, la cual empleó mucho tiempo y sustituyó a la de 1916 de Stanford. Kuhlmann, en 1912, 1922 y 1939, llevó a cabo exhaustivas revisiones basadas en estandarización, calificación y rango por edades.

Simultáneamente, otros psicólogos trabajaban en la publicación de la revisión de la escala de Stanford-Binet de 1916, bajo la dirección de M. L. Terman, a fin de adaptarla y estandarizarla a los Estados Unidos.

Como una muestra de 2300 casos (1700 normales, 200 deficientes mentales y superiores, y 400 adultos) trabajaron durante mucho tiempo.

La escala de 1916 consta de 90 reactivos que abarcan un grupo de 3 a 14 años. En ésta se añadieron una serie de pruebas para nivel adulto medio y otra para adultos superiores.

De los 90 reactivos, 54 fueron tomados de la escala de 1911, cinco de revisiones anteriores a la de Binet, cuatro de pruebas americanas y 27 fueron nuevos reactivos. Los grupos de 3 a 10 años tienen seis pruebas-reactivos, con una, dos o tres alternativas, dando crédito de dos meses a cada prueba-reactivo, obteniéndose así un año de incremento en edad mental para cada grupo. No se elaboró ninguna prueba para los 11 años, porque se consideró que el desarrollo en este período de un año es muy lento y no permite establecer claras diferencias.

Se elaboraron ocho pruebas para niños de 12 años (no hubo para los de 13 años), acreditando tres meses por cada prueba, abarcando así 24 meses. En la prueba para 14 años se acreditaron cuatro meses por cada reactivo (6), dando como total 24 meses.

El nivel adulto tiene seis reactivos y califica un valor mental medio de 16 años, ya que acredita cinco meses por cada reactivo. De esta manera, quien aprueba todos los reactivos tendrá una edad mental media de 16.5 a nivel adulto.

Fueron los adultos superiores quienes mostraron una edad mental superior a 17, siendo 19.5 el máximo (seis nuevos por cada reactivo). En esta prueba se otorga una edad basal de 3 años y una terminal de 19.5.

En 1937 se hizo una segunda revisión de la escala, formando dos series equivalentes, *I* y *M*, con 129 reactivos. En esta revisión se eliminaron aquellos reactivos que no resultaron adecuados en la de 1916.

La escala se extiende desde los dos años hasta tres niveles de adultos superiores (I, II y III), incrementando así su utilidad. Los inferiores de 5 años y los superiores a 14 fueron estandarizados y calificados cuidadosamente. De los 2 a los 5 años hay grupos de pruebas con un año de intervalo entre cada grupo. Las escalas de 1916 y 1937 son esencialmente verbales; en ellas hay más ejecución y mejor manejo de material no verbal para niveles bajos.

La estandarización fue más cuidadosa y extensa en sujetos blancos nativos únicamente. La muestra constó de 3000 sujetos, distribuida en 11 estados, con cierta distancia entre las áreas territoriales utilizadas. El método de calificación es igual al utilizado en la escala de 1916, teniendo en cuenta la edad mental y el cociente intelectual, y se diferencia únicamente en la edad mental máxima, que es de 22.10 meses, en vez de 19.6 meses.

Los reactivos de las formas *L* y *M* han sido analizados por medio del análisis factorial. McNemar, quien hizo el primero y más completo análisis factorial, concluyó que en cada uno de los diversos niveles de edad las pruebas-reactivos poseen un factor común (G) y que este factor común es el mismo en todos los niveles de edad. El peso del factor común difiere un poco entre los diversos niveles de edad; sin embargo, este factor influye en el promedio de calificaciones (aproximadamente el 40% de las diferencias –varianza-, es decir, aproximadamente el 40% de las diferencias en ejecución entre un grupo de examinados).

Los resultados estadísticos sugieren también la presencia de factores de grupo en las siguientes edades: 2, 2 1/2, 6, 19 y tal vez 7 y 11. Estos son factores secundarios (factores de grupo) que representan de 5% a 11% de las diferencias, mientras que un tercer factor (otro factor de grupo) contribuye en un 4 a 7%.

Los factores de grupo no parecen ser idénticos en todos los niveles de edad, y de ninguna manera están bien definidos desde el punto de vista de los procesos psicológicos implicados en ellos. Tentativamente, McNemar sugiere que algunos de estos factores de grupo, en los diferentes niveles, pueden denominarse “memoria de diseños”, “motor” o “verbal”. Sin embargo, la conclusión más significativa y definida es que un factor (G) es suficiente para ser tomado en cuenta para las interrelaciones de las pruebas-reactivos, con las excepciones anotadas.

La última revisión, hecha en 1968, agrupa a los sujetos en niveles de edad, desde la edad II hasta el adulto superior. Entre II y V siguen intervalos de un año. Los niveles adultos son: adulto promedio y adulto superior I, II y III. Cada nivel consta de seis pruebas, a excepción del adulto promedio, que tiene ocho. En cada nivel, las pruebas presentan una dificultad uniforme. Asimismo, en cada nivel hay una forma alternativa de la prueba, que es equivalente y sirve para utilizarla en aquellos casos en que por alguna razón se anule la prueba original.

El material de la prueba es simple: juguetes para niños pequeños, y tarjetas impresas; los juguetes representan objetos comunes. Las tareas abarcan y exigen una amplia gama de conductas, desde las más sencillas hasta el razonamiento abstracto; coordinación visomanual (por ejemplo, la prueba de ajuste de formas), construcción con cubos (la prueba de dibujos, en la cual el sujeto debe

reproducir figuras geométricas) y emparejamiento de formas geométricas y longitudes comparadas (prueba de discriminación visual).

Las pruebas de observación e identificación de objetos comunes exigen que se dé el nombre a objetos concretos, así como sus partes y usos. Para edades superiores, los dibujos de objetos se emplean con los mismos fines. Asimismo, hay pruebas de completamiento de dibujos, de analogías o diferencias; una prueba de juicio práctico o sentido común, y una prueba de comprensión. Además, las pruebas de memoria se encuentran a través de toda la escala, así como las pruebas de orientación espacial se hallan en niveles separados. Las pruebas numéricas abarcan desde los conceptos más simples hasta los problemas de razonamiento aritmético más complejo. En los niveles superiores son comunes las pruebas de contenido verbal, como: vocabulario, analogías, completación de oraciones, frases desordenadas, definición de términos abstractos e interpretación de proverbios. También algunas pruebas hacen hincapié en la fluidez verbal, como aquellas que exigen nombrar, tan rápidamente como sea posible, palabras no relacionadas, dar rimas o construir frases que contengan tres palabras dadas. Aún cuando algunas pruebas no son predominantemente verbales en su contenido, requieren comprensión e instrucciones verbales bastante complejas.

La escala Stanford-Binet, en conjunto, mide en gran parte las mismas funciones que la prueba de vocabulario.

Aplicación y puntuación

Al igual que la mayoría de las pruebas de inteligencia individuales, la Stanford-Binet exige un examinador con amplia experiencia. La aplicación y la puntuación son muy complicadas para la mayoría de las pruebas. Por tanto, se requiere una gran familiaridad y conocimiento de la escala, para que la ejecución sea completa. Las dudas y búsquedas suelen ser peligrosas para la armonía. Los ligeros e inadvertidos cambios en la dicción pueden alterar la dificultad de los elementos. Otra complicación consiste en el hecho de que las pruebas deben calificarse medida que se aplican, ya que la realización subsecuente del examen depende de la actuación que el sujeto haya tenido en los niveles previamente aplicados.

Al contestar la prueba Stanford-Binet, ningún sujeto debe resolver todos los elementos. A cada sujeto se le prueba solamente en una amplitud de niveles de edad apropiada a su nivel intelectual. La aplicación de las pruebas no suele requerir más de 30 a 40 minutos para los niños más pequeños, y es inferior a hora y media para los sujetos de más edad. El procedimiento típico consiste en aplicar primero pruebas que se encuentren a un nivel ligeramente inferior a la edad mental que se espera tenga el sujeto. De este modo, las primeras pruebas deben ser lo suficientemente fáciles para despertar confianza, pero no tanto que causen hastío y fastidio. Si el sujeto fracasa en alguna prueba de las que se encuentran dentro del nivel de edad que se aplicó en primer lugar, entonces se le aplica la prueba inmediatamente inferior. Se continúa este procedimiento, hasta que se alcance un nivel en el que se ejecuten con éxito todas las pruebas. Este nivel se denomina *edad basal*. La aplicación de pruebas se continúa después en orden creciente, hasta un nivel en el que se fracase en todas las pruebas, el cual se designa como *edad tope*. Al llegar a este nivel, se suspende la aplicación.

La puntuación de los elementos individuales o de las pruebas sigue un sistema de todo o nada. Para cada prueba se especifica en un manual adjunto la ejecución mínima que se requiere para la aprobación.

Los elementos que el individuo apruebe o falle presentarán cierta dispersión entre los niveles de edad adyacentes. Los individuos no aprueban todas las pruebas que abarcan su nivel de edad mental o el inferior al mismo, ni fracasan en todos los niveles superiores. En lugar de ello, las pruebas que se aprueban satisfactoriamente, se extienden a varios niveles de edad, limitados por la edad basal del sujeto en un extremo y por su edad tope en el otro. En la prueba Stanford-Binet la edad mental del sujeto se halla partiendo de su edad basal y añadiendo a ésta los meses correspondientes por cada prueba que aprueba más allá de ese nivel. La edad mental más alta que puede alcanzarse teóricamente con la prueba Stanford-Binet es de 22 años y 10 meses. Desde luego, esta puntuación no constituye una verdadera edad mental, sino una puntuación numérica que indica un grado de superioridad por encima de la ejecución del adulto medio.

Una innovación importante incluida en la prueba Stanford-Binet de 1960 fue la sustitución del *CI* de proporción por el *CI* de desviación, empleado en las formas anteriores. Estos *CI* de desviación son puntuaciones típicas con una media de 100 y una desviación estándar de 16. La ventaja principal de este tipo de *CI* radica en que proporciona puntuaciones comparables en todos los niveles de edad, eliminando las extravagancias de los *CI* de proporción.

Otro cambio introducido en la forma de 1960 surge del reconocimiento de que la mejora en la ejecución de pruebas continúa hasta la edad de 18 años, en vez de cesar a los 16, como se suponía en la revisión de 1937. Para toda persona cuya edad sea de 18 años o más, la *EC* se considera igual a 18 al calcular el *CI*.

En general, la prueba Stanford-Binet tiende a ser más confiable para las edades altas que para las bajas, y más para los *CI* inferiores que para los superiores. En conjunto, los datos indican que la prueba Stanford-Binet es una prueba muy confiable, ya que la mayoría de los coeficientes de confiabilidad que se dan para los distintos niveles de *CI* y edades son superiores a 0.90.

En relación con la validez del contenido de la prueba Stanford-Binet, las tareas que se eligieron desarrollan claramente funciones como la exactitud de observación, el juicio práctico, la memoria para diversas clases de material, la aptitud para seguir direcciones, la capacidad de ver en el espacio y el razonamiento y manejo de conceptos abstractos. Las aptitudes verbales predominaban claramente en las edades mentales superiores. Los conocimientos que se adquieren en la escuela, como la lectura y la aritmética, son indispensables para la actuación satisfactoria en los niveles de edad superiores. En la medida en que todas estas funciones son pertinentes a lo que comúnmente se considera como inteligencia, cabe decir que la escala tiene validez de contenido.

La continuidad en las funciones medidas en las escalas de 1916, 1919 y 1960 quedó asegurada, reteniendo en cada revisión sólo aquellos elementos que correlacionaban satisfactoriamente con la edad mental en la forma precedente. La diferenciación con la edad representa el principal criterio en la selección de los elementos de la prueba Stanford-Binet. De aquí que exista la seguridad de que esta prueba mide aptitudes que aumentan con la edad durante la niñez y la adolescencia, en nuestra cultura. En cada forma la consistencia interna constituye otro criterio para la selección de elementos. Existe una gran homogeneidad funcional en la prueba Stanford-Binet, a pesar de su aparente variedad de contenido, como queda indicado por la correlación media

elemento-escala de 0.66 para la revisión de 1960. El predominio de las funciones verbales en la escala queda demostrado por la correlación superior de los elementos verbales sobre los no verbales respecto a la actuación en la escala entera.

Otros datos relativos a la validez se hallan en una serie de análisis factoriales de los elementos de la prueba Stanford-Binet. Los datos sobre ambos tipos de validez de la prueba Stanford-Binet, la predictiva y la concurrente, se han expuesto principalmente tomando como criterio el rendimiento académico.

Al igual que la mayoría de las pruebas de inteligencia, la Stanford-Binet presenta una alta correlación con la actuación en casi todos los cursos académicos, pero sus correlaciones son más elevadas en los cursos predominantemente verbales, como el inglés y la historia.

La estabilidad de largo alcance de los CI en la prueba Stanford-Binet puede considerarse también como testimonio de validez predictiva.

Valoración

Una de las principales ventajas de la prueba Stanford-Binet se deriva de la cantidad de datos interpretativos y de la amplia experiencia clínica que se ha acumulado con relación a esta prueba. Para muchos médicos, educadores e investigadores que se interesan en la valoración del nivel de aptitud general, el CI de la prueba Stanford-Binet ha llegado a ser sinónimo del concepto de inteligencia.

El amplio uso de esta calificación de niveles de CI, aún cuando es de indudable utilidad para tipificar la interpretación que se haga de la ejecución de la prueba, lleva en sí mismo ciertos peligros. Como toda clasificación de personas, no debe aplicarse rígidamente ni debe utilizarse para excluir otros datos relativos al individuo. No hay, desde luego, claros índices de separación entre el deficiente mental y el retrasado, o entre el superior y el muy superior.

Al interpretar el CI debe tenerse presente que la prueba Stanford-Binet es principalmente una medida de la aptitud escolar, saturada, además, de funciones verbales, especialmente en los niveles superiores. Los individuos con un impedimento de lenguaje, así como aquéllos cuyas aptitudes más destacadas no se encuentran a lo largo de líneas verbales, obtendrán calificaciones relativamente bajas en esta prueba. Asimismo, hay indudablemente cierto número de campos en los que la aptitud escolar y la comprensión verbal no son de primordial importancia. Evidentemente, al aplicar una prueba a situaciones para las que no es apropiada, dicha prueba sólo reducirá su efectividad. A causa de la identificación corriente de la prueba Stanford-Binet con el concepto de inteligencia, ha existido la tendencia a esperar demasiado de esta prueba. La Stanford-Binet es, igualmente, inapropiada para la medición de las aptitudes diferenciales. Del mismo modo, no es una prueba adecuada para los adultos, especialmente dentro de los rangos normal y superior.

Un gran número de médicos consideran a la prueba Stanford-Binet no sólo como una prueba tipificada, sino también como una entrevista clínica. Las mismas características que hacen que esta escala sea difícil de aplicar, crean también las oportunidades de interacción entre el examinador y el sujeto y proporcionan otras pistas al médico experimentado. En mayor proporción que la mayoría de las pruebas, la Stanford-Binet posibilita la observación de los métodos de trabajo del sujeto, su toma de contacto con el problema y otros aspectos cualitativos de la ejecución. El

examinador puede también tener la oportunidad de juzgar ciertas características de la personalidad, como la buena voluntad, la confianza en sí mismo, la confianza social y la atención. Toda observación cualitativa efectuada en el curso de la aplicación de la prueba Stanford-Binet debe, desde luego, reconocerse claramente como tal, y no ha de interpretarse de la misma manera que las puntuaciones objetivas de la prueba. El valor de las observaciones cualitativas depende en gran medida de la habilidad, experiencia y auténtica formación psicológica del examinador, así como de su conocimiento de las fallas y limitaciones de este tipo de observación.

En cuanto a los fines para los que fue elaborado, la prueba constituye indudablemente un instrumento muy satisfactorio. Los procedimientos seguidos en la determinación de la confiabilidad de la prueba son bastante minuciosos y la confiabilidad de la escala es grande. Se han acumulado muchos datos para dar significación al CI de la prueba.

Funciones subyacentes de la prueba de Stanford-Binet

De los dos a los cinco años de edad, mediante los ítemes que constituyen esta parte de la escala, mide (una cosa es la validez de facie, que es el nombre del test o del ítem, y otra la función que mide):

- a)* Percepción y análisis visual.
- b)* Análisis visual y desarrollo motor.
- c)* Percepción visual más principios de formación de conceptos.
- d)* Recuerdo inmediato.
- e)* Desarrollo de la comprensión y del lenguaje.
- f)* Razonamiento con abstracciones y formación de conceptos.

De los seis a los doce años de edad, mide:

- a)* Análisis visual.
- b)* Análisis visual más desarrollo motor.
- c)* Recuerdo inmediato.
- d)* Desarrollo del lenguaje.
- e)* Razonamiento con abstracciones y formación de conceptos.
- f)* Formación del concepto de número y razonamiento con abstracciones.

De los trece años de edad hasta los adultos III superiores, mide:

- a)* Percepción visual y análisis y razonamiento con materiales no verbales.
- b)* Recuerdo inmediato.
- c)* Desarrollo del lenguaje.
- d)* Razonamiento con abstracciones.
- e)* Formación de conceptos y razonamiento con abstracciones.

MODERNAS APROXIMACIONES AL CONCEPTO DE INTELIGENCIA

En los últimos años se han establecido dos nuevas definiciones acerca de lo que entendemos por inteligencia, las cuales combinan y amplían las definiciones ya presentadas.

Los trabajos de Binet han tenido una gran influencia en la psicología aplicada, fundamentalmente en los Estados Unidos, donde han sido utilizados en forma intensiva en escuelas y clínicas psicológicas, en el ejército y en la industria.

Por otra parte, en Inglaterra, donde las teorías de Binet habían dejado honda huella, surge un marcado paralelo al principio de este siglo entre el trabajo de este investigador y el de Charles Spearman, el cual se encuentra resumido en su obra *Las habilidades del hombre*, publicada en 1927.

Spearman, como Binet, se interesó en las habilidades humanas, pero con una aproximación basada en la naturaleza de las relaciones existentes entre las teorías psicológicas y apoyado en modelos matemáticos para el estudio de tales habilidades. Los esfuerzos de Binet para medir la inteligencia se fundaban en la hipótesis de que “existe sólo un factor de inteligencia, o la inteligencia está dominada por un solo factor”. La teoría de Spearman, por otra parte, plantea como hipótesis sujeta a comprobación la presencia de un “factor general”. En uno de sus primeros trabajos, Spearman concluye que las diferencias individuales en todas las pruebas de habilidad pueden considerarse a partir de dos factores: un factor *g* y un factor específico. En el desarrollo de sus investigaciones llegó a la conclusión de que *g* no era suficiente para poder establecer las correlaciones entre todas las pruebas de habilidades. Además, se dio cuenta de la necesidad de participación de factores específicos correlacionados con “factores de grupo”. Como quiera que se les desee llamar, es obvio que debe ser necesario algo más que un mero análisis factorial. En la actualidad, en la que nosotros mismos hemos “fragmentado” las habilidades humanas en múltiples factores, hemos advertido que mucha de la varianza común entre las pruebas de habilidades puede deberse a un factor general.

Entonces, Spearman propuso un método más refinado para medir el factor *g*. Sería necesario, primero, determinar cuáles tareas pueden ser las apropiadas para la medición del factor *g*; y después, establecer el peso apropiado de cada tarea intelectual. Spearman criticó a Binet por no probar primeramente la presencia de un factor general, y determinar después las cargas de los mismos sobre el factor; sin embargo, Spearman estuvo más interesado en desarrollar la investigación básica acerca de las habilidades humanas, que en el desarrollo de medidas prácticas para medir la inteligencia.

El trabajo de Spearman es importante porque, además de desarrollar modelos matemáticos para el estudio del factor *g*, sentó las bases del análisis factorial, al mismo tiempo que estableció una tradición en la investigación de las habilidades; asimismo, estuvo más interesado en la comprensión de las habilidades humanas que en su medición. Centró su interés en las teorías del factor *g* relacionadas con sus bases biológicas; en la influencia de la cultura; en la interacción del factor *g* con las expresiones de habilidades en el comportamiento diario, y en la relación del factor *g* con la fatiga, la velocidad y otras variables.

La teoría desarrollada por Spearman ha sido denominada *bifactorial* porque sostiene la presencia del factor *g* y de factores específicos. El factor *g*, para él, es un tipo de “energía mental”.

La hipótesis de Spearman puede explicarse más ampliamente mediante la afirmación de que las pruebas psicológicas contienen un solo factor común y que todo lo demás se debe a factores específicos para cada prueba; es decir, que parte de la varianza de la prueba puede deberse a un factor común o general, y que lo demás no tiene nada en común. Asimismo, pensó que las pruebas podrían tener alguna particularidad o especificidad relacionada con los reactivos.

La teoría bifactorial de Spearman establece el concepto de inteligencia, diciendo que ésta consiste en la educación de relaciones y la educación de correlatos, es decir, la capacidad de cada individuo de establecer relaciones, desde las más simples hasta las más complejas.

La Prueba de Matrices Progresivas de Raven está construida sobre las bases de la teoría bifactorial de Spearman. Dicha prueba es un test no verbal, diseñado para valorar la capacidad del sujeto, a fin de establecer relaciones y correlatos mediante figuras y diseños geométricos.

En esta prueba las operaciones que se le exigen al examinado son las relativas a la percepción de la estructura de la matriz y de la figura necesaria para completar cada sistema de relaciones presentada. De tal suerte, se trata de medir la habilidad de la persona para discriminar y utilizar lógicamente las relaciones presentadas en un material no verbal. Los problemas requieren, para su solución, operaciones analíticas e integrativas, abstracción de relaciones y, hasta cierto punto, factores de verbalización si el sujeto requiere utilizarlos para facilitarse la tarea.

El análisis factorial ha demostrado que las pruebas de matrices miden con suficiente extensión el factor *g* (factor general) con poca aplicación del factor de percepción espacial. Raven, autor de este test, considera y traduce este factor como la capacidad de un sujeto para educir relaciones y correlatos establecida por Spearman.

Sólo por ejemplificar otro test, construido sobre las mismas bases, y para eliminar la idea de que son pruebas paralelas, mencionaremos la prueba de Anstey (de dominós). El análisis factorial ha demostrado que el grado de correlación existente entre los factores que miden ambas pruebas no es suficientemente alto para considerar equivalentes a dichas pruebas.

Después de múltiples investigaciones, tanto en Inglaterra como en los Estados Unidos, los psicólogos consideraron que un solo factor común no es suficiente para influir en las interrelaciones de las pruebas mentales; sin embargo, estadísticamente se ha demostrado que Spearman no estaba equivocado en cuanto a la importancia, aunque no omnipotencia, del factor general.

Actualmente, las teorías de las habilidades humanas de Spearman sirven de estímulo y pueden mostrar a aquellos factorialistas modernos una buena lección a su constante afán de buscar cada día más factores. Para los psicólogos británicos, los estudios de Spearman son aún preponderantes. Así, aún cuando no aceptan totalmente la teoría del factor *g* como el único factor de inteligencia, sí reconocen la presencia de un factor general, además de factores de grupo.

Los seguidores de Spearman han puesto mayor atención en los factores de grupo que en el factor *g*. A medida que se incrementa el descubrimiento en el número de factores, se plantea el problema de disponerlos en alguna forma lógica de interrelación. En 1949, sir Cyril Burt concibió un tipo jerárquico que puede aplicarse a la totalidad de la mente humana, con la mayor dicotomía entre las características intelectuales o factor *g* y las características prácticas o conductuales. Este

investigador identifica varios niveles de bifurcación: en el primero y más alto nivel existen relaciones, en el segundo asociaciones, en el tercero percepciones, y en el cuarto sensaciones. Sin embargo, en su modelo, Burt tuvo que separar todas aquellas categorías que contenían más de dos factores. Al nivel de asociaciones, reconoció que había divisiones en lo que se ha denominado *memoria* con un factor de retentividad general, donde fundamentalmente existen factores de grupo como subcategorías, que podrían ser factores visuales, auditivos, cinestésicos, de memoria verbal, etc., y asociación productiva con un factor general de “inventiva”, donde quedan supeditados factores de fluidez y originalidad. Otra asociación general puede considerar factores de habilidad verbal, habilidad de lenguaje y habilidad aritmética, bajo cada una de las cuales se encuentran agrupados dos o tres subfactores.

Antes del modelo de Burt, debe mencionarse la teoría de Vernon: jerárquicamente, bajo el factor *g* se fundamentan dos grandes factores, el educacional-verbal y el *K-M*, que implica aspectos pragmáticos, como en el modelo de Burt. Este modelo se subdivide en tres tipos: habilidad espacial, habilidad manual e información mecánica.

En términos generales, los estudios experimentales y sus consecuencias se han pronunciado contra el factor *g*, lo cual constituye una clave conceptual de los modelos jerárquicos de Burt y Vernon. Uno de los argumentos para sustentar este punto de vista es que lo que busca y exige el factor *g* no es una variable invariante, sino un cambio efectuado con gran facilidad que dependa de la batería de pruebas psicológicas que se estén analizando; es decir, de que el primer factor (en el análisis factorial) sea considerado como factor *g*. Al cambiar la batería, cambia una rotación a partir de esta dimensión para localizar el factor *g*, se dificulta la observación de la invarianza de un análisis a otro.

Tanto las teorías de Spearman en Inglaterra como las de Binet en Francia fueron rápidamente aplicadas en los Estados Unidos; sin embargo, debe hacerse notar que los psicólogos norteamericanos mostraron mayor interés por el análisis factorial de Spearman. De 1920 a 1930, estos psicólogos estudiaron exhaustivamente los métodos matemáticos de análisis factorial. Destacaron en esta actividad personalidades como T. L. Kelly y K. J. Kalzinger. Posteriormente, L. L. Thurstone se concentró en el desarrollo matemático del análisis factorial, a fin de explorar las habilidades humanas. Por esta razón, al igual que en Inglaterra, a Spearman se le considera el “padre del análisis factorial”; a Thurstone se le ha asignado en los Estados Unidos el título de “padre de la escuela factorialista americana”, caracterizada, cuando menos hasta hace poco, por un empirismo bien fundamentado que, aún cuando aparentemente no ha tenido teorías suficientemente aproximadas, sí ha tenido purismo técnico en la medición de las habilidades humanas y en el análisis matemático de sus correlaciones.

Thurstone, antes de ser psicólogo, fue ingeniero. Durante algún tiempo trabajó con Thomas Edison. Su capacidad inventiva recorrió los campos del aprendizaje, la percepción, la personalidad y otros tópicos. Su laboratorio de la Universidad de Chicago fue el centro mundial del estudio de las habilidades humanas, y aún existen muchos estudiosos egresados de esa universidad que se dedican a esta área de la psicología. Le interesó profundamente el problema de las diferencias individuales, y su intención fue la de abordar problemas empíricos.

En 1930, aún se discutía si Spearman estaba en lo cierto por lo que respecta a su enfoque acerca de la estructura de la inteligencia. Thurstone rechazó la afirmación de Spearman en este

sentido, e intentó contestar a la pregunta: ¿cuántos y qué tipos de factores presentan correlación en los tests de inteligencia? Para dar una respuesta satisfactoria llevó a cabo minuciosos estudios, a fin de mejorar los métodos de análisis factorial existentes. Para esos fines específicos, empleó más de sesenta pruebas psicológicas desarrolladas. Sólo aquellos que hayan participado en la elaboración de instrumentos de medición psicológica pueden comprender la trascendencia de la amplitud de posibilidades que Thurstone debió considerar para ejecutar tales magnitudes, y el conocimiento y comprensión del problema que esto implica. Por otra parte, cuando realizó sus investigaciones no tuvo a su disposición un equipo suficiente para el procesamiento de datos, es decir, para computar el material obtenido. Esta fue una de las razones principales por las que utilizó el método centroide y de rotación de factores.

No analizaremos aquí detenidamente las múltiples investigaciones de Thurstone, sino que sólo expondremos sus resultados últimos, ya que han sido fundamentales para la comprensión del problema acerca de la inteligencia y, específicamente, de su concepción de la estructura de la misma. Así, una vez que se aplicaron los instrumentos elaborados para la investigación, administrados adecuadamente, tabulados y llevados a los procedimientos estadísticos específicos, se obtuvo una tabla de interrelaciones factorializada, en el cual aparecieron diez factores mediante el método centroide. Al revisar la matriz factorial rotada aparecieron los siguientes factores: *a)* comprensión verbal, *b)* fluidez verbal, *c)* factor numérico, *d)* factor espacial, *e)* velocidad perceptual, *f)* memoria (memoria de asociación y memoria de significado de relaciones) y *g)* factores de razonamiento general (deducción o diseño de conclusiones, y educación de relaciones –es decir, observar las relaciones entre dos objetos, ideas o eventos, a fin de utilizarlas para encontrar otras ideas, objetos u eventos -). En términos generales, el factor de razonamiento general puede considerarse como la capacidad o habilidad de encontrar soluciones a problemas, razonamiento de problemas aritméticos, etc.

Al analizar las sesenta pruebas elaboradas por Thurstone, se puede observar el mismo grupo de factores primarios encontrados en otros estudios previos de tipo factorial. De esta manera, seis factores han mostrado estabilidad en un lapso de edad, aún cuando algunos de ellos no han podido ser clarificados.

Los factores que han podido mostrar su presencia con alto grado de estabilidad temporal son:

<i>a)</i> Comprensión verbal.	Factor V
<i>b)</i> Fluidez verbal.	Factor W
<i>c)</i> Factor numérico	Factor N
<i>d)</i> Factor espacial	Factor E
<i>e)</i> Memoria de asociación	Factor M
<i>f)</i> Inducción o razonamiento	Factor R

El factor perceptual y el factor deductivo no fueron suficientemente determinados en los estudios de Thurstone. El factor V, comprensión verbal, se refiere a la comprensión de la palabra hablada: pruebas de vocabulario, sinónimos, antónimos, completamiento de frases, comprensión de la cultura, etc. El factor W, fluidez verbal, hace referencia a la habilidad para expresarse mediante un repertorio adquirido y que puede ser probado a través de discursos preparados o espontáneos, anagramas con rima, etc.

El factor numérico N, se refiere a la habilidad para ejecutar cálculos numéricos con exactitud y rapidez, y parece tener restricciones que no se relacionan con la solución de problemas, sino que son de naturaleza más simple.

El factor espacial E, representa el manejo de las dimensiones por parte del sujeto, mismas que pueden ser dos o tres, contenidas en un sinnúmero de habilidades mecánicas.

El factor M, memoria de asociación, ha aparecido sistemáticamente a través de toda la batería y depende de la habilidad que tenga el sujeto para memorizar rápidamente. Este factor probablemente puede ser desglosado en más factores específicos.

El factor R, inducción o razonamiento, está contenido en tareas que exigen al sujeto que establezca una regla o principio en una situación dada. Parece ser que en todos los estudios se presentan dos factores de razonamiento independiente: inducción y deducción. La mayoría de las pruebas elaboradas hasta ahora parecen estar saturadas de factores intuitivos que, por el momento, denominamos razonamiento.

Después de los estudios de Thurstone, se ha impulsado y desarrollado ampliamente el estudio e investigación de las habilidades humanas. Hacia 1950, French (1951) informó que existían más de cuarenta factores de habilidades bien establecidos.

El número de factores componentes del intelecto ha seguido en proceso de investigación, y se puede decir que actualmente existen entre cincuenta y cien factores, dependiendo de la cantidad de la acuciosidad de los análisis.

En la búsqueda gradual de factores parece que se pueden forzar artificialmente en diversos tipos. Puede hacerse forzando las pruebas o forzando a los sujetos, o bien, calificando de diversas formas la misma prueba. En el primer caso, se aplica tres veces la prueba en momentos diferentes, pidiendo a los sujetos que, por ejemplo, encierren en un círculo todas las aes que encuentren, concediéndoseles un tiempo de 60 segundos para el primer intento; para el segundo intento, que encierren igualmente las bes y para el tercero la ces. Las tres pruebas estarán bien correlacionadas para relacionarlas con otras pruebas, mediante sus factores. En el segundo caso, se interrelacionan factores, que son heterogéneos en diversos aspectos, e interrelacionan variables que se relacionan con esta heterogeneidad; por ejemplo, puede registrarse un factor de edad que adquiere su mayor carga en la edad cronológica, y su menor carga en pruebas que cambian con la edad; otro ejemplo puede ser el registro de un factor sexual, que relaciona las diferencias en la tendencia central entre hombres y mujeres, en pruebas de habilidades. No obstante, es posible prevenir que esto ocurra si se agrupan los sujetos por la variable que se intenta medir: sexo, edad y otras, antes de aplicar el análisis factorial.

Independientemente de que muchos de los factores que supuestamente se han encontrado pueden ser “artefactores” (por explicar de alguna manera el que en un momento dado puedan ser artificios), para poder explicar algunas operaciones intelectuales complejas, es menester determinar claramente la importancia que puede tener cada uno de ellos. Así, un ejemplo sería aquel en el que un factor se encuentra actuando en una cantidad considerable de pruebas que miden las habilidades humanas mediante las correlaciones entre todas las pruebas. Por tanto, un factor será importante en la medida en que se refiera a un factor general y a un grupo de factores. Otro ejemplo sería el del

factor de comprensión verbal, el cual está relacionado casi siempre con todo tipo de prueba que directa o indirectamente se relacione con la comprensión o entendimiento de las palabras, de las cláusulas, del discurso, etc. Dado que ejerce gran influencia en múltiples tareas intelectuales, es difícil reducir sus efectos en pruebas construidas para la medición de otros factores. Asimismo, podríamos afirmar que es difícil delimitar la “generalidad” de un factor: si lo colocamos en el extremo de la generalidad, diremos que el factor no es suficientemente analítico y tiende a ocultar subdimensiones de las habilidades en un área específica.

Otro aspecto importante de este tipo de factores radica en la predicción de variables de criterio significativo; por ejemplo, en el éxito que pueda haber en el entrenamiento de un sujeto para realizar alguna tarea específica.

Flishman y Hempel (1954) llevaron a cabo una serie de estudios destinados a encontrar las relaciones existentes entre los factores de habilidad y progreso en diferentes pruebas de aprendizaje; así, concluyeron que la importancia de los factores dependía del estadio en que se encontrara el aprendizaje.

Guilford (1961) ha enfatizado las relaciones que existen entre los factores de habilidad humana y los procesos, los cuales han sido investigados en experimentos controlados, como formación de conceptos, aprendizaje, memoria, motivación, etc. Este investigador ha hecho importantes aportaciones respecto a los factores del intelecto y ha contribuido notablemente al desarrollo del análisis factorial; sin embargo, lo fundamental de su trabajo ha sido su aproximación sistemática acerca de la naturaleza de los factores del intelecto.

El conocimiento que se tiene acerca de la inteligencia del hombre data desde aproximadamente 25 años a la fecha. Como hemos visto, uno de los investigadores más notables en este campo ha sido Thurstone, junto con sus colaboradores.

El descubrimiento de los componentes de la inteligencia se ha llevado a cabo a través de experimentos con el método del análisis factorial. Nunnally indica que cada uno de los componentes o factores intelectuales constituye una habilidad única y necesaria para “hacer algo bien”, cuando un sujeto tiene que ejecutar alguna tarea (o prueba).

Guilford, utilizando ampliamente el análisis factorial, ha desarrollado una teoría acerca de estos problemas a la que ha denominado *estructura del intelecto*. Así, aún cuando Guilford considera que cada factor es perfectamente distinto de otro para ser apreciado mediante el análisis factorial, en los últimos años se ha considerado necesario clasificarlos, dado que son similares en algunos aspectos. El criterio de clasificación empleado por Guilford ha sido en función del tipo de proceso o de operación ejecutada. Así, este investigador obtuvo cinco grupos de habilidades intelectuales: factores de cognición, de memoria, de pensamiento convergente, de pensamiento divergente y de factores de evaluación.

Los factores de cognición se refieren al conocimiento, descubrimiento o redescubrimiento de lo conocido. La memoria representa la retención de lo conocido. El pensamiento convergente y el divergente, en sus operaciones, originan dos tipos de información: el conocimiento de información y el recuerdo de información. En las operaciones de pensamiento divergente encontramos direcciones diferentes, algunas veces en forma molecular y en otras molar. En el pensamiento

convergente, la información se transmite de una respuesta correcta o exacta a una reconocida o convencional. En la evaluación, se califica mediante decisiones como “correcto”, “bueno”, “conveniente” o “adecuado” para el conocimiento, para aquello que recordamos o producimos en el pensamiento productivo.

Desde otro punto de vista, a los factores intelectuales los clasifica Guilford de acuerdo con la clase de material o contenido que utilizan. Así hasta ahora se han encontrado tres clases de materiales: figuras, símbolos y un contenido de tipo semántico. El contenido de figuras es concreto y se percibe por los sentidos en forma directa; por ejemplo, el material de tipo visual tiene propiedades como el tamaño, la forma, el color, la localización o la textura. El material simbólico se forma con signos convencionales que representan objetos, situaciones o dígitos que, organizados en sistemas generales, forman sistemas numéricos, alfabetos u otra clase de estructuras representativas de objetos concretos. El contenido semántico se refiere a significados o a ideas verbales.

Estas tres clases de materiales pueden ser representadas por un modelo sólido (en forma de cubo) que simboliza el “intelecto humano” o “estructura del intelecto” en donde se pueden observar las dimensiones, y a través de cada una de ellas operan diversos elementos: en la primera, distintos tipos de operaciones; en la segunda, productos de tipo variado, y en la tercera, diversas clases de contenido. Además se ha agregado una cuarta categoría, a la cual se le ha denominado “contenido conductual” para representar en cierto modo la llamada “inteligencia social”.

La cognición es básica para cualquier tipo de operaciones intelectuales, por lo que, en orden de clasificación de las dimensiones, aparece primero. Así, sin cognición no hay memoria, y sin ésta no hay producción. Asimismo, sin cognición ni producción, evidentemente, no hay evaluación. El diseño del cubo-modelo está hecho visto de frente, con base en el incremento de la dependencia de un tipo de operación a otro. A continuación mencionaremos los elementos de dicho diseño.

Las unidades de los productos se consideran básicas aún cuando aparezcan hacia el fondo del cubo. Estas unidades se dividen en clases, relaciones, sistemas, transformaciones e implicaciones, y van de lo particular (clases) a lo general (implicaciones). El concepto de “producto” es la forma en que ocurre todo tipo de información; podría llamarse también “concepción”, para referirnos a formas de conocimiento o comprensión. Las informaciones pueden entenderse en forma de unidades. Las unidades son cosas a las que normalmente les aplicamos nombres. Las clases, como tipo de producto de información, se acercan al significado más común del término; así, una clase es un conjunto de objetos que participan de una o más características comunes. Una relación es una forma de conexión entre dos cosas. Los sistemas son patrones u organizaciones complejas de interdependencia o interacción de partes. Las transformaciones son cambios, revisiones, redefiniciones o modificaciones mediante las cuales un producto de información que se encuentra en una situación dada pasa a otra diferente. Finalmente, una implicación es algo esperado, anticipado o predicho a partir de una información. (Véase la “estructura del intelecto”, figura 3).

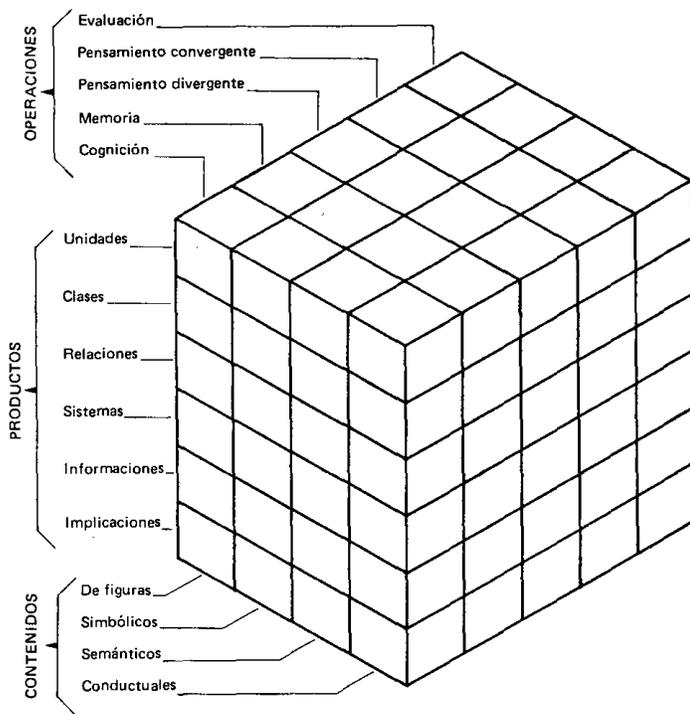


Figura 3. Estructura del intelecto, de J. P. Guilford. (Tomado del libro *The Nature of Human Intelligence.*)

Para una mejor comprensión del modelo, se describirá resumidamente paso a paso, pero de la manera más simple y clara posible. Así, cada celdilla del modelo corresponde a cierta clase de habilidad, que se puede describir mediante operaciones, contenidos y productos; es decir, cada celdilla es la intersección de una combinación de tipos de operaciones, contenidos y productos. Un test destinado a probar esta habilidad debe tener las tres propiedades.

Para analizar el modelo debemos tomar un estrato vertical, empezando por la frente. La primera capa muestra 18 matrices (eliminando la columna destinada a “contenido conductual”) de celdilla, cada una de las cuales contiene una habilidad cognoscitiva.

Hasta donde se sabe, sólo se ajustan 15 de las 18 celdillas de las habilidades cognoscitivas. Cada fila representa una tríada de habilidades similares, que tienen en común una clase especial de productos. Los factores de la primera fila están en relación con el conocimiento de unidades; los de la segunda se hallan en relación con las clases, y así sucesivamente. El hecho de que aparezcan tres factores en una celdilla indica que en la columna de figuras puede haber, cuando menos, más de una habilidad.

Por lo que respecta a la habilidad para conocer unidades simbólicas, para conocer habilidades semánticas (factor de comprensión verbal), para probar la habilidad, para conocer clases de unidades, etc., es necesario elaborar una serie de ítemes que contengan operaciones que deben ejecutarse de acuerdo a la concepción de los elementos que conforman la estructura del intelecto.

Durante algún tiempo, la habilidad para comprender un sistema semántico ha sido conocida como el factor de razonamiento general. Si este factor se coloca en una celdilla de la estructura del

intelecto, aparece un nuevo concepto de su naturaleza. Esta podría ser una habilidad para captar todo tipo de sistemas que son considerados como conceptos verbales.

Las transformaciones son cambios de tipo variable, que comprenden ordenamiento, organización o significado. En la columna de figuras para las hileras de transformaciones encontramos un factor de visualización.

En el conjunto de habilidades que se relacionan con la cognición de implicaciones, podemos notar que el individuo va más allá de la información dada, pero no en la extensión de llegar a conclusiones; más bien, podríamos decir que extrapola. De tal suerte, a partir de la información dada, espera ciertas consecuencias. Por otra parte, los dos factores que se encuentran en la hilera de la matriz de cognición se denominan “factores de prevención”, en relación con el material concreto de las figuras, con ideas o con eventos.

Respecto al área de las habilidades de la memoria, debemos decir que ésta es la menos explorada; en ella, sólo siete de las celdillas de la matriz de la memoria han demostrado ciertos factores conocidos, los cuales se reducen a unidades, relaciones y sistemas. La primera celdilla está “ocupada” por dos factores paralelos a la matriz de cognición: memoria visual y memoria auditiva. La diferencia entre estas dos habilidades implica que una persona puede ser capaz de señalar, por ejemplo, un objeto que vio sobre una página, pero incapaz de especificar cuál, después de haber pasado algunas de ellas. Teniendo en cuenta la hilera en la matriz de la memoria, podremos encontrar habilidades para recordar clases, transformaciones e implicaciones, así como unidades, relaciones y sistemas.

Por lo que se refiere a las habilidades de pensamiento divergente, el único aspecto de pensamiento divergente consiste en que produce una gran variedad de respuestas, pero no determinadas por la información dada; es decir, el pensamiento divergente no desempeña un papel importante en el proceso total de obtención de una conclusión, sino más bien su función abarca un tipo de pensamiento de ensayo y error.

La habilidad del buen conocimiento de la fluidez de la palabra puede ser probada mediante el examen de listas de palabras que satisfagan un requerimiento de letras específico, por ejemplo, palabras que empiecen con “s” o que terminen en “ción”. Esta habilidad se conoce con el nombre de *unidades simbólicas* (como una facilitación divergente). La habilidad semántica paralela se designa “fluidez ideacional”; por tanto, se piensa que este tipo de fluidez puede ser el único aspecto de su factor, llamado “flexibilidad espontánea”.

Un estudio actual poco conocido, pero que revela habilidades divergentes, contiene pruebas acerca de si hay habilidades simbólicas y de figuras que también producen múltiples clases de habilidades.

La única habilidad que contiene relaciones se denomina “fluidez de asociación”, debido que relaciona un aspecto determinado con una variada producción de ellos.

Otro factor importante es el llamado “fluidez de expresión”, o sea, la rápida formación de frases u oraciones que prueban la existencia del factor.

Asimismo, en la hilera de la producción de la matriz que se refiere a las transformaciones encontramos factores relevantes como el que se conoce con el nombre de “flexibilidad adaptativa”, que puede localizarse también en la columna de figuras.

El factor llamado “originalidad” se reconoce ahora como flexibilidad adaptativa con material semántico, en el cual hay un cambio de significados.

Las habilidades de producción convergente pueden verse en las tres columnas de las 18 habilidades, 12 de las cuales se reconocen ahora. En la primera hilera, que pertenece a las unidades, aparece una habilidad para designar propiedades de contenido figural (formas o colores) y una habilidad para designar abstracciones (clases, relaciones, etc.). Es probable que la rapidez para designar colores o formas no está colocada en su sitio en la matriz de pensamiento convergente. Tal vez lo esperado en la producción del pensamiento convergente de unidades de contenido de figuras debería ser en forma de figuras, en vez de palabras. La producción convergente, en sus relaciones, está representada por tres factores conocidos que implican “la educación de correlatos”, como Spearman los designaba. La información dada incluye una unidad y una relación establecida. Las pruebas de analogías para complemento, más que para elección de respuestas alternativas, pone de relieve este tipo de habilidad.

Sólo conocemos un factor de producción convergente de sistemas y éste se encuentra en la columna de contenido semántico. Dicho factor puede medirse por medio de pruebas llamadas “de ordenamiento”.

En el tipo de transformaciones producidas por una variedad, podemos localizar tres factores, reconocidos como habilidades de redefinición, implicando en cada caso la redefinición del cambio de funciones o del empleo de partes de una unidad y dándoles nuevas funciones o empleos en alguna nueva unidad. Puede ser probado esto por medio de tareas concretas y simbólicas. Para el factor de redefinición semántica puede aplicarse la prueba de transformación gestáltica.

La producción convergente de implicaciones significa el diseño de conclusiones totalmente determinadas, a partir de una información dada. El factor bien conocido de facilidad numérica se encuentra a través de la columna de simbolismos. La habilidad paralela en la columna de figuras, contiene la prueba de razonamiento de figuras, que define operacionalmente la manera en que utilizan las figuras. En la columna semántica encontraremos probablemente el factor llamado “deducción”.

El área evaluativo aparece al final de todas las categorías operacionales. Dentro de la matriz de evaluación se han reconocido solamente ocho habilidades, pero cinco hileras, al menos han tenido uno o más factores, así como tres de las columnas correspondientes a las categorías de contenido. La evaluación considera en cada caso la búsqueda de decisiones, como seguridad, beneficio, conveniencia o funcionalidad de la información.

En la primera hilera de la evaluación de unidades, la decisión importante debe hacerse en función de la identidad de una unidad. En la columna de simbolismos se localiza una habilidad de identificación de juicios de unidades simbólicas, en forma de series de letras, de series de números o de nombres de individuos.

Ninguna de las habilidades evaluativas pertenece a la clase que ya ha sido reconocida. Las habilidades que tienen relación con evaluaciones que implican relaciones, deben tener un criterio de consistencia lógica, y la evaluación de sistemas tiene ingerencia con la consistencia interna de aquellos sistemas en la medida en que partimos del conocimiento de cada factor. El factor ha sido denominado “evaluación experimental”.

La habilidad semántica para la evaluación de transformaciones es lo que por un tiempo se ha llamado “juicio” o “enjuiciamiento”. Consiste en la capacidad de los sujetos para responder, de cinco soluciones probables, aquella que resuelva el problema de la manera más fácil y práctica. La solución implica, en general, improvisaciones o adaptaciones de objetos familiares a usos no frecuentes.

Por otra parte, existe un factor que se conoce como “sensibilidad a los problemas”. Este factor se ha distinguido por considerársele como una habilidad evaluativa que se relaciona con implicaciones.

En un principio, los tests de ejecución fueron diseñados como suplementos o, a veces, como sustitutos de la escala Stanford-Binet, a fin de examinar niños sordos, analfabetos y personas que no hablaran inglés. Asimismo, estos tests fueron elaborados para aplicarlos a niños en los cuales se sospechaba que tenían dificultades para la lectura o que expresaban conductas que denotaban déficit de habilidades intelectuales, producidas por influencia ambiental o por problemas inherentes de tipo biológico.

Un punto de vista que ha provocado una gran división entre los teóricos de este campo es el de si los aspectos cognoscitivos y no cognoscitivos de las diferencias individuales podrían separarse. Todo mundo puede estar de acuerdo en que ambos aspectos interactúan en la vida diaria del sujeto; sin embargo, históricamente se sabe que existe una distinción entre dos características de la naturaleza humana: el razonamiento y la emoción, la habilidad y la motivación, etcétera. Muchos psicólogos han estado de acuerdo con que tales divisiones no son más que artificios llevados a la práctica y que quienes evalúan la inteligencia tienen que admitir plenamente que la motivación y la personalidad son aspectos no eliminables de la medición mental.

El aprendizaje de habilidades, particularmente en las escuelas, ha conducido a la consideración de que si la función de esas instituciones es la estimular el logro del desarrollo intelectual, obliga a diferenciar los “deseos y las actitudes” como partes componentes de la inteligencia o como elementos totalmente independientes de ella.

David Wechsler, quien ha elaborado instrumentos para niños y adultos, a fin de poder evaluar el nivel intelectual, durante mucho tiempo ha insistido en que la inteligencia y otros tipos de características (a las cuales se les conoce con diversas denominaciones, por ejemplo el mal llamado “temperamento”) no pueden estar separadas.

Wechsler, trabajando en el Hospital Psiquiátrico de Bellevue, Nueva York, desarrolló su primer instrumento de medición intelectual, el cual difería en muchos aspectos de los elaborados hasta entonces, particularmente del de Binet. Este instrumento, denominado *escala Wechsler-Bellevue*, cuyos fundamentos lógicos están sustentados en el concepto de “inteligencia general” (un tanto erróneo) y en la validez racional, más que en la empírica, fue publicado en 1939.

Wechsler, como todos sus antecesores, inició su tarea intentando definir el concepto de inteligencia de la siguiente manera: la inteligencia es el *agregado o capacidad global del individuo para actuar propositivamente, para pensar racionalmente y para conducirse adecuada y eficientemente con su ambiente*. Este investigador, apoyándose en los estudios de Spearman, pensaba que existía un factor fundamental en las funciones intelectuales. Le denomina a la inteligencia agregado o capacidad global porque está compuesta de elementos o habilidades que, aún cuando no son totalmente independientes, pueden ser cualitativamente diferenciables; por tanto, a través de la medición de estas habilidades evaluamos, en última instancia, la inteligencia, entendiendo que la inteligencia no es la mera suma de habilidades.

Wechsler considera tres importantes razones para fundamentar lo anterior: a) que los productos finales de la inteligencia no están sólo en función del número de habilidades, sino también de la forma en que se combinan para operar, es decir, lo que él llama “su configuración”; b) que existen, en cierto sentido, otros factores independientes de las habilidades intelectuales, que influyen en la conducta intelectual, como son los incentivos y los impulsos o motivaciones, intenciones, etcétera, y c) el hecho de que en ciertas tareas se exija una mayor “cantidad” de habilidad intelectual específica no implica un “aumento” en la efectividad o eficiencia de la conducta intelectual como un total. Como quiera que sea, la única forma obtenible de evaluación de la inteligencia es mediante la medición de diversos aspectos de esas habilidades. No existe contradicción alguna entre los términos “inteligencia general” y “habilidad intelectual”.

Thorndike fue el primero en desarrollar la idea de que la medición de la inteligencia consistía en la evaluación cuantitativa de los productos mentales en función del número, la eficiencia y la velocidad con que se realizaran. Afirmaba que las habilidades son producciones mentales ordenadas en diferentes clases o tipos de operaciones. Así, por ejemplo, sugirió que la inteligencia debía agruparse en tres categorías: a) inteligencia abstracta o verbal, mediante el manejo fundamental de símbolos; b) inteligencia práctica, por medio de la manipulación de objetos, y c) inteligencia social, facilitadora de las confrontaciones con los seres humanos (relaciones humanas adecuadas y positivas).

Como se recordará, el profesor Spearman, en su teoría bifactorial, afirma que la inteligencia contiene algunas habilidades o factores independientes uno de otro, y sus investigaciones se dirigen a determinar si los resultados de los tests reafirman el punto de vista de que la inteligencia “práctica” y la inteligencia “verbal” son capacidades distintas e independientes o si (como él sostiene) no son capacidades distintas e independientes y representan esencialmente un mismo aspecto, difiriendo sólo en que el tipo de factores operantes son no intelectivos o factores específicos.

De esta pequeña disertación, puede ahora comprenderse por qué Wechsler elaboró sus pruebas en la forma conocida. Después de algunas investigaciones previas, eligió doce tests y los aplicó a aproximadamente 1000 sujetos. El rango de edades osciló entre 10 y 60 años, aunque las normas finales incluyeron desde los 7.5 años de edad. Esta escala de inteligencia fue diseñada con el propósito de que en una sola “forma” del instrumento pudiera hacerse la evaluación de un amplio rango de edades. En 1955 aparece una forma revisada de esta escala, conocida como WAIS (escala de inteligencia Wechsler para adultos), la cual no contiene cambios significativos en cuanto a su contenido, construcción, organización y calificación. Lo significativo fue su ampliación en cuanto a las muestras que originalmente se habían utilizado para su estandarización, algunas modificaciones prácticas en cuanto a las instrucciones y calificación y, sobre todo, una revisión minuciosa del

contenido, para situaciones especiales (por ejemplo, en la evaluación de sujetos con un bajo nivel de eficiencia mental). Asimismo, se incrementó la dificultad progresiva entre cada reactivo; se eliminaron unos y se aumentaron otros, ya sea por su baja validez o por su ambigüedad, etc.

En 1949, Wechsler había planteado la necesidad de elaborar una escala destinada especialmente a la evaluación de la inteligencia en los niños de cinco a quince años de edad. Sobre las mismas bases y estructura de los anteriores, construyó la escala de inteligencia Wechsler para niños, WISC, y posteriormente la WPPSI, escala de inteligencia Wechsler para preescolares y para niños de primer año de enseñanza primaria, que cubre un rango de edades que va de los cuatro a los seis años y medio.

CLASIFICACIÓN DE LA INTELIGENCIA

Para explicar y poder entender el por qué de la clasificación de la inteligencia, podríamos hacer un parangón con las leyes físicas que controlan y predisponen al individuo a discriminar los colores del arco iris. La inteligencia genera un continuo semejante al espectro del arco iris. La conducta humana se expresa en diversas formas, es decir, presenta patrones específicos que dependen de las exigencias del ambiente. La denominada conducta inteligente tiene también expresiones que plantean patrones de comportamiento, mismas que en cierto momento los psicólogos deben determinar y localizar en ese continuo que hemos denominado inteligencia.

La contribución de la psicología moderna en este campo ha sido, por lo menos, dar nuevas configuraciones a la conducta intelectual y precisar conceptos de ella mediante la introducción de métodos cuantitativos. Las ventajas teóricas de la clasificación cuantitativa son obvias, aún cuando en la práctica no son reales, debido a que una clasificación cuantitativa depende no sólo de la confiabilidad de los datos empleados, sino también de la validez de la interpretación asignada a ellos. Los datos cuantitativos se basan generalmente en medidas derivadas de diversas pruebas de inteligencia, y la clasificación dependerá del tipo de unidad psicométrica que se utilice (percentiles, CI, calificaciones T, etcétera) a partir de las cuales la significación podrá ser diferente.

Por ejemplo, en la clasificación de Terman, los CI Stern obtenidos inferiores a 70 clasifican a las personas como deficientes mentales; aquellos que obtienen entre 80 y 90, como subnormales; aquellos que obtienen de 90 a 110, como normales o promedio; etcétera.

Kuhlmann considera que el límite inferior para las categorías arriba descritas es: inferiores a 75, deficientes normales; de 75 a 84, limítrofes o borderlines (también se les denomina *fronterizos*); de 85 a 94, subnormales; de 95 a 104, normales; etcétera.

Otros autores han utilizado distintos criterios para la clasificación de la inteligencia. Sin embargo, lo importante es saber por qué utilizan tal amplitud de intervalo.

Wechsler considera necesaria una redefinición de las categorías básicas de la inteligencia, en función de criterios estadísticos explícitos. Así, propone que para cada nivel de inteligencia se tenga un intervalo de clase que marque un rango de CI que se encuentre a una distancia medida a partir de la media y que se expresará en términos de la desviación estándar. De este modo, un deficiente mental es una persona cuyo CI se halla a tres o más desviaciones abajo de la media, lo que significaría, en términos de rangos percentilares, que se encontraría alrededor del 0.2% de la parte

inferior de la población total. Una persona con inteligencia fronteriza o limítrofe, en términos de CI, se hallaría entre las desviaciones 3 y 2 a partir de la media; etcétera.

La elección de los puntos límites es relativamente arbitraria. En el caso de los deficientes mentales, existen varias formas de evaluación. Estas estimaciones tienen en consideración la amplitud de variación entre ellas mismas, las cuales se aproximan bastante al 2.5% de la población total. Para definir a los grupos deficientes mentales, es necesario y razonable hacerlo a partir de aquellos individuos que obtienen un CI cuya distancia es de -3 o más, a partir de la media.

El esquema de clasificación de Wechsler es simétrico y comprende muchas clases, tanto hacia arriba como hacia debajo de la media. Wechsler elaboró una clasificación final de la inteligencia, en la cual incluye porcentajes para cada categoría dada. La tabla siguiente expresa la forma en que han sido clasificados, de acuerdo con este investigador, los diversos niveles de inteligencia, los cuales se refieren a las escalas de WAIS y de WISC.

Bajo este criterio, Wechsler encuentra que el concepto estadístico de la inteligencia, para la clasificación mental, es necesario en la ciencia psicológica. El concepto y sus implicaciones lógicas son difíciles de aceptar para algunos psicólogos, debido a que pueden conducir a decisiones imprácticas y absurdas; sin embargo, estas decisiones no provienen del concepto en sí mismo, sino de la *no comprensión de tal concepto*.

Tabla 1. Clasificación de la inteligencia, de David Wechsler

CI	CLASIFICACIÓN
De 130 o más	Muy superior
De 120 – 129	Superior
De 110 – 119	Normal brillante
De 90 – 109	Normal
De 80 – 89	Subnormal
De 70 – 79	Limítrofe (borderline o fronterizo)
De 69 – 50	Deficiente mental superficial
De 49 – 30	Deficiente mental medio
De 29 o menos	Deficiente mental profundo

Realmente la definición estadística de inteligencia, obtenida sobre cualquier muestra particular o específica, es válida sólo para los grupos que representen la población muestreada. No limita la medida de los grupos de referencia, para los que pueden ser útiles las normas; éstos pueden ser tan grandes como la representatividad de la muestra lo propicie; pero las restringe a partir del tipo de individuos que queden incluidos y para los que se tengan otros propósitos de clasificación.

Para Wechsler, la gran ventaja que representa la utilización del CI como una base fundamental de clasificación es que indica clara y constantemente la relatividad de las medidas. A veces, en la práctica, es necesario utilizar los resultados de las pruebas como si representaran cantidades absolutas, por ejemplo, cuando aplicamos pruebas de aptitudes como medidas de

eficiencia mental. En una prueba de aptitudes podemos dar puntuaciones utilizando un mínimo como estándar para calcular los índices de eficiencia. Lo mismo podemos hacer con una prueba de inteligencia. Podemos decir, por ejemplo, que para ser un buen maestro o buen mecánico, un sujeto deberá tener determinada cantidad mínima de inteligencia en una prueba dada. Si el CI utilizado como medida de inteligencia del sujeto es tan claro que permita percibir el papel de una calificación mínima en el caso de la prueba de aptitudes, y si este denominador es constante, podrá considerarse que, debido a sus propiedades, constituye una medida absoluta. Esta forma de aplicación del CI puede ser aceptada siempre y cuando se tenga en consideración que se ha transformado en un CI de eficiencia.

Al hablar de las diferencias entre ambos cocientes, nos referimos a la medida del cociente de inteligencia de una persona en relación con el cociente de inteligencia de los individuos de su propio grupo, sobre la base de que se trata de un grupo homogéneo. Cuando estudiamos el cociente de eficiencia, no nos interesa la edad de la persona ni otros factores que influyen en la obtención del CI, sino sólo en las habilidades de esa persona, comparadas con la ejecución estándar del resto del grupo.

El CI no es la única calificación, ni la mejor, para clasificar esa entidad tan compleja y definirla con un simple número. Esta construcción es función de otros factores (además de la habilidad intelectual) como: impulsos, balances emocionales, persistencia, etcétera, que no son fácilmente apreciables, ni mucho menos mensurables, en situaciones concretas.

Un problema poco estudiado cuando nos referimos a la inteligencia es el de la deficiencia mental o retraso mental. Este, como muchos conceptos en psicología, sufre modificaciones con el transcurso del tiempo. Así, presenta dos problemas fundamentales: a) la definición abstracta del término, y b) su aplicación práctica.

Se han dado muchas definiciones psicológicas a partir de la interrogante ¿quiénes son los deficientes mentales? Se han diseñado métodos específicos para diferenciarlos. El problema de la definición es de orden moral y médico-legal; así, hay definiciones abstractas de deficiencia mental, aspectos cuantitativos y estadísticos y métodos prácticos de su medición. De acuerdo con esto, un deficiente mental es aquel que, debido a un desarrollo intelectual restringido, es incapaz de manejarse por sí mismo y de resolver sus propios problemas, constituyéndose no sólo en una carga familiar sino también social al tener que ser ayudado a sobrevivir. Podríamos agregar que el deficiente mental es una persona cuya dotación básica, intelectual, es insuficiente para asimilar y retener las experiencias necesarias que le brinda el ambiente. Esta situación hace que pueda producir una calificación mínima o un registro de edad mental o CI en ciertas pruebas de inteligencia estandarizadas.

En seguida haremos una breve descripción de las escalas de Wechsler a fin de ejemplificar una de las formas mejor construidas para medir la inteligencia (o el rendimiento intelectual). Estas escalas nos interesan particularmente, así como cada uno de los tests con que hemos tratado de ejemplificar cuando ha sido necesario, porque no sólo se han venido empleando en México desde hace mucho tiempo, sino que en la actualidad ya han sido estandarizadas, cuando menos para sujetos del Distrito Federal, es decir, de la ciudad de México. En algunos lugares de la República Mexicana también ya se han hecho estandarizaciones para grupos. Sin embargo, aún falta mucho por hacer para poder llegar a establecer estas escalas, con el fin de emplearlas confiablemente en todo el país.

Mucho de este trabajo ha sido ejecutado por el Centro de Ciencias del Comportamiento, antes CICC, ahora INCCAPAC (Instituto Nacional de Ciencias del Comportamiento y de la Actitud Pública, A. C.), en donde desde hace cerca de quince años se empezó a trabajar tanto en la formación de investigadores en el campo de la psicología educativa, de la psicología social y de la psicología evolutiva como en la estandarización de instrumentos de medición en esos campos de la psicología.

CONSTITUCIÓN DE LAS ESCALAS DE WECHSLER

En general, las escalas de Wechsler están basadas en la teoría bifactorial de Spearman en el sentido de que, para Wechsler, existe un factor fundamental en las funciones intelectuales; no obstante, al igual que Binet, Wechsler analiza la idea de la “inteligencia general”. Empero, debe aclararse que el factor *g* de inteligencia, que para Spearman era un factor primario, para Wechsler es un factor *g* secundario, o sea, hace la función de unificador para la expresión unitaria de una conducta que requiere la participación de varias operaciones e, incluso, de una serie de habilidades específicas que al combinarse dan un producto intelectual.

Las escalas de Wechsler son escalas compuestas, verbales y de ejecución, constituidas por una serie de tests específicos que, al pasar a formar parte de una escala, se denominan “subtests”. Son, por una parte, escalas heterogéneas, intertest, debido a que miden diferentes funciones y, por lo tanto, distintos factores en cada uno de ellos; y, por otra parte, son escalas homogéneas intratest, porque cada subtest mide un solo factor en toda su amplitud.

Los subtests están estructurados en la siguiente forma: los reactivos se clasifican en tres grupos; el primero corresponde a aquellos reactivos que toda persona puede contestar, tenga o no instrucción escolar; el segundo abarca reactivos que se espera sean contestados por sujetos que han adquirido instrucción escolar media, y el tercero se refiere a aquellos reactivos que sólo pueden ser contestados por sujetos con cierta especialización (instrucción especial).

En cuanto a los subtests, éstos se clasifican en varios grupos, atendiendo a los factores específicos que evalúan cada uno de los subtests. Así, tenemos que: el primer grupo es el denominado *estrictamente verbal*, al cual corresponden los subtests de información, comprensión, semejanzas y vocabulario, debido a que son verbales tanto las operaciones necesarias para resolver las tareas implicadas como las respuestas.

El segundo grupo se denomina *de atención y concentración*, al cual pertenecen los subtests de aritmética y retención de dígitos. En la escala de ejecución también encontramos dos grupos de subtests: los visuales y los visomotrices. A los visuales pertenecen los subtests de completamiento de imágenes y ordenamiento de imágenes, y los visomotrices abarcan los de ensamble de objetos, diseño con cubos, símbolo de dígitos, y laberintos en la escala de niños.

Desde el punto de vista de la vulnerabilidad de cada uno de los subtests (sensibilidad), podemos decir que el menos sensible de ellos, y por tanto el que constituye un parámetro de comparación de la ejecución de los otros subtests, es el vocabulario; y los más sensibles son la retención de dígitos (escala verbal) y el completamiento de figuras (escala de ejecución); de éstos el primero es más vulnerable que el segundo.

Desde el punto de vista de la validez de las escalas de Wechsler, la escala ha sido determinada sobre un amplio rango de edades y sujeta a una serie de análisis a fin de establecerla. Se han efectuado interrelaciones entre subtests, para determinar la presencia o ausencia del factor *g*. Las respuestas más significativas son las encontradas entre calificaciones en cada subtest cuando se correlacionan con calificaciones del resto de la escala.

Se ha establecido la correlación entre la escala y la escolaridad o calificación de logro escolar, y correlaciones con juicios de los maestros, quienes un tanto subjetivamente evaluaron el logro escolar de sus alumnos. El incremento y decremento de las calificaciones en la escala, en función de la edad y la constancia de medias y desviaciones estándar de los cocientes intelectuales en diversas edades han sido otros de los criterios de la determinación de la validez de la escala.

Un criterio significativo de validez concurrente ha sido el establecido con la prueba de Stanford-Binet, mediante la cual se ha obtenido una correlación de 0.82, la cual es muy alta e indica que ambas pruebas están midiendo lo mismo en un alto grado. Por otra parte, esto ha demostrado validez predictiva, toda vez que, utilizando el juicio empírico como árbitro de decisión, ha sido capaz de justificar las afirmaciones hechas mediante la prueba acerca de la conducta futura de un sujeto o de un grupo.

Respecto a la confiabilidad, las correlaciones test-retest (destinadas a la obtención de coeficientes de estabilidad temporal y los errores estándar de medida para todos los subtests y para los tres tipos de coeficiente intelectual) han dado resultados satisfactorios, ya que se obtuvieron coeficientes de confiabilidad desde 0.60 hasta 0.80. El CI total tiene una confiabilidad comparable a la del Stanford-Binet, o sea, aproximadamente de 0.90. Como suele suceder, los subtests tienen menos confiabilidad que la escala total.

Wechsler, en su instrumento de medición de la inteligencia, realizó todo tipo de análisis estadísticos a fin de asegurarse de que su utilización tenía la validez y la confiabilidad adecuada para emplearla con absoluta seguridad en los resultados que arrojará. Así, este instrumento de medición es aplicable en estudios de confiabilidad temporal, de consistencia interna, calificación por jueces, validez predictiva, validez concurrente, validez de contenido y análisis factoriales (como el método centroide de Thurstone de extracción de factores), análisis ortogonales y oblicuos, la modificación de Burt del procesamiento ortogonal bifactorial, y el de Cohen, quien utilizó la solución centroide completa. Se han realizado, hasta ahora, por lo menos más de seis análisis factoriales, algunos de ellos con las muestras originales y otros con muestras nuevas pero equivalentes y, por tanto, comparables.

DESCRIPCIÓN DE LA ESCUELA DE INTELIGENCIA WECHSLER

En la elaboración de su primera escala de inteligencia, que será la que aquí tomemos como ejemplo, ya que las otras no varían en los fundamentos teóricos, ni en los formales y ni siquiera en sus presentaciones publicadas, Wechsler llevó a la práctica los siguientes procedimientos:

1. Realizó un análisis minucioso de los diversos tests de inteligencia que se encontraban ya en uso, específicamente en lo relativo al tipo de funciones que medían las características de las muestras sobre las cuales habían sido estandarizadas, y la realidad de la confiabilidad de las pruebas.

2. Intentó establecer la validez de cada una de ellas con base en: a) otros tests ya probados, y b) registros empíricos de inteligencia. Estos tests contenían estimaciones y registros de los maestros, ejecutivos, oficiales del ejército, etcétera.

3. Intentó hacer un registro de su propia experiencia clínica, así como de los otros investigadores.

4. Dedicó dos años a preparar el trabajo experimental preliminar de entrenamiento sobre grupos de un nivel de inteligencia conocido.

Sobre estas bases, seleccionó doce pruebas, de las cuales quedaron finalmente once, ya que eliminó la de análisis de cubos. La escala final contiene seis subtests verbales y cinco de ejecución para adultos. Dichos tests, en el orden siguiente, se denominan:

Escala verbal

- a) Subtest de información.
- b) Subtest de comprensión general.
- c) Subtest de aritmética.
- d) Subtest de semejanzas.
- e) Subtest de retención de dígitos.
- f) Subtest de vocabulario

Escala de ejecución

- Subtest de completamiento de figuras.
- Subtest de diseño con cubos.
- Subtest de ensamble de objetos.
- Subtest ordenamiento de figuras.
- Subtest de símbolos de dígitos.

Es necesario indicar aquí que para la elaboración de la escala para niños consideraron doce subtests; es decir, se agregó uno en la escala de ejecución, como suplementario, que consiste en la prueba de laberintos. En la escala verbal se consideró como subtest suplementario la prueba de retención de dígitos. Tener pruebas suplementarias en la escala para niños tiene por objeto equilibrar la calificación global del sujeto, cuando por variables no controladas, en alguno de los subtests baja el rendimiento.

Separar en dos escalas, verbal y de ejecución, el trabajo intelectual del sujeto no significa que sólo ciertas habilidades estén implicadas en el test, sino más bien implica las diferentes formas en que se expresa la inteligencia. Los subtests constituyen medidas diferentes de la inteligencia y no tipos distintos de inteligencia; por otra parte, los dos tipos de escala significan solamente la forma en que pueden ser agrupados los tests de acuerdo a su naturaleza.

DESCRIPCIÓN DE LOS SUBTESTS DE LA ESCALA DE WECHSLER

Escala verbal

Subtest de información

Uno de los problemas que se presentan en la utilización del subtest de información, que forma parte de la escala de Wechsler, fue el hecho de que durante mucho tiempo este tipo de pruebas fue utilizado por los psiquiatras como el mejor “detector” del nivel intelectual de sus pacientes. Los psicólogos hicieron durante muchos años lo contrario. Este tipo de pruebas sólo fueron tomadas en cuenta cuando, haciendo uso de ellas para probar las habilidades de grupo, demostraron su presencia en las pruebas de inteligencia estandarizadas. Además presentan facilidades para la obtención de datos; por ejemplo, permiten obtener fácilmente calificaciones; muestran, asimismo, que la cantidad de conocimiento que posee una persona depende de sus oportunidades educativas y culturales, fundamentalmente.

El principal soporte del grado de información de una persona, considerado como una buena medida de inteligencia, fue obtenido de un instrumento denominado Army Alfa. El valor de una prueba de información depende de la validez de los ítemes incluidos en ella. No existen principios universales que “legislen” si las preguntas son buenas o no. En general, los ítemes deben ser considerados como la media del tipo de conocimiento que un promedio de individuos con oportunidades similares pueden ser capaces de adquirir.

La prueba de información de Wechsler consta de 29 ítemes seleccionados de una larga lista, cuyos resultados fueron obtenidos de su aplicación a grupos de individuos de nivel intelectual conocido, y a partir de los aciertos o errores entre varios grupos. El orden en que se presentan es el que indicó como necesario la muestra sobre la cual se hizo la estandarización. Es obvio que el orden de presentación de los ítemes dependerá de los grupos a los que se les aplique, analizados desde un punto de vista cultural, educacional y de niveles socioeconómicos. Lo importante es que quienes utilicen los tests tengan siempre en cuenta esta situación, tan fundamental para la obtención de datos válidos.

Siendo la prueba de información uno de los subtests más completos de la batería, declina con la edad y presenta altas correlaciones con otros tests de la batería (como el de retención de dígitos –memoria rutinaria -) o con pruebas visomotoras (como la del ordenamiento de figuras o diseño con cubos).

Las suposiciones de que las preguntas de información cubren un amplio rango de material cuya muestra proporciona la información de lo que una persona debió aprender dentro de las oportunidades otorgadas por el grupo social al que pertenece (la de que el rango de información de un individuo constituye un índice de su capacidad intelectual, y la de que a mayor inteligencia, mayor amplitud de intereses, mayor “curiosidad” y una búsqueda mayor de estimulación) sólo tienen validez si el sujeto a quien se ha aplicado la prueba ha tenido las mismas oportunidades de experiencia y aprendizaje, y si los ítemes de la prueba son una muestra representativa de las oportunidades de adquisición de la información.

Esta prueba, además, como casi todas, está sujeta a factores importantes en su efectividad de desempeño, como la motivación, la estabilidad emocional (actual y cotidiana) que influye directamente no sólo en este test, sino en cualquier tarea experimental o que por exigencias del ambiente se le plantee al individuo en la vida diaria.

Las principales funciones que contiene la prueba de información se presentan a partir de la composición factorial de cada una de ellas; en este caso, en primer lugar una carga considerable del factor memoria, particularmente en grupos de adultos. Esto varía, como es lógico, de individuo a individuo. En forma más analítica, podríamos decir que fundamentan a la función “gruesa” llamada memoria, operaciones de asociación y organización de la experiencia y un amplio rango de retención del material aprendido. Como cabe suponer por el tipo de material, los factores externos influyentes son ambientales, fundamentalmente.

Subtest de comprensión general

Este subtest depende, en individuos normales, del factor de comprensión verbal y del factor *g*. Desde el punto de vista factorial, un factor importante en el desempeño de esta prueba es el llamado “estereotipo social”, así como el de “sentido común”. Contiene catorce situaciones-problema, en las que el sujeto debe comprender cuál es el problema contenido en la situación y debe dar la o las respuestas adecuadas, dentro de lo que se espera en un grupo “normal” de comportamiento. El éxito en esta prueba depende de la obtención de información práctica, aunada a la habilidad para evaluar y utilizar la experiencia pasada. Además, se ha demostrado que el factor verbal contribuye considerablemente al buen desempeño de estas tareas.

De tal suerte, notamos que la prueba de comprensión general consta de funciones u operaciones de razonamiento abstracto, organización del conocimiento y capacidad de utilizar en la práctica las experiencias y aprendizajes obtenidos. Desde luego, esta prueba depende ampliamente de oportunidades culturales y de la capacidad para colocarse en las situaciones reales.

Subtest de aritmética

Sorpresivamente, este subtest consta de una carga factorial muy alta del factor memoria, aunque, como cabe esperar, presenta una gran saturación del factor *g*. De acuerdo con algunos análisis factoriales (por ejemplo, el de Davis), parecen encontrarse cargas significativas no sólo en el razonamiento general, sino también en factores de fluidez numérica y de conocimiento mecánico e información. El factor numérico se advierte en todos los niveles de edad en esta prueba. Los problemas que se le presentan al sujeto son aquellos que exigen concentración y atención específica. En este tipo de subtests se requieren operaciones del razonamiento con abstracciones, formación de conceptos y retención de procesos aritméticos; además, dicho subtest se ve grandemente influido por atención inmediata y automática, así como por la oportunidad de adquirir los procesos aritméticos fundamentales.

Subtest de semejanzas

La carga factorial de este subtest se ubica en el factor *g*, fundamentalmente a través de todos los niveles de edad. No obstante, es de vital importancia la carga del factor verbal, lo que hace que esta prueba conduzca, primeramente, a la medición de habilidades de generalización y abstracción.

Por esta razón, con frecuencia se ha pensado que evalúa la formación de conceptos. Asimismo, se ha demostrado que la ejecución de esta prueba no depende exclusivamente de la capacidad verbal, como podría suponerse por su contenido de “facie”. Esto se corrobora con las altas correlaciones que presenta esta prueba en el razonamiento abstracto, en el cual el lenguaje hablado no desempeña una función importante.

La prueba de semejanzas consta de trece reactivos apareados, es decir, trece pares de palabras en las cuales se le pide al sujeto que dé una respuesta en relación con la razón del por qué están o deben estar juntas. En este punto, el sujeto se ve obligado a ejecutar operaciones de análisis de relaciones y de formación de conceptos verbales, para lo cual influyen las oportunidades culturales que le proporcione el ambiente.

Subtest de retención de dígitos

Desde el punto de vista operacional, este subtest consiste de la repetición de números como única habilidad o habilidad específica para su ejecución. Además, se encuentra fuertemente influido por el factor memoria y con baja correlación en el factor *g*.

Parece ser que la habilidad para atender y concentrarse es fundamental para realizar correctamente esta prueba. Wechsler afirma que la habilidad o inhabilidad para repetir o no los dígitos depende de lo que él ha llamado “factores no intelectivos de la inteligencia”. Tal vez la posibilidad o libertad de poder distraerse y no responder correctamente, constituye uno de esos factores. Tanto esta prueba como la de símbolos de dígitos tienen en común una gran influencia del factor memoria, y una función de la memoria general o neutral en sus modalidades. Dicha memoria general es trascendental en las funciones intelectuales.

Las principales funciones que se advierten en esta prueba son: el recuerdo inmediato, la imaginación auditiva y la imaginación visual simultánea; y como factores que influyen notablemente en la ejecución de las tareas que implican esta prueba, se encuentran la atención inmediata y la atención dirigida.

Esta prueba consta de series de dígitos que varían en longitud, desde tres hasta nueve dígitos (en orden directo) y hasta ocho dígitos (en orden inverso). La evidencia experimental ha demostrado que existen bajas correlaciones con otros tests de inteligencia. Este subtest se utiliza, además, para detectar deficiencia mental o defectos orgánicos. Una baja memoria de dígitos, especialmente en orden inverso, se encuentra frecuentemente en personas incapaces de mantener la atención necesaria cuando las tareas intelectuales son de difícil solución. Cuando se obtienen bajas calificaciones en este subtest y no están asociadas con problemas de deficiencia u organicidad, ello puede deberse a la presencia de cierto grado de ansiedad, que impide automáticamente las operaciones de atención inmediata.

Subtest de vocabulario

Este tipo de subtest no sólo señala el índice de la escolaridad de una persona, sino que constituye una buena y confiable medida de inteligencia, debido a que el número de palabras que una persona conoce constituye una medida tanto de su habilidad para aprender, como de su

información verbal y de la amplitud general de sus ideas. Desde luego, este tipo de actividad está influida en gran cantidad por las oportunidades culturales y educacionales de las personas.

Esta prueba consta de cuarenta palabras seleccionadas de un diccionario sobre una base de “grado de dificultad” y elegidas al azar. Desde el punto de vista cualitativo, una prueba que exige como tarea la definición de palabras resulta de mucho valor. Generalmente, cada persona define de manera diferente las palabras; así mientras una define *burro* como un animal, otra puede definirlo como “tiene cuatro patas”, “sirve para cargar bultos”, etcétera. La definición es de gran importancia porque muestra el nivel cultural y educacional de las personas. La amplitud de vocabulario de un sujeto permite la evaluación cuantitativa, es decir, denota cuántas palabras conoce, no importando la elegancia o “pulcritud” del lenguaje, sino el reconocimiento del significado de la palabra.

ESCALA DE EJECUCIÓN

Subtest de completamiento de figuras

Este subtest consta de veintiún reactivos consistentes en tarjetas que muestran una figura a la cual le falta una parte importante, que el sujeto debe encontrar y dar su nombre; por ejemplo, una cara a la cual le falta la nariz. Factorialmente, en este subtest aparece una especial situación: presenta una carga sobre factores no verbales en la mayoría de los análisis efectuados, y muestra poca influencia del factor *g*, la más baja de todos los subtests de ejecución; sin embargo, surge un nuevo y único factor que parece importante en el comportamiento intelectual y que no ha podido ser identificado con precisión. Intervienen en él factores como la percepción visual (análisis) y la imaginación visual.

Dadas las operaciones a ejecutar en esta tarea que impone el test, puede, obviamente, considerarse que influye de manera notable la experiencia ambiental y la agudeza visual. Asimismo, mide ostensiblemente habilidades perceptuales y conceptuales implicadas en el reconocimiento e identificación de objetos y formas conocidos. El sujeto, para ser capaz de darse cuenta de qué le falta a una figura, debe saber primero qué es lo que ella representa. Debemos hacer notar que la habilidad de un individuo para apreciar qué le falta a una figura faltante sea importante para completarla, ya sea en su función o en su forma.

Subtest de diseño con cubos

Este subtest consta de diez diseños diferentes, los cuales, al igual que los demás subtests de esta escala, son progresivos en grado de dificultad. Cada diseño está constituido por nueve cubos de colores (en la prueba para adultos, los cubos son rojos y blancos, y en la de niños son amarillos, azules, rojos y blancos); el sujeto debe emplear sólo cuatro de ellos para construir o reproducir los ejemplos que le son presentados en la primera parte de la prueba. Cuando el examinador lo indique, debe emplear los nueve.

Por otra parte, la carga factorial indica un factor de organización no verbal, y tiene una gran influencia del factor *g*. Según Wechsler, este tipo de prueba no es menos restrictivo que el resto de los subtests de la escala. Asimismo, Davis, en sus análisis factoriales, encuentra que este subtest muestra una carga mayor no sólo en visualización, sino también en factores de conocimiento e información mecánica. Desde el punto de vista individual, estos factores no son muy importantes,

en relación a la varianza total de la prueba; pero, desde el punto de vista colectivo, adquieren cierta trascendencia. Las principales funciones en este subtest son: percepción de forma (análisis) e integración visomotora, las cuales constituyen factores que influyen en la ejecución motora y en un mínimo de capacidad discriminativa de color.

Subtest de ordenamiento de figuras.

Este subtest consta de ocho series de figuras. Cada serie se le presenta al sujeto desordenadamente, para que la ordene en una secuencia correcta y forme una historia. De este modo, se intenta medir la comprensión y la evaluación de una situación por parte del sujeto, sin la intervención del lenguaje hablado. Factorialmente, presenta una carga de factores no verbales, aunque menos significativamente que el subtest de completamiento de figuras o ensamble de objetos. Muestra, también, una carga importante del factor memoria, y parece tener relaciones con factores específicos no intelectivos.

Otro aspecto fundamental en este subtest lo constituye la carga del factor E en ciertos grupos de edad (45-54 años). Dicho factor, al cual Wechsler lo designa como “relevancia” y que debemos interpretarlo como “aplicabilidad” cuando no está presente, indica la falla de un sujeto para obtener una secuencia correcta, aunque él crea que la obtuvo. Las principales funciones implicadas en la solución de los problemas son: percepción visual de relaciones, y síntesis de material no verbal, las cuales están influidas grandemente por la agudeza visual y por un mínimo de oportunidades culturales.

Subtest de ensamble de objetos

Este subtest está constituido por cuatro figuras, las cuales le son presentadas al sujeto en forma de rompecabezas, es decir, cortadas de tal manera que puedan ser acomodadas para obtener la forma que represente un objeto familiar al examinado. Al sujeto se le pide que perciba y reconstruya las partes del objeto y lo complete.

Al respecto, se ha encontrado que dicho subtest correlaciona poco con el resto de los subtests de la escala y produce poca diferenciación entre los individuos, pero ayuda al examinador a comprender los estilos de percepción, el grado de confiabilidad que se puede tener en los métodos de ensayo y error, y la manera en que se reacciona a los errores. Asimismo, este subtest expresa un factor analítico: con excepción del factor g, casi toda la varianza está en función de factores de organización no verbales: percepción visual (síntesis), integración visual y precisión de la actividad motora.

Subtest de símbolos de dígitos

Este subtest consta de un rectángulo dividido en nueve cuadros o celdillas dobles. En las celdillas de la parte superior se encuentran los números del uno al nueve, y en las de la parte inferior (las cuales corresponden a cada uno de los números) se halla un símbolo. (A esta prueba Wechsler la denominó “test de sustitución”, el cual es uno de los más antiguos utilizados en los tests psicológicos). En seguida, se encuentran cuatro rectángulos divididos en cien celdillas dobles, en las cuales están distribuidos al azar los números en la parte superior y se hallan vacíos los cuadros de la parte inferior.

En este subtest, la tarea que se le asigna al sujeto consiste en asociar los números con los símbolos que aparecen en la parte superior. En realidad, lo que se le pide es que asocie símbolos con símbolos en un tiempo dado, con lo cual determinará la medida de habilidad intelectual en esta área. Las principales funciones en esta prueba se hacen más relevantes en los adultos, en quienes la agudeza visual y la coordinación motora desempeñan un papel importante para determinar sus niveles de habilidad. Desde luego, las funciones fundamentales que implica esta prueba son: el recuerdo rutinario inmediato, la integración visomotora y la imaginación visual, mismas que se habían visto influidas considerablemente por el nivel de actividad motora del sujeto.

Factorialmente, a partir del factor *g*, este subtest muestra una fuerte carga factorial en memoria y organización no verbal. Asimismo, la velocidad perceptual puede considerarse como organización verbal, a diferencia de la organización numérica, misma que se considera como un factor separado, lo cual hace que este subtest sea factorialmente ambiguo y su exclusión sea frecuentemente discutida; sin embargo, existe una interacción muy interesante entre la facilidad numérica y la velocidad perceptual.

Tabla 2
Clasificación de las funciones subyacentes y los factores que influyen en la ejecución de la escala de Wechsler

SUBTEST	FUNCIONES	FACTORES INFLUYENTES
<i>Información</i>	a) Capacidad de retención (memoria)	a) Ambiente cultural. b) Intereses
<i>Comprensión</i>	b) Asociación y organización de experiencias. a) Razonamiento con abstracciones (análisis y síntesis, símbolos, números y lenguaje).	a) Oportunidades culturales. b) Respuestas a situaciones reales.
<i>Aritmética</i>	a) Razonamiento con abstractos. b) Formación de conceptos. c) Retención (procesos aritméticos)	a) Atención inmediata. b) Oportunidades para adquirir los procesos aritméticos fundamentales.
<i>Similitudes</i>	a) Análisis de relaciones. b) Formación de conceptos verbales (calidad).	a) Mínimo de oportunidades culturales.
<i>Vocabulario</i>	a) Desarrollo del lenguaje.	b) Oportunidades culturales (M. A. facilitador).
<i>Retención de dígitos.</i>	a) Recuerdo inmediato. b) Imaginación auditiva. c) Imágenes visuales (simultáneas)	a) Atención inmediata.

INFORMACIÓN ADICIONAL ACERCA DE LAS ESCALAS DE WECHSLER

Como indicamos anteriormente, Wechsler elaboró primero la escala Wechsler-Bellevue, la cual era aplicada en un amplio rango de edades que iban desde los diez hasta los sesenta años, aún cuando las normas resultaban útiles para sujetos de 7.5 años de edad en adelante. Esta escala fue probada por su autor y estandarizada en la población para la cual debía aplicarse, a partir de variables fundamentales, como distribución ocupacional de los adultos en los Estados Unidos, nivel educacional (desde analfabetos hasta graduados), niveles de edad (desde niños hasta adultos), etcétera.

Desde el punto de vista de los estudios de validez, se ejecutaron interrelaciones de los subtests, correlaciones con la escolaridad, correlaciones con el juicio de maestros respecto de los alumnos, incremento y decremento de las calificaciones en función de la edad y correlaciones con la escala Stanford-Binet. Asimismo, se proporcionó el rango y distribución de los cocientes intelectuales, y se aplicó un criterio pragmático de la eficiencia de la escala en cuanto al pronóstico.

Tabla 3
Escala de Ejecución

SUBTEST	FACTORES INFLUYENTES	FUNCIONES
<i>Completamiento de figuras</i>	<ul style="list-style-type: none"> a) Percepción visual: análisis b) Imágenes visuales. 	<ul style="list-style-type: none"> a) Experiencia ambiental. b) Exactitud visual (en el momento).
<i>Diseño con bloques</i>	<ul style="list-style-type: none"> a) Percepción de forma. b) Percepción visual: análisis y síntesis. c) Integración visomotora 	<ul style="list-style-type: none"> a) Nivel de actividad motora. b) Mínimo de visión del color.
<i>Ordenamiento de figuras</i>	<ul style="list-style-type: none"> a) Percepción visual de relaciones (insight visual). b) Síntesis de material no verbal. c) Planeación y anticipación. 	<ul style="list-style-type: none"> a) Mínimo de oportunidades culturales. b) Exactitud visual.
<i>Ensamble de objetos</i>	<ul style="list-style-type: none"> a) Percepción visual: síntesis. b) Integración visomotora. c) Anticipación. 	<ul style="list-style-type: none"> a) Velocidad de actividad motora. b) Precisión de actividad motora.
<i>Símbolos de dígitos</i>	<ul style="list-style-type: none"> a) Recuerdo inmediato. b) Integración visomotora. c) Imaginación visual. d) Capacidad de reproducción e imitación. 	<ul style="list-style-type: none"> a) Velocidad de actividad motora.

Respecto a la confiabilidad, se obtuvieron coeficientes de estabilidad temporal, tanto como subtests como con la escala completa, además del índice de confiabilidad absoluta mediante la obtención del error estándar de medición. Estos estudios revelaron que la prueba tenía un alto coeficiente de confiabilidad.

Sin embargo, en 1955, al hacerse una revisión de la escala de Wechsler-Bellevue, se encontró que aún cuando desde 1949 ya se había dividido en dos escalas (la de niños y la de adultos) esta última, que era casi igual que la original de Bellevue, tenía serias objeciones, entre ellas: problemas acerca del contenido de algunos reactivos, la longitud de la población, y en la administración y calificación del instrumento. Así pues, luego de realizar las modificaciones pertinentes, se construyó la escala que se conoce con el nombre de *escala de inteligencia Wechsler para adultos* (WAIS). La estandarización de esta escala fue hecha en 1700 sujetos (850 mujeres y 850 hombres) tomados de las cuatro áreas geográficas en que se dividió a los Estados Unidos. El rango de edades fluctuó entre 16 y 64 años, clasificados en siete subgrupos que contenían proporcionalmente el número de sujetos de cada edad según el censo de 1950 en los Estados Unidos. La definición de la muestra se hizo en términos de raza, ocupación, zonas urbana y rural, y nivel educacional.

Como en el caso de la escala Wechsler-Bellevue, en esta escala se hicieron principios que las demás escalas para adultos (escalas verbal y de ejecución, CI parciales, y CI total como unidad de calificación). La diferencia que existe entre la escala de Wechsler para niños respecto de las demás escalas, es que esta consta de doce subtests, de los cuales dos son opcionales o suplementarios (test de retención de dígitos y test de laberintos). Asimismo, contiene una prueba que sustituye al test de símbolos de dígitos para niños menores de ocho años y que consiste en una prueba de las llamadas de código (correspondencia de figuras geométricas con números).

La estandarización de la escala se llevó a cabo en cien niños y cien niñas de cada uno de los once niveles de edad, a partir de los cinco años. La muestra constó de 2200 niños, definida por: a) zona de residencia (rural o urbana); b) nivel ocupacional del padre, y c) zona geográfica. La proporción de la muestra se consideró a partir del censo de 1940. Los datos de confiabilidad y los errores estándar de medición son importantes en este tipo de escalas. Los datos para los Estados Unidos muestran buenos coeficientes de confiabilidad aunque no óptimos, lo que puede deberse (al igual que en México o en cualquier otro país) a que el error de medición es debido, en parte, al desarrollo psicológico y normal del niño. Sin embargo, en relación a cada una de las escalas, los índices son más confiables que cuando se aplica subtest tras subtest, a excepción del subtest de vocabulario.

Los estudios de validez demuestran que tanto los subtests como la escala completa miden factores comunes. Cada subtest verbal correlaciona significativamente con la calificación de la escala verbal total, en tanto que los no verbales correlacionan más bajo con la calificación total de la escala de ejecución. Esto demuestra que, mientras que cada subtest tiene poco en común con los demás, los subtests verbales combinados tienen más en común con cada subtest verbal individual. Lo mismo ocurre con los subtests de ejecución: en ellos, la validez concurrente alcanza altos coeficientes de correlación tanto en la escala de Stanford-Binet como en la de Raven. En cuanto a la validez predictiva, los subtests de ejecución han demostrado su eficiencia predictiva con respecto al nivel educacional.

ESTANDARIZACIÓN DE LA ESCALA DE INTELIGENCIA WECHSLER PARA ADULTOS Y PARA NIÑOS, EN LA CIUDAD DE MÉXICO

El equipo de investigaciones del INCCAPAC (Instituto Nacional de Ciencias del Comportamiento y la Actitud Pública, Asociación Civil), antes CCC (Centro de Ciencias del Comportamiento), bajo la dirección del doctor Rogelio Díaz-Guerrero, profesor e investigador decano de la psicología mexicana, ha venido realizando desde 1964 una serie de investigaciones tendientes, entre otras metas, a lograr la estandarización de un número –ahora ya muy significativo– de instrumentos de medición psicológica, entre los cuales se encuentran las escalas de Wechsler tanto para niños como para adultos. Estas actividades se han circunscrito al Distrito Federal, es decir, a la ciudad de México. La estandarización para México de la escala WISC (escala de inteligencia para niños) se inició en 1964 en un intento por proporcionar instrumentos de medición acerca de los factores de inteligencia que tuviesen una utilización más amplia en el área mexicana.

La muestra que se empleó fue la correspondiente al primer año de aplicación del estudio longitudinal sobre el desarrollo escolar mexicano. Como metodológicamente –para hacer una estandarización válida – era imposible seguir utilizando la misma muestra en los siguientes años, en vista de que se tenían los datos demográficos obtenidos a través de una exploración de la población escolar del D. F., se obtuvieron muestras representativas de las variables que definían la muestra original de los años subsecuentes (es decir, segundo, tercero, quinto y sexto grado de primaria, y segundo y tercer grado de secundaria), ya que el estudio longitudinal del IDPEM, en su primer año de investigación, sólo incluía al primero y cuarto año de primaria y al primero de secundaria. Obviamente, la decisión de excluir en la estandarización del WISC a los mismos sujetos que fueron estudiados a través de la investigación, fue tomada en función de las probables variables que podían influir negativamente en los resultados válidos de tal estandarización; por ejemplo, el aprendizaje puede influir fundamentalmente en la solución de los problemas que todo instrumento de medición psicológica plantea para producir la conducta deseada.

Los pasos específicos que se siguieron en esta investigación fueron:

- a) La obtención de una muestra suficientemente extensa y representativa de la población sujeta a estudio.
- b) La muestra contendría a escolares urbanos “estables”, de ambos sexos y de tres niveles de sistemas escolares, los cuales representan tres niveles económicos de la ciudad de México.

La estabilidad escolar se refiere a que, dada la meta de la investigación del IDPEM (llevar a cabo un estudio longitudinal), los escolares deberían, como requisito indispensable, permanecer en la ciudad de México, aún cuando durante los seis años de la investigación no vivieran en un mismo domicilio ni concurrieran a una misma escuela. La exigencia de este requisito tenía por objeto poder, en un momento dado, obtener de parte de los padres del escolar la información necesaria acerca de éste, en caso de que cambiara de domicilio o de escuela. (Como dijimos anteriormente, esto sólo ocurrió y fue tomado en cuenta para los sujetos de primer año de primaria y primero de secundaria. El resto de la muestra de estandarización se obtuvo a partir de la delimitación de la muestra original, o sea considerando las variables controladas para la elección de tal muestra, pero en sujetos que nunca habían sido examinados en los segundos, terceros, quintos y sextos grados de primaria, y segundo y tercero de secundaria.

- c) Se tomó la traducción de Puerto Rico y, comparándola con la original de Wechsler, se corrigió la traducción y se adaptaron aquellos reactivos que se consideraron no adecuados a nuestra cultura, cuidando de no alterar el contenido de los mismos. Por ejemplo, en vez de preguntar, como en la prueba original, ¿qué se conmemora el 4 de julio?, se preguntó ¿qué se celebra el 16 de septiembre? Esta operación se ejecutó en cada uno de los reactivos de los subtests que forman la escala de WISC que requirieron tales modificaciones, cuidando siempre, además de no alterar el grado de dificultad.
- d) Se administró la escala completa siguiendo las instrucciones exactamente como lo indica el autor, pero con las modificaciones para México en relación con la sintaxis, y respetando siempre el significado de tales instrucciones (el contenido).
- e) Se hizo un análisis de dificultad de los reactivos, a fin de establecer el orden pertinente para México. Es menester apuntar aquí los resultados dramáticos observados. Se presentaron diferencias importantes entre la escala original (norteamericana), la traducción de Puerto Rico y la muestra para México.
- f) Se hizo una recalificación con el orden obtenido para México.
- g) Se ejecutó un análisis por edades.
- h) Se procedió a la transformación de las puntuaciones crudas a calificaciones *z* para, a partir de ahí, transformar a CIs –en función del porcentaje de respuestas – la media (*X*) y la desviación estándar, primero en cada subtest, utilizando una media de 10 y una desviación estándar de 3, y para los CIs de las escalas completa, verbal y de ejecución, así como para el CIs total, con una media de 100 y una desviación estándar de 15.
- i) Se estableció la confiabilidad mediante el test-retest, obteniendo un alto coeficiente de confiabilidad temporal significativa al 0.01 y al 0.05 como sigue:

Tabla 4

Test-retest del WISC en un intervalo de 12 meses para tres grupo de edades

<i>Subtest</i>	<i>Grupo I</i>	<i>Grupo II</i>	<i>Grupo III</i>
Información	0.60	0.62	0.63
Comprensión	0.31	0.25	0.50
Aritmética	0.40	0.53	0.52
Semejanzas	0.46	0.47	0.47
Vocabulario	0.36	0.41	0.40
Retención de dígitos	0.18	0.37	0.62
Figuras incompletas	0.41	0.33	0.64
Ordenamiento de dibujos	0.43	0.23	0.43
Diseño con cubos	0.64	0.64	0.63
Ensamble de objetos	0.50	0.37	0.33
Símbolos de dígitos	0.21	0.41	0.47
Puntuación de escala verbal	0.72	0.74	0.76
Puntuación de escala de ejecución	0.70	0.60	0.64
Puntuación de escala de ejecución	0.79	0.77	0.79

Grupo I = *r* entre 6 y 7 años (133 casos).

Grupo II = *r* entre 9 y 10 años (118 casos).

Grupo III = *r* entre 12 y 13 años (133 casos).

r significativa: al 0.01 = 0.228

al 0.05 = 0.174

- j) Con la primera aplicación del WISC se comprobó la validez de construcción mediante el análisis factorial denominado de *eje principal*, es decir, se logró el máximo grado de varianza de cada factor.

Se ejecutó un análisis factorial por edades y se realizaron análisis de varianza (ANOVAS) para probar el poder discriminativo por edad, de cada subtest, así como las hipótesis de diferenciación por sexo y por nivel socioeconómico.

Como dato importante, es preciso enfatizar que por una parte hay una diferenciación significativa por edades, en tanto que por otra se van perdiendo los efectos del nivel socioeconómico.

Resumiendo, el análisis factorial comprueba los resultados obtenidos por Wechsler y por otros investigadores con la escala WISC.

4

MEDICIÓN DE HABILIDADES MÚLTIPLES

CONCEPTO DE HABILIDAD

Las habilidades humanas no han sido totalmente generales ni absolutamente específicas. En los últimos años la utilización del análisis factorial ha permitido entender, cuando menos, algunos de los factores fundamentales que operan en la solución de pruebas que exigen determinadas tareas y, por lo tanto, cierto tipo de habilidad o habilidades.

Lo importante de los hallazgos es que las correlaciones entre las habilidades humanas casi siempre son positivas, aunque bajas en algunos casos.

De este modo, sería difícil encontrar una habilidad humana cuyo alto nivel de ejecución determinara o indicara que el sujeto se comportará deficientemente en otro tipo de tarea.

Las habilidades humanas guardan una relación muy estrecha con las diferencias individuales, problema en el cual los psicólogos han estado muy interesados.

Podemos definir las habilidades mediante la pregunta: ¿qué tan bien ejecutan las personas las diversas tareas que se les presentan?, es decir, ¿qué habilidades son necesarias para ejecutar una tarea a partir de un criterio estándar mínimo de ejecución?

No se puede estudiar un tipo particular de habilidad sin definir antes que entendemos por “qué tan bien es la ejecución de una tarea” dada en forma objetiva. La definición implica que “la persona intenta hacer algo”, lo cual pone de manifiesto la necesidad de medir las habilidades en

aquellas situaciones en las que suponemos que los individuos están motivados para hacer el trabajo “lo mejor que puedan”.

Las diferencias entre las personas, de una a otra habilidad, son infinitas. A través de las investigaciones realizadas en este campo se han encontrado sólo algunas habilidades que merecen la atención. De este modo, se distinguen las *habilidades intelectuales* de las *habilidades especiales*. Las primeras, por darles un significado que nos permita entenderlas, son aquellas que se requieren para el trabajo científico y que tienen una presencia absoluta y demostrable, aunque sea una mínima cantidad. Las segundas son la destreza o exactitud con que se ejecuta una tarea particular. No obstante estas definiciones, es difícil hacer una clara distinción entre ambas clases de habilidades y establecer una separación entre ellas.

Los diversos análisis factoriales producen ciertas diferencias significativas en el tipo de factores que se han encontrado y difieren ampliamente en el número de ellos, el cual depende de qué tan profundamente se quiera abordar el problema; por ejemplo, muchos factorialistas británicos se han conformado con la exploración de factores fundamentales, es decir, solamente con aquellos más importantes que permitan explicar la mayor parte de la varianza de la prueba. Así, por ejemplo, Thurstone, en 1941, en su primera investigación informa de la presencia de nueve factores en una muestra de estudiantes de preparatoria. Posteriormente, este mismo investigador realizó un segundo análisis factorial, con el fin de probar si los mismos nueve factores encontrados en sujetos de preparatoria se hallaban también en niños pequeños. Entonces, utilizando sesenta pruebas psicológicas, especialmente construidas para este propósito y administradas y calificadas en forma estandarizada, Thurstone encontró siete factores detectables. Sin embargo, investigaciones anteriores realizadas en Inglaterra y los Estados Unidos, hasta los estudios de Thurstone, habían sostenido la teoría de la existencia de siete factores fundamentales, más otros sin número limitado, de segundo orden. (Estos factores son analizados en el capítulo 3, dentro del tema “Desarrollo del concepto de inteligencia”.)

Los factores de habilidad para predecir el comportamiento en situaciones de la vida real son, potencialmente, múltiples. En general, han sido utilizados en relación con el éxito obtenido en la escuela, en todos sus niveles, y con el buen desempeño en algunas ocupaciones específicas.

Ferguson, en 1954, consideraba que la transferencia de entrenamiento en experimentos de aprendizaje era regida por factores de habilidad humana; es decir, si a una persona se le da entrenamiento en cierto tipo de tarea que tenga relación con un tipo específico de habilidad, habrá transferencia positiva a otra clase de problemas relacionados con el mismo factor; por ejemplo, si a una persona se le entrena para que ejecute una tarea específica de razonamiento, habrá transferencia de aprendizaje para algunas tareas, aparentemente diferentes, pero relacionadas con el mismo factor.

Guilford, en 1961, puso de manifiesto las relaciones existentes entre los factores de habilidad humana y los procesos, investigados en experimentación controlada, de formación de conceptos, aprendizaje, reforzamiento, memoria y otros. Ya en 1954, Fleishman y Hempel habían hecho experimentos para relacionar factores de habilidad y el progreso en pruebas de razonamiento o de aprendizaje. Al respecto, encontraron que la diferencia fundamental entre los diversos factores radicaba en el estadio del aprendizaje. Al aprender una tarea psicomotora, en los primeros tests, aparecía simultáneamente un factor de tipo espacial, de gran trascendencia, que iba desapareciendo de manera gradual, hasta que en las últimas pruebas se presentaba con una correlación muy baja, con

la ejecución específica de la tarea. Cuando observamos estas situaciones necesariamente tenemos que recordar que, independientemente de la predisposición biológica que trae consigo cada individuo, es muy importante el aprendizaje; es decir, ¿qué hacer con esas habilidades en potencia?; ¿cómo hacer para obtener de ellas lo máximo que, hipotéticamente, podrían producir?

Para la ejecución de una tarea, necesariamente se tiene que dar la combinación de una serie de operaciones, tanto intelectuales como motoras. Dependiendo del tipo de ejecución que se exija, serán las operaciones que jerárquicamente aparezcan cuando podamos probar las habilidades individualmente o agrupadas en factores. Estos factores pueden hacer su aparición ya sea matizando en forma general toda la ejecución de la tarea u operando solamente con algunos de sus elementos, es decir, con aquellos que sean de mayor necesidad para la realización del trabajo.

De la manera como el sujeto utilice sus habilidades o de aquellos factores generales necesarios para la ejecución de tareas similares, dependerá la capacidad que se tenga para predecir cómo actuará en situaciones posteriores semejantes. Cuando lo hacemos con la ayuda de tests apropiados para ello (es decir, mediante el uso correcto de la instrumentación adecuada), puede establecerse la predicción de la conducta.

A menudo las investigaciones acerca de estos problemas se han enfocado más bien a aspectos particulares; por ejemplo, interesa mucho el desarrollo escolar del individuo, o bien, su éxito ocupacional o profesional. Sin embargo, la vida cotidiana, en su totalidad, exige la combinación no sólo de las habilidades para ejecutar las tareas necesarias para la realización en la escuela o en el trabajo, sino también de las habilidades para desempeñarse en cualquier otro campo. Así pues, la tendencia de los investigadores ha sido el estudio de habilidades escolares y habilidades especiales. No obstante, necesitamos saber mucho más que eso. Claro que esas investigaciones no son inútiles; por supuesto que muchas de esas habilidades o factores se necesitan y operan, de hecho, cotidianamente. Sin embargo, lo importante es saber cómo se desempeñan los sujetos en tareas aparentemente diferentes.

La predicción sobre los factores de habilidad humana depende fundamentalmente del entrenamiento y de la experiencia en las tareas. Es muy importante encontrar esa validación predictiva en la vida diaria, ya que casi todo lo que se ha hecho ha sido en los laboratorios. Hay que comprobar su realidad; hay que observar y seguir detenidamente a las personas en las tareas que emprenden.

Las investigaciones más acertadas acerca de los elementos componentes de la inteligencia humana no tienen más de treinta años. Uno de los investigadores más notables en este campo es Luis L. Thurstone, con sus estudios sobre las habilidades primarias; es el iniciador del empleo de aplicaciones experimentales del análisis factorial para poder descubrir y entender la complejidad del intelecto. De este modo, se ha podido elaborar una teoría unificada de este aspecto: “la estructura del intelecto”. Guilford, autor de esta teoría, llegó a diseñarla a partir de una clasificación de los factores, en relación con el proceso u operación ejecutada.

Es evidente que en toda tarea participan múltiples elementos que producen el comportamiento adecuado y necesario; pero, además, se encuentran presentes otros factores que influyen determinantemente en el resultado. Así, pues, los factores motivacionales, afectivos,

físicos, fisiológicos, etcétera, deben tenerse en cuenta al hacer valoraciones sobre el comportamiento de un individuo, ya sea de tipo general o de aptitudes especiales.

En el capítulo 3 se describe con cierta amplitud la teoría de la estructura del intelecto, diseñada por J. P. Guilford, la cual permite entender mejor lo explicado hasta aquí. Basta añadir que la idea de Guilford, en parte, tiende a explicar la estructura del intelecto, primero por un gran número de factores separados, y luego por la búsqueda de más factores y elementos que lo complementan.

Algunos investigadores han propuesto diseñar un modelo jerarquizado que conduzca a niveles cada vez más generales, es decir, a factores que abarquen un mayor número de elementos. Al respecto, en 1962, Humphreys elaboró un modelo en el cual hipotetiza cuatro niveles discriminativos que podrían encontrarse en un test sobre información mecánica: a) información de implementos específicos; b) capacidad para discernir o comprender; c) información acerca de las áreas de ejecución mecánica (plomaría, carpintería, mecánica automotriz, etcétera) de tipo especializado, y d) información mecánica de tipo general.

Guilford también intenta diseñar un modelo jerárquico de las posibles relaciones entre diferentes tipos de factores, colocando primero el factor *g* de Spearman, luego el pensamiento, después el pensamiento productivo y, más adelante, dos tipos de pensamiento productivo. Aún así, habrá que considerar otros tipos de factores por cuanto toca a la jerarquización de las operaciones y productos del pensamiento.

Por otra parte, Nunnally indica que sería más adecuado trabajar partiendo de la base hacia arriba, y no a la inversa; por ejemplo, en lugar de empezar por el final, el intelecto, tal vez sería mejor comenzar por el pensamiento productivo y buscar las jerarquías subyacentes a éste.

MEDICIÓN DE MULTIHABILIDADES. TESTS Y BATERÍAS DE TESTS

La medición de grupos de habilidades mediante un solo instrumento parte de la base de que la inteligencia es una capacidad general que se puede medir a través de una buena muestra de actividades mentales.

Esta muestra de actividades mentales, para su medición, es estructurada dentro de lo que se llama *tests de multihabilidades o escalas de multihabilidades*. En ellas se pueden observar dichas actividades, distribuidas en formas diferentes, ya sea formando subtest conteniendo cada uno una muestra de los ítemes con que se intenta medir la habilidad correspondiente y que juntos forman una escala de multihabilidades, o bien, en tests separados, pero con las mismas características.

El principio de distribución de los ítemes en cada test o subtest es el de “ser progresivamente difíciles”, es decir, ir desde el de más sencilla ejecución hasta el más complicado, sobre la idea de que cada vez que el individuo ejecuta satisfactoriamente un grupo de ítemes se va acercando a su máximo de ejecución, en un tipo especial de actividad intelectual.

La mayor parte de las escalas de multihabilidades se califican primero en cada uno de los subtests, y luego se obtiene una calificación general; además, son calificadas con mayor rigidez que las escalas especiales. También resulta importante que los subtests de la escala sean de poder o de

velocidad; de este modo, se impone el tiempo límite para cada uno de ellos, dependiendo de sus objetivos. Cabe agregar que hay escalas verbales, de ejecución y también combinadas o mixtas.

Este tipo de tests se utiliza para la selección y clasificación de personal, ya sea para diversas clases de ocupación, profesión o estudios, o simplemente para buscar cuál puede ser el campo en donde mejor se desempeñe un sujeto. Al darles esa utilización, con el fin de distinguir su uso como detectoras de inteligencia general, de otros usos o para saber que son tests específicamente diseñados con estos objetivos, se les denomina *escalas o tests de aptitudes múltiples*.

Es conveniente establecer aquí el por qué de la diferencia entre el uso o la construcción de estas pruebas psicológicas. Al respecto podemos decir que la aptitud es diferente de la habilidad y del perfeccionamiento. Así pues, consideramos a la habilidad como la capacidad de ejecutar una acción con facilidad y precisión, y hablamos del perfeccionamiento cuando la habilidad incluye no sólo actividades motoras y manuales, sino también cognoscitivas concretas, como el manejo del lenguaje, de los números y, más exactamente, de las matemáticas, de la historia, de la biología, etcétera. Hablamos de perfeccionamiento en cualquier tipo de ejecución. Por otra parte, cuando nos referimos a las aptitudes que tiene un sujeto para desarrollar cierto tipo de actividad, estamos considerando la capacidad *adquirida* para desempeñarse con perfección en condiciones adecuadas, es decir, estamos teniendo en cuenta sus posibilidades en un momento dado.

Este tipo de tests son de uso común en el campo de la psicología industrial y en la educacional principalmente; sin embargo, también tienen cierta aplicación en el área de la psicología clínica para valorar la conservación de las habilidades o aptitudes de los pacientes, y así trazar su programa de rehabilitación.

En la selección y clasificación de personal en el campo de la psicología del trabajo se utilizan múltiples escalas con fines evaluativos, los cuales están formadas por tests construidos independientemente uno de otro, pero que son susceptibles de proporcionar información acerca de cuáles son las áreas de funcionamiento más eficientes de un individuo, a fin de expresar el “perfil” de éste, el cual es muy útil en estos casos.

Resulta muy difícil tener a la mano la información acerca de cuáles son las baterías de pruebas más utilizadas en esta área en México, por la forma en que se trabaja; no obstante, hay algunas escalas de uso muy común, aún cuando ninguna está estandarizada ni validada realmente para nuestra cultura; en consecuencia, el desempeño mediante estos instrumentos se ve afectado considerablemente por ella.

Un ejemplo de lo anterior lo constituye el manejo del tiempo. Así tenemos que, en investigaciones hechas en México y en los Estados Unidos, al establecer una comparación entre los tiempos de reacción y de ejecución registrados en diversos grupos caracterizados por variables diferentes, las pruebas de significancia muestran que, estadísticamente, existen diferencias muy significativas entre ambas culturas. Esto da lugar a plantearnos una serie de interrogantes, entre las cuales la más importante sería: ¿con qué confianza estamos utilizando tests no estandarizados en México para seleccionar y clasificar personas, si el patrón a seguir no es el nuestro, culturalmente hablando?; asimismo, ¿por qué no hemos podido construir o manejar correctamente tests apropiados, para que su validez y confiabilidad nos permitan estar seguros de lo que hacemos? Estas son preguntas que tienen una sencilla contestación, pero que implican una difícil resolución.

Al respecto, citaremos un ejemplo de una batería de aptitudes múltiples que ha sido adaptada, hasta cierto punto, en México, pero que es de uso común para determinados tipos de ocupación o profesión. Esta es la *Batería Diferencial de Aptitudes*, mejor conocida en México como *DAT*. De esta prueba se ha dicho que es una de las mejores contribuciones que se han dado al campo de la medición de aptitudes múltiples. Así Carroll, en 1960, dijo: “Los autores han hecho un trabajo tan completo y técnicamente satisfactorio, que cualquier revisor lo encontraría difícil de hacerlo parecer suficientemente crítico para él mismo.”

Uno de los aspectos más importantes que pueden advertirse en la forma en que el manual está estructurado es aquel que expresa con todo cuidado los fundamentos de construcción y los principios que sostienen la escala.

En el *DAT* se define a la aptitud como “aquellos síntomas de capacidades para un aprendizaje específico”. Las aptitudes seleccionadas para ser medidas fueron aquellas que se consideraron de suficiente interés para demostrar su éxito en la gran variedad de objetivos.

El desarrollo de cada uno de los tests fue realizado en función de ciertas condiciones especiales. Así tenemos que: *a)* cada uno de los tests, en forma y utilidad, es independiente de todos los demás, incluyendo su validez y su confiabilidad; *b)* todas las pruebas fueron diseñadas como tests de poder, con tiempo límite, por conveniencia en la administración; *c)* la batería puede producir un perfil, y *d)* la batería es de fácil aplicación. Además, el *DAT* no fue construido como un conjunto de medidas factorialmente puras. De ser así, esto dependería de los procedimientos desarrollados, más que de la base sobre la que están construidos. Además, son intencionalmente homogéneos, es decir, utilizan un solo tipo de ítemes en cada uno de los subtests.

Ahora bien, el *DAT* está constituido por los siguientes tests:

- a) Razonamiento verbal:* tiene un tiempo de duración máxima en la ejecución, 30 minutos, y mide la habilidad para comprender conceptos verbales.
- b) Habilidad numérica:* mide la comprensión y la facilidad en el manejo de números. Los ítemes son de aritmética de elección múltiple. Tiene un tiempo de duración de 30 minutos.
- c) Razonamiento abstracto:* es un test no verbal de habilidad para razonar en dónde está formado cada ítem por cuatro figuras que representan posiciones diferentes. Tiene un tiempo de duración de 25 minutos.
- d) Relaciones espaciales:* mide la capacidad para trabajar con objetos concretos mediante la visualización. Aquí los ítemes son una combinación de factores espaciales rotados con la visualización del ítem patrón. Tiene un tiempo de duración de 30 minutos.
- e) Razonamiento mecánico:* este es el test de comprensión mecánica de Bennett. Aquí, durante 30 minutos, el sujeto tiene que ejecutar una tarea consistente en seleccionar, de una serie de respuestas de opción múltiple, las tareas que se le muestran en tarjetas fotografiadas.

- f) *Velocidad y exactitud*: mide la velocidad de la respuesta en una tarea simple, perceptual. Este test está constituido por cinco pares de letras o números colocados arriba de una línea de puntos. En una hoja de respuestas, en la cual estas letras o números se encuentran invertidos, el sujeto debe marcar en dónde considere correcto hacerlo.

Es el ambiente escolar y, en general, en el académico suelen medirse habilidades intelectuales; por ejemplo, una situación que es importante advertir, desde los niveles escolares más inferiores hasta los universitarios y profesionales, es la capacidad de “comportamiento inteligente general” en todos aquellos individuos que sea necesario por los requerimientos de sus actividades en un momento dado; o sea, es importante investigar si las tareas que intenta desarrollar el individuo o que el medio le impone que desarrolle son las adecuadas en ese momento, en primer lugar, para desarrollar sus “habilidades generales”; luego, se harán los estudios específicos para advertir sus aptitudes especiales. (En la psicología del trabajo y en la selección y clasificación de personal se procede del mismo modo).

Desde luego, a niveles escolares inferiores, en México no se trabaja de esta manera en forma consistente y planeada. Desafortunadamente, sentimos que esas actividades las podemos ejecutar desde nuestra particular observación de los alumnos y desde nuestro particular punto de vista. Así pues, es rara la escuela que lleva a cabo un registro completo de cada alumno desde sus primeros años de instrucción. Por otra parte, estos estudios deben ser realizados por especialistas, ya que en ocasiones la administración de los instrumentos psicológicos puede ser sencilla y hasta su forma de calificación; pero lo importante no es sólo la correcta aplicación de los tests, sino la interpretación, con base en conocimientos especializados, de los resultados numéricos que se obtienen de casi todos ellos.

Obviamente, uno de los principales objetivos que persigue la instrucción escolar es la enseñanza por parte de los maestros, y el aprendizaje de lo enseñado, por parte de los alumnos. Así, los objetivos educacionales son la medición y la evaluación de los resultados de la instrucción que se encuentran dentro de las actividades del maestro. El es quien tiene que efectuar esa medición, para descubrir el grado de aprendizaje que ha logrado el alumno y para solucionar los problemas de la enseñanza que se presenten. Ahora bien, uno de los problemas que sistemáticamente se encuentran aquí es la forma de llevar a cabo esa medición y esa evaluación, que permite juicios muy subjetivos por parte de los maestros, debido no sólo al número de alumnos que generalmente tienen a su cargo, sino también a que tradicionalmente se ha empleado una serie de métodos tanto en la enseñanza como en los hábitos que el alumno debe adquirir para aprender, los cuales están tan arraigados que difícilmente se pueden tener en cuenta sugerencias que mejoren esta situación. Los maestros de cualquier nivel escolar pueden construir objetivamente sus instrumentos a fin de medir el aprendizaje, con lo cual no sólo facilitarán sus labores, sino también tendrán la seguridad de ser más justos en sus evaluaciones y, como consecuencia, los resultados de la enseñanza cada día se verán incrementados positivamente.

En el siguiente capítulo expondremos las formas y tipos de tests psicológicos relacionados con el desarrollo del infante, del preescolar y del niño escolar. En él quedan incluidos los tests adecuados para medir y evaluar las habilidades fundamentales de esta clase de niños, así como para desarrollar su actividad escolar y también algunas formas estandarizadas en México para medir el rendimiento escolar.

Existen baterías de aptitudes que pueden ser utilizadas en los diversos campos de la actividad humana con el fin de seleccionar, clasificar y evaluar a las personas para que ejecuten diversas tareas. No obstante, como ya indicamos anteriormente, estas baterías son de “importación”, lo cual pone en entredicho los datos que arrojan y, por tanto, la injusticia que pudieran producir, pues en muchos casos no han sido siquiera adaptados a la situación mexicana que se trata de valorar. Sin embargo, tenemos que aceptar traerlas y utilizar algunas de ellas, en vía de economía de tiempo, dinero y esfuerzo, pues fundamentalmente no difieren de las construidas en México; así, pueden usarse, siempre y cuando sean estandarizadas y adaptadas a nuestro medio. Esto sólo podrá ocurrir cuando valoremos aspectos generales. Por el contrario, cuando se trate de valorar situaciones específicas, lo mejor será construir instrumentos adecuados a nuestras necesidades culturales. Por ello, todo psicólogo que quiera ser eficiente profesionalmente, debe conocer la metodología, el orden y los requisitos necesarios para construir tests más útiles para nuestro medio.

En nuestra experiencia diaria encontramos que, a veces, es más conveniente utilizar lo elaborado por otros, pero teniendo en cuenta los cambios a que habrán de someterse los instrumentos, dadas las circunstancias, con el fin de ahorrar tiempo y trabajo, y no venir a descubrir lo que otros ya han descubierto desde hace tiempo, pero que nos puede ser de utilidad sabiendo qué hacer con ello. Por ejemplo, las baterías de aptitudes pueden no ser de absoluta utilidad para nosotros; sin embargo, si tomamos su estructura y sobre ella construimos ítems apropiados a lo que en un momento dado se nos exige, habremos avanzado más rápidamente que aquellos que tuvieron que sentarse a pensar “cómo” podrían medir aquello que necesitaban. Por supuesto, sería mejor si somos capaces de crear lo particular para solucionar los problemas que se nos plantean en un momento dado. Por ello, siempre haremos hincapié en que los conocimientos del psicólogo deben conducirlo a resolver todo lo necesario para ejecutar sus tareas con éxito.

MEDICIÓN DE LAS HABILIDADES ESPECIALES

En párrafos anteriores se dio el concepto de habilidad de un individuo. Lo ratificaremos aquí para una mejor comprensión de la información acerca de los tests de aptitudes especiales.

Así, entendemos por habilidad o aptitud especial de un individuo la capacidad que éste tiene para adquirir, mediante entrenamiento especial, cierto tipo de conocimiento; o sea, la habilidad potencial de una persona para ejecutar cierto tipo de actividad especializada que, mediante su ejercicio, pueda rendir al máximo.

Ahora bien, como sabemos, existen diferencias individuales que se pueden advertir con facilidad y dignas de subrayar, como son: las habilidades sensoriales, motoras, mecánicas, artísticas, etcétera. Es frecuentemente factible encontrar que dos personas tienen una calificación igual o similar en tests de inteligencia general, los cuales, como hemos visto, pueden estar constituidos por un gran número de subtests que valoran independientemente cada uno de los factores que, de acuerdo con la construcción teórica del instrumento, son componentes fundamentales de la estructura intelectual. Sin embargo, a pesar de que la calificación total sea similar para ambos, su comportamiento en términos del rendimiento en cada uno de los tests seguramente en términos del rendimiento en cada uno de los tests seguramente que será diferente. Quizá uno de los individuos tuvo una mejor ejecución en comprensión o se desempeñó mejor en sus funciones analítico-sintéticas, y el otro manejó mejor los factores numéricos, su análisis perceptual o su capacidad de retención.

No debe olvidarse que, a pesar de que los tests psicológicos tienen un nombre para designarlas, las funciones fundamentales siempre estarán presentes y el conocimiento de ellas permitirá una interpretación realmente certera de los resultados que obtengamos al efectuar la medición.

Sí, como ya dijimos, las habilidades especiales son producto de un entrenamiento especial en ciertas tareas, entonces tendremos que hablar de aprendizaje. En este sentido, los tests de inteligencia general son índices de la habilidad para aprender; es decir, medir la capacidad de aprendizaje implica obtener datos sustraídos de una evaluación previa a un periodo de práctica estudio o instrucción. Esto trae como consecuencia la necesidad de determinar la validez predictiva de aquellos instrumentos que utilizemos para ejecutar la tarea. Conviene recordar aquí que los tests de inteligencia propiamente dichos, y los tests de rendimiento o logro –que son o pueden ser los primeros, sólo que manejados en forma diferente – dan resultados que hipotéticamente deben tener altas correlaciones respecto a la conducta que se mide. Sin embargo, en la práctica esto no sucede. Por ello, pensamos que aplicar un pretest y un postest después de una o más tareas impuestas a un sujeto para que las aprenda, es lo más sensato y completo para probar en qué grado ha aprendido. No obstante, la confiabilidad en pruebas de logro suele ser baja y sin un alto coeficiente de estabilidad temporal; por ello, estos instrumentos y sus resultados se convierten en obsoletos por irreales, y sus correlaciones con tests de inteligencia, o con cualquier otra medición independiente, son también muy bajas.

Las mediciones de rendimiento en habilidades deben ser altamente confiables; además, si la medición del aprendizaje es el criterio mediante el cual pruebas de inteligencia u otras medidas son validadas como mediciones de habilidad para aprender, tienen que exigirse requisitos para la situación de aprendizaje. Uno importante es dar a cada quien la oportunidad de aprender de acuerdo con su habilidad. De otra manera, la medición del logro obtenido deja de ser criterio de habilidad para aprender, puesto que ha existido una limitación de oportunidades. Esto sucede muy frecuentemente en el ámbito escolar, sobre todo por la estructura académica-administrativa, que intenta abstraerse y olvidar al individuo, para centrarse en las instituciones.

Los límites impuestos a la enseñanza-aprendizaje de años escolares, semestres, horas-clase, etcétera, presionan a maestros y alumnos, limitando la ayuda instruccional. Además, la posibilidad de validación no debe restringirse en la situación de enseñanza-aprendizaje, ni mucho menos en la situación de test. No es suficiente que cada quien tenga oportunidad de aprender de acuerdo con su habilidad, sin que debe dársele también la oportunidad de tener a su alcance la totalidad de lo que tiene que aprender, a fin de que pueda reflejarse al medir el logro. En estos casos hay una tendencia a errar en la obtención de la confiabilidad de la medición, pues se piensa que una buena prueba inicial puede ser una buena prueba final y, en general, que un buen test dará buenas medidas de logro. Esto puede no resultar cierto, porque al final de la situación de evaluación del logro se pueden producir medidas de aprendizaje muy pobres, las cuales no debemos, si somos honestos, considerar, a menos que tengamos en cuenta una buena distribución de medidas de aprendizaje. Si esto es así, solamente necesitaremos un buen test para medir el aprendizaje tantas veces como se requiera.

Si entonces hipotéticamente, hiciéramos un conjunto de ítemes con un rango de 0 a 100 y con un 50% de dificultad, la problemática del examen inicial y del final desaparece. El problema se reduce, entonces, a la dificultad de aprender. De tal suerte, el análisis de ítemes en estos casos debe

basarse *no* en el porcentaje de respuestas correctas, sino en el porcentaje de lo aprendido para responder correctamente.

Por último, diremos que, entre los dos grupos de datos obtenidos, las calificaciones de prueba de inteligencia constituyen buenos elementos de predicción del aprendizaje, y lo expuesto anteriormente debe ser aplicado a otras áreas de la medición de habilidades, ya que en cualquier tarea, no importa cuál sea su índole, se requiere la medición del aprendizaje en sus diferentes niveles o momentos del proceso, con el fin de poder alcanzar el objetivo principal de nuestra labor: clasificar, predecir y validar las conductas evaluadas mediante los instrumentos adecuados y útiles para medir lo que queremos.

Los tests de aptitudes especiales son de muchos tipos, de múltiples formas de administración, de formas muy diversas de calificación, etcétera. Como hemos venido haciéndolo, en este capítulo trataremos de transmitir lo que sabemos que se hace en México en las diversas áreas de aplicación de la psicometría.

Los tests sensoriales o de tipo sensorial pueden ser, como su nombre lo indica, en formas tan diferentes como distinto sea el órgano de los sentidos que entre *preponderantemente* en juego. Subrayamos el término preponderantemente porque en la mayoría de los tests psicológicos, y no sólo en los sensoriales, interviene, en situaciones de normalidad, el órgano visual en primer lugar y el auditivo en segundo; los demás órganos son muy importantes en la ejecución de estos tests, pero nunca en la misma proporción que hay para el tipo de ejecuciones cotidianas.

Las pruebas construidas para evaluar aptitudes visuales son bastante completas. En ellas debemos distinguir la agudeza visual para ver de cerca y la agudeza visual para ver de lejos, así como la percepción de profundidad (es decir, la habilidad para distinguir la proximidad de unos objetos con otros) y la percepción de color, la cual es cuestión de grados y de una complejidad importante para su medición.

Los tests de audición están constituidos por diferentes funciones, pero, en términos generales, miden la habilidad para distinguir sonidos finos (Esto también es cuestión de grados).

Los tests motores y manuales son de los instrumentos más antiguos de la medición de las diferencias individuales. Uno de ellos es el dinamómetro, que mide la fuerza del puño de la mano o, más bien, la presión de los dedos contra la mano; la fuerza es medida en kilogramos. Aparentemente este instrumento intenta medir también el grado de agilidad manual y el nivel de fatiga, y está diseñado para ser utilizado en todas aquellas ocupaciones que requieren estas dos características fundamentales.

Otros de los tests que suelen incluirse en esta clasificación son los de la medición del tiempo de reacción. (Llamamos tiempo de reacción al tiempo transcurrido entre la presentación de un estímulo y el momento en que el sujeto inicia su respuesta). Como es sabido, este tipo de pruebas se aplican para medir la velocidad de respuesta en situaciones que exigen una reacción inmediata, como podría ser: operaciones con ciertas máquinas, manejar un automóvil, etcétera.

Otro instrumento es el dexterímetro, que como su nombre lo indica, mide la destreza manual, la velocidad de los movimientos gruesos de manos y brazos, el ritmo y la coordinación

manual y la coordinación y control de los dedos en diversos grados. La precisión de la mano es medida por la seguridad con que el sujeto es capaz de introducir una pequeña barra cilíndrica de metal en pequeñas perforaciones circulares que se encuentran en una plancha de metal conectada a la electricidad. El contacto de las barritas con los arillos de las perforaciones registra la medida.

Al igual que este instrumento, existen múltiples aparatos diseñados para medir las habilidades motoras y manuales. Así, hay pruebas de habilidades mecánicas que son instrumentos combinados de capacidades motoras y sensoriales, algunos de los cuales contienen también la medición de habilidades perceptuales de relaciones espaciales, información sobre aspectos mecánicos y la capacidad de comprender relaciones mecánicas. Estas pruebas se han diseñado para medir la capacidad y la ejecución a niveles más altos y complejos que los de un simple dexterímetro.

Por otra parte encontramos tests de aptitudes especiales para medir otro tipo de habilidades, como son las artísticas y las profesionales. Entre los primeros podemos destacar aquellos que miden aptitudes musicales, aptitudes en artes gráficas, tests de aptitudes estéticas, como la escultura, la arquitectura, los productos industriales, etcétera. Como ejemplo de los segundos tenemos las pruebas de aptitudes para la medicina, para la jurisprudencia, para la enseñanza, para la ingeniería, etcétera.

Generalmente, las pruebas de habilidades no son tan específicas como para darnos una guía sobre las bases en que se ha construido un test de aptitudes, ya que interfieren la motivación dada por los intereses, valores y preferencias de una persona. Sin embargo, la motivación puede ser un factor decisivo para la selección de una carrera o de una ocupación. A menudo las personas se ven influidas por situaciones externas, más que por su propia evaluación y conocimiento del campo. Por esta razón se ha intentado construir una serie de instrumentos, con el fin de detectar los intereses y preferencias de un individuo en áreas limitadas.

En esta sección mencionaremos algunos inventarios de intereses ocupacionales y profesionales que pueden servir como instrumentos para ayudar a la elección, tanto por parte del psicólogo como del examinado, de aquellas tareas en las que supuestamente el sujeto puede desempeñarse mejor y que obedecen a sus valores y preferencias.

INVENTARIOS DE INTERESES, VALORES Y PREFERENCIAS

Inventario de intereses vocacionales de Strong

Este inventario, de uso común en México sin que aún se haya hecho una buena adaptación, cuando menos para utilizarlo en todas aquellas situaciones en que sea necesario, tiene por objeto detectar gustos y desagrados sobre ocupaciones, materias escolares, actividades cotidianas y rasgos peculiares de personalidad. El orden de preferencia en las actividades que cada quien ha de desempeñar, la importancia de los factores que afectan al trabajo, el orden de preferencia del hombre o de la mujer, lo que a uno le gustaría haber sido, la posición que a uno le gustaría ocupar dentro de una organización, la comparación de intereses entre pares de ítems y el autorregistro de habilidades y rasgos actuales con los objetivos fundamentales de este inventario. De los elementos mencionados que los constituyen, intenta medir la extensión en que cada uno de ellos (los intereses y preferencias de un individuo) concuerdan con los éxitos personales de un sujeto en sus ocupaciones específicas.

Este inventario se ha diseñado en dos formas, uno para hombres y otro para mujeres, y es aplicable desde los 17 años. Cada ocupación se califica por separado y los registros obtenidos se consideran como los “indicadores” de los intereses del sujeto en cada ocupación. Puede agruparse también en seis grupos ocupacionales, más que por ocupaciones específicas. Además, califica los intereses no ocupacionales que pueden ser de utilidad en la orientación al sujeto, tales como maduración en el área de intereses, masculinidad-femineidad, nivel ocupacional (diferencias entre sujetos que trabajan como obreros y otros del mismo tipo, y sujetos profesionistas o de negocios), nivel de escolaridad, etcétera.

Inventario de Kuder y Strong

El registro de preferencias de Kuder es uno de los inventarios elaborados por este autor. Puede ser aplicado a sujetos que tengan desde 15 años de edad. Está constituido por 168 ítemes distribuidos en tríadas, es decir, en cada respuesta se debe indicar cuál de los tres ítemes agrada más y cuál menos.

Este inventario cubre un rango de actividades muy amplio y su calificación produce un perfil que representa diez áreas: mecánica, científica, persuasiva, artística, literaria, musical, de servicio social, de oficina, trabajo al aire libre y de computación o cálculos numéricos. La prueba original utiliza los rangos percentilares como unidad psicométrica de calificación para cada una de las áreas. El perfil se analiza con la intención de determinar en cuál de las áreas los intereses y preferencias del sujeto presentan cargas significativas.

La identificación de las áreas no determina de ninguna manera que el sujeto sea canalizado en sus preferencias por alguna ocupación específica detectada por el instrumento. En cada área hay múltiples ocupaciones específicas relacionadas entre sí que pueden ser especialmente tomadas en cuenta. Así pues, en un perfil es frecuente encontrar más de un área de preferencias. Al respecto, Kuder ha formado pares de preferencias de posibles intereses ocupacionales: la mecánica-artística, la mecánica-científica, la científica-artística, la científica-social, la persuasiva-literaria, etcétera.

Los inventarios de intereses y preferencias ocupacionales como el de Kuder y el de Strong no son tests de aptitudes, sino que sólo indican la extensión en que se expresan intereses y preferencias por quienes son sometidos a ese tipo de examen y que tienen empeño en el éxito dentro de sus áreas específicas ocupacionales.

El principio fundamental que rige a estos inventarios es el de que un individuo que tiene un patrón de preferencias y desagradados similares a los detectados en un grupo u ocupación dada, tiene una alta probabilidad de encontrar el tipo de actividad deseada y, por supuesto, probar posteriormente que tiene las aptitudes requeridas.

El inventario Kuder es diferente del Strong en cuanto a su aproximación al problema. Así, aquél intenta detectar aspectos importantes de intereses vocacionales en un amplio rango de ocupaciones, en tanto que a este último le interesa establecer patrones de preferencias que permitan diferenciar una ocupación de otra y luego analizarlos para cada sujeto estudiado.

La confiabilidad y validez de estos inventarios son satisfactorias para los Estados Unidos, en donde se emplean frecuentemente. En México, como indicamos al principio, queda en duda el

que los coeficientes de confiabilidad y validez sean tan altos como en los Estados Unidos (0.85 para Kuder y 0.95 para Strong), ya que no hay, o al menos desconocemos que existan, estudios de este tipo.

Por otra parte, es importante tener en consideración la estabilidad de los intereses. Diversas investigaciones en este sentido han demostrado, sobre todo en niveles de estudios profesionales, que ocurren cambios en intereses y preferencias durante los años de pregrado; en cambio, parece haber una estabilidad muy importante después de que se ha obtenido el título profesional.

COMENTARIOS ACERCA DEL EMPLEO DE LOS TESTS DE HABILIDADES MÚLTIPLES Y DE HABILIDADES ESPECÍFICAS EN MÉXICO

En México, en cada una de las instituciones que hasta el presente han considerado la necesidad de hacer una selección y clasificación de personal mediante la instrumentación psicológica, se advierte la siguiente situación: o se utilizan pruebas psicológicas construidas para grupos no mexicanos, las cuales se han intentado adaptar, y se establecen sus propias normas para sus grupos respectivos, o sobre esos patrones se construyen pruebas que se han considerado apropiadas para medir las aptitudes especiales que supuestamente tendrán que exigírseles a las personas que soliciten desempeñar alguna de las ocupaciones que la institución o empresa puede ofrecerles. También, por lo general, cuando el psicólogo que se ocupa del trabajo en esta área llega a desempeñar sus funciones, la institución le indica lo que debe hacer, en vez de que sea él, como especialista, quien tenga la posibilidad de hacerlo. Cuando hemos preguntado quién ha proporcionado la batería que en tal o cual lugar se está utilizando y cuáles fueron los motivos que facilitaron su elección, no encontramos ninguna información ni explicación al respecto.

Hacemos hincapié en esta área porque es una de las más criticadas en el país por su forma de funcionamiento. Evidentemente, no podemos cerrar los ojos y declararnos ajenos a lo que ocurre, porque tenemos una gran responsabilidad al respecto. La intención al subrayar esto aquí es la de hacernos responsables de esas instituciones y tratar de modificarlas, mostrando nuestro conocimiento acerca de ello.

Definitivamente, como lo indicamos en la introducción, el hecho de que en este libro no aparezcan nombres de tests que se manejan en otros países y que en México no se conocen con ese nombre aunque hayan servido de “inspiración” parcial o total para construir aquí algunos, los cuales posteriormente mencionaremos en forma muy general puesto que no conocemos sus nombres y quizá no los tengan, no quiere decir que no exista en el mundo una gran cantidad de pruebas construidas para esos fines, sobre todo en aquellas sociedades muy industrializadas.

Los tests psicológicos para selección de trabajadores y empleados que más frecuentemente encontramos son formas pequeñas de tests que miden la comprensión de instrucciones, palabras, números o letras que se encuentran mal colocadas a lo largo del test para medir la rapidez perceptual, el manejo de las operaciones aritméticas simples, algunas pruebas acerca de la información general de las personas, pruebas de similitudes que lo requieren, pruebas de destreza manual, de coordinación visual, de coordinación visomotora, etcétera. Es difícil que se tenga integrada realmente una buena batería que abarque todos los niveles ocupacionales y mucho menos los profesionales que la organización contenga.

En la actualidad muchos psicólogos han hecho fructíferos esfuerzos en algunas empresas que les han permitido realizar investigaciones simultáneamente con las actividades prácticas propias de su ramo.

Por otra parte, en muy pocas empresas se utilizan pruebas expresamente construidas para sus propias necesidades, porque las circunstancias en las cuales se realiza el trabajo impiden, en general, utilizar tiempo, dinero y energía para administrar instrumentos de medición que implicaría una merma en el tiempo de los trabajadores.

MEDICIÓN DE LAS HABILIDADES ESPECIALES EN EL ÁMBITO ESCOLAR

En la psicología educacional parece haber un poco más de desarrollo en el área de los tests, ya que muchos de éstos han sido traídos a México, adaptados y estandarizados para los diversos niveles de escolaridad. Así, por ejemplo, se ha trabajado con tests de inteligencia general, de maduración perceptual, de formación de conceptos, con pruebas sensorio motrices, etcétera, y se han estandarizado escalas de inteligencia, como las de Wechsler, que bien pueden servir como medidas de inteligencia general o como medidas de aptitudes especiales (si sólo utilizamos subtest por subtest).

En muchas instituciones escolares se emplean las pruebas de admisión, las cuales, obviamente, se han construido en base a la información que se supone debe tener el examinado en el momento de ser probado para ingresar al nivel escolar inmediato superior. (Solamente como dato informativo, debemos hacer notar que la mayoría de estas pruebas no han sido construidas por psicólogos; por tanto, cabe dudar de su validez y de su confiabilidad).

5

LA MEDICIÓN PSICOLÓGICA APLICADA AL DESARROLLO DEL NIÑO EN SUS DIVERSAS ETAPAS: INFANTE, PREESCOLAR Y ESCOLAR

DESDE EL NACIMIENTO HASTA LOS QUINCE AÑOS

La comprensión del desarrollo de la conducta del niño es uno de los problemas a los que se enfrenta el psicólogo, ya que, como se sabe, está influido por una gran variedad de factores que determinan estilos específicos de comportamiento, los cuales sólo pueden entenderse, jerarquizarse o explicarse si se tienen en cuenta las circunstancias bajo las cuales tales procesos de desarrollo se realizan. Estas circunstancias o factores determinantes pueden resumirse en: *a)* factores genéticos y factores congénitos, que determinan las potencialidades biológicas con que nace el niño, y *b)* el medio en que éste nace, el cual influirá considerablemente en los aprendizajes que adquirirá en su interacción con su campo vital. Este medio hay que considerarlo desde el ambiente social y cultural en que se desarrolla, es decir, la cultura y la estructura social a que pertenece por raza o nacionalidad y su medio social inmediato, que puede ser desde el familiar y el escolar, hasta el de grupos de pertenencia por compatibilidad de intereses, creencias y valores que, desde luego, le fueron transmitidos por los progenitores y por la cultura.

Así, el desarrollo conductual del niño y de sus expresiones está determinado por la acción de los aspectos biológicos y naturales, y los efectos del medio externo en donde crece física y psicológicamente.

Se sabe que los rasgos físicos del individuo, en su mayor parte, dependen de la herencia. Algunas veces, durante el proceso del desarrollo fetal y en el momento del nacimiento, pueden ocurrir alteraciones que deformen de alguna manera y en grados diferentes esos rasgos. Sin embargo, también puede suceder que el individuo varíe en sus rasgos físicos a través del proceso de adaptación, el cual no es sólo biológico sino también social, psicológico y aún geográfico, nutricional, etcétera; por ejemplo, algunas deformidades anatómicas pueden ser de origen genético; otras, como algunos aspectos del desarrollo corporal, pueden ser producidas por factores ambientales.

Existe una íntima relación entre los procesos fisiológicos y los psicológicos, así como variaciones de esa interacción de sujeto a sujeto. Al respecto, es importante para el psicólogo saber hasta qué grado las diferencias del funcionamiento fisiológico y neurológico son de origen genético.

Se han llevado a cabo múltiples investigaciones acerca de este aspecto y se ha podido constatar que, en muchos casos, ciertos factores genéticos influyen en la predisposición para adquirir determinado tipo de alteraciones fisiológicas que tienen sus expresiones psicológicas en el comportamiento del individuo. De este modo, existen defectos en el funcionamiento intelectual debidos, por ejemplo, a una alteración de tipo hereditario que se localiza en las células nerviosas del encéfalo y de la médula espinal, que produce el retraso mental “amaurótico” infantil, debido a una inflamación de esas neuronas que se saturan de grasa y que producen ceguera, parálisis y deficiencia mental. Este tipo de padecimientos llevan al individuo a la muerte a muy temprana edad.

Otro tipo de alteraciones que repercuten en el funcionamiento intelectual es la llamada fenilcetonuria, que es la carencia, en el niño, de una enzima que resulta absolutamente necesaria para su metabolismo normal, y el sistema nervioso se lesiona por la acumulación de una sustancia tóxica que debería ser eliminada por la enzima mencionada.

En resumen, el papel que desempeñan los factores genéticos en las alteraciones del comportamiento del individuo, a pesar de ser muy discutido en el campo de la psiquiatría y de la fisiología, debe tenerse en consideración para entender con más claridad cómo se efectúan los procesos de evolución o desarrollo psicológico del niño y del individuo.

El desarrollo del niño puede dividirse en varias etapas, las cuales dependerán del punto de vista del que se intente partir. Desde luego, debe recordarse que cuando se habla de desarrollo físico del niño, se está haciendo referencia al crecimiento que normalmente se verifica en todo el organismo humano, con sus variaciones naturales a lo largo del tiempo. A este tipo de desarrollo se le denomina *crecimiento* y se refiere a todos aquellos cambios físicos, químicos y en general fisiológicos que son concomitantes, hipotéticamente, al desarrollo conductual. Conforme el sistema nervioso va desarrollándose, se va modificando y diferenciando la conducta; del mismo modo, conforme el cuerpo crece, evoluciona la conducta. Así, el crecimiento es un proceso de moldeamiento. Su detección permite determinar si el crecimiento y los fenómenos concomitantes a él son correlativos o se encuentran de alguna manera alterados. Por ejemplo, los reflejos primarios del niño, como el parpadeo, la prensión, el reflejo de la rótula y posteriormente el sentarse, pararse, arrastrarse, andar, etcétera, son formas de conducta, iguales a las reacciones oculares que ocurren cuando el niño percibe un objeto y lo sigue con la mirada, hasta donde sus capacidades lo permitan. La manera de asir los objetos con ambas manos o el estirarse a obtenerlos con una sola mano, son formas de conducta que denotan evolución y crecimiento en todas las esferas de actividad del individuo.

De acuerdo con Arnold Gesell y Amatudra, la conducta se expresa en cuatro aspectos fundamentales, que no son excluyentes, pero que tienen su momento oportuno de surgimiento y que perduran, aún sin estar aparentemente activos, a lo largo de la existencia humana. Esas cuatro áreas de entrenamiento son las siguientes: *a)* el área Sensoriomotriz; *b)* el área correspondiente a aquella conducta adaptativa; *c)* el área que corresponde al lenguaje, y *d)* el área llamada *personal-social*. Estos grupos de conducta estudiados por Gesell no implican un desligue o rompimiento, en un momento dado, entre unas y otras, ni un funcionamiento autónomo de cada una; por lo contrario, desde el principio del nacimiento tienen una íntima relación y permanecen así a lo largo de la vida, sólo que cada una de ellas va teniendo su momento principal durante el desarrollo desempeñando su papel correspondiente, sujeta a lo que el medio le exija al individuo en su interacción.

En un principio, lo único que se puede detectar en la evolución y desarrollo del individuo es la actividad preponderante del área o esfera Sensoriomotriz, y es importante porque denota la forma más o menos normal en que se va realizando la madurez fisiológica y psicológica del niño. A ella corresponden todas aquellas ejecuciones de movimientos corporales, desde los más globales hasta los más finos, que implican actividad muscular y que paulatinamente van coordinándose con un sinnúmero de procesos sensoriales y perceptuales que producirán la preponderancia de la siguiente etapa, a la cual se la ha llamado *adaptativa*. Esta consiste de la coordinación adecuada de los movimientos oculares y manuales, es decir, la coordinación visomotora, la cual permitirá al niño desarrollar su habilidad para emplear su dotación motora con el fin de solucionar problemas prácticos, así como desarrollar su capacidad para que en situaciones posteriores pueda resolver sus problemas, que irán en creciente dificultad de acuerdo con su edad.

En estas dos etapas del desarrollo del niño, como puede observarse, tienen lugar fundamentalmente actividades sensoriales, perceptuales y motoras que lo irán conduciendo paulatinamente al desarrollo de otro aspecto muy importante de la actividad humana, en el cual aparece como fundamental el lenguaje. La aparición de éste revela cómo ha ido organizándose el sistema nervioso central del niño. (Aquí cabe hacer notar que cuando se habla de lenguaje, se hace referencia no sólo al sentido que comúnmente se le da, que es el verbal, sino también a su más amplia concepción, que incluye toda forma de comunicación observable, como pueden ser gesticulaciones, expresiones motoras, balbuceos, formación de palabras u oraciones –a veces un tanto confusas –, conductas de imitación de expresiones de personas que se encuentran en su convivencia vital, más o menos cercanamente, etcétera.)

Uno de los principales problemas que se presentan en el campo de la psicología es la comprensión del desarrollo cognoscitivo del niño, que viene a hacerse patente cuando en un momento determinado del desarrollo surge el lenguaje. Los procesos anteriormente descritos, conforme van ocurriendo, le permiten al niño ir asignando nombres a los objetos que lo rodean e ir acumulando experiencias que pondrá en práctica cuando el medio se lo exija. Conforme crece en edad, el niño va siendo capaz de hacer una mayor diferenciación y precisión en la percepción de las características particulares de los objetos, la cual puede ser cualitativa o cuantitativa o una combinación de ambas, que le hacen diferenciar unos objetos de otros o encontrar las semejanzas que entre ellos existen, iniciándose así lo que se denomina *conceptos de conceptualización del medio*; es decir, empiezan las funciones de categorización, ordenamiento, valoración y jerarquización, etcétera, del niño, que le permitirán codificar y situar, en sus esferas conceptuales correspondientes, a todos los elementos con que interactúa en su ambiente.

La capacidad lingüística del niño se desarrolla a pasos no secuenciados, es decir, “a saltos”, de tal manera que sus habilidades cognoscitivas se ven afectadas radicalmente por esta forma de desarrollo.

Cuando el niño ha adquirido suficiente cantidad de lenguaje y puede utilizarlo con eficiencia para informarse, codificar y decodificar la información, entonces las palabras, como elementos primordiales del lenguaje, se convierten en mediadoras de sus ejecuciones.

Sus progresos en el ámbito escolar dependerán directamente de la adquisición de habilidades cognoscitivas de alto nivel. Fomentar tal desarrollo exige la investigación de la forma en que las condiciones familiares, sociales y, en general, educativas y de enseñanza escolar influyen en

su adiestramiento cognoscitivo. Un gran número de investigaciones al respecto han revelado que la calidad del ambiente, sobre todo en el aspecto lingüístico dentro del cual se desarrolla el niño en sus primeros años, es un factor externo de fundamental importancia que influye en el incremento del lenguaje: un ambiente estimulante en lo lingüístico, incrementa el aprendizaje del desarrollo cognoscitivo; lo contrario inhibe tal desarrollo.

El descubrimiento más importante sobre este punto es que el desarrollo del lenguaje está íntimamente relacionado con las diferencias de clases sociales. Al respecto, el sociólogo inglés Basil Bergstein, catedrático de la Universidad de Londres, en investigaciones relativas a las clases sociales y sus diferencias, hizo notar lo que él llama lenguaje “restringido” de la clase socialmente inferior y los códigos complejos de la clase media, poniendo de manifiesto que las madres que pertenecen a ese nivel social ejercen poca influencia sobre el niño para encaminarlo a expresar sus experiencias en forma verbal, de tal manera que inhiben el lenguaje, produciéndose bajos niveles de conceptualización y de diferenciación, llevándolo a que se interese solamente por los aspectos “concretos”, en el aquí y ahora, en lo directo, lo inmediato, lo global en su medio.

El lenguaje complejo es más individualista y específico, más diferenciado y preciso, además de que comunica una mayor variedad de conceptos y de contenidos cognoscitivos y afectivos que es capaz de distinguir y de expresar.

La aparición del lenguaje, en su más amplia concepción, marca una de las etapas de desarrollo más relevantes: la conjugación de las anteriormente mencionadas, misma que perdurará a lo largo de la vida del individuo, expresándose en diferentes modalidades, las cuales dependerán de la calidad y facilidad que el medio le haya proporcionado al individuo, expresándose en diferentes modalidades, las cuales dependerán de la calidad y facilidad que el medio le haya proporcionado al individuo. Esta etapa se halla determinada por el moldeamiento de la conducta influida considerablemente por factores intrínsecos del crecimiento. Por ejemplo, el medio habrá de exigirle al sujeto una serie de requisitos para que pueda interactuar socialmente dentro de su grupo de pertenencia. Así, él tendrá que aprender a controlar sus esfínteres, lo cual no sólo es una exigencia social, sino que primordialmente dependerá del desarrollo neuromotriz del niño. También le exigirá que aprenda a alimentarse, a asearse, a relacionarse con los demás, y a colaborar y reaccionar de acuerdo con las normas y convencionalismos sociales. Desde luego, esto dependerá de las variaciones individuales, pero siempre dentro de ciertos límites.

Todo lo indicado anteriormente conduce a hablar del desarrollo psicológico del niño a partir de ciertas dimensiones, tales como la adquisición de experiencias, la adquisición de la capacidad para emplear las experiencias, la capacidad para relacionar unas con otras, la habilidad para solucionar problemas mediante diferentes formas de respuesta, etcétera.

Contra todas las creencias antiguas, ahora se sabe que el niño inicia su aprendizaje desde el momento en que nace. Se sabe también que nace con una “carga potencial” que lo capacitará para aprender nuevas conductas y para modificar conductas reflejas conforme vaya efectuándose su maduración. Al nacer, el niño es capaz de establecer una relación entre un estímulo externo y una respuesta; al psicólogo sólo le atañe determinar qué estímulos son los que producen las respuestas y cuáles de éstas se adquieren más fácilmente.

Aquí es necesario mencionar la labor de un personaje muy importante en el campo de la psicología, sobre todo en lo relativo al estudio e investigación del desarrollo psicológico del niño; se trata de Jean Piaget, quien si bien en un principio no fue considerado como un investigador rigurosamente metodológico, por la forma en que eligió para realizar sus estudios, quizá por las dificultades con que tropezó, en la actualidad ha llevado sus investigaciones a conclusiones que pueden constituir ejemplos de los estudios sistemáticos y metodológicos que ha realizado.

Piaget divide las etapas del desarrollo psicológico del niño en forma más analítica que como lo hace Gesell. Piensa que la capacidad de adaptación del niño depende de su inteligencia y que atraviesa por una serie de etapas de maduración. Considera que existen dos grandes etapas del desarrollo intelectual; el desarrollo intelectual Sensoriomotriz (desde el nacimiento hasta los 2 años de edad hasta la madurez). En la primera etapa, el niño, para adaptarse, no utiliza un gran número de símbolos o de verbalizaciones, sino que casi toda su conducta es preverbal. En esta etapa, Piaget distingue varios periodos, determinados por lapsos de edades, aún cuando apunta de antemano que no todos los niños se desarrollan al mismo ritmo, aunque la sucesión de etapas sean en el mismo orden, pero no con la misma velocidad. Estos lapsos son seis, al cabo de los cuales su conducta se convierte definitivamente en propositiva; es capaz de reproducirse con la conducta de un modelo o con la mención de un objeto en su ausencia; es decir, mediante la memorización hecha, inicia una manipulación intencional de la realidad.

Al final de esta primera etapa y al principio de la siguiente, la continuación del desarrollo se caracteriza por el uso del lenguaje y por la capacidad del niño para operar con símbolos. De ahí en adelante continuará su desarrollo hasta la madurez, incrementando su capacidad de manipulación de esos elementos adquiridos, y seguirá aprendiendo continuamente. El autor considera que el individuo siempre está aprendiendo nuevas experiencias hasta que finalice su vida. Ahora la calidad de lo aprendido dependerá, por una parte y fundamentalmente, de las diferencias individuales, y por otra, de las facilidades y la calidad del medio en el cual estos procesos se llevan a cabo.

EL DESARROLLO FÍSICO

El desarrollo físico, con sus variaciones relativas, es mucho más rápido en los primeros meses de vida. Al llegar el niño a su segundo año de edad, disminuye esa rapidez, sin dejar de ser notables los cambios conductuales diarios que el niño va mostrando. Sus cambios físicos, desde los de estructura ósea, su desarrollo fisiológico, etcétera, le van permitiendo conducirse, cada día, de diferentes formas en un intento de adaptarse paulatinamente a su medio. Conforme esto va ocurriendo, sus capacidades de respuestas motoras se van ampliando. (No es necesario explicar aquí paso a paso cómo van ocurriendo estos cambios. Quienes estén interesados en estos problemas, pueden revisar la bibliografía correspondiente). Sin embargo, se considera como antecedente importante, antes de analizar el contenido que nos ocupa –o sea, el de la medición aplicada al desarrollo psicológico del niño -, hacer énfasis en un punto que es vital para que se puedan comprender con más claridad los resultados que arrojen en este campo los instrumentos de medición apropiados. Este aspecto fundamental es la socialización del niño.

Con la aparición del lenguaje verbalizado y con el desarrollo más o menos completo de su condición física, el niño va adquiriendo la capacidad de interactuar con el medio que lo rodea. Puede ya manipular objetos, dirigirse hacia ellos, intentar darles un nombre, iniciar sus procesos verbales de conceptualización –ya que éstos se han venido desarrollando en edades anteriores a

niveles preverbales -, empieza a ejercitar sus capacidades exploratorias, a resolver problemas en diferentes formas de las que utilizaba anteriormente; en resumen, “está aprendiendo”, sabe que tiene grandes limitaciones, pero que también, a veces, puede desenvolverse y solucionar problemas sin la ayuda de los mayores.

La socialización se refiere a la adopción y aprendizaje, por parte del individuo, de los valores, creencias y formas de percibir el mundo que son compartidos por un grupo específico, con el fin de conducirse como los demás esperan que él lo haga, al formar parte, como miembro responsable, de ese grupo; es decir, la socialización ocurre cuando el individuo se ha ajustado a los patrones de un nuevo grupo. De este modo, el niño está aprendiendo a ser un miembro aceptable de un grupo.

Conforme el niño va estableciendo contacto con mayor información, va aprendiendo a discriminar y a jerarquizar esa información en términos de su credibilidad y relevancia que represente para él.

Durante el tiempo en que el niño permanece bajo la manipulación y protección parental, ciertos factores van siendo importantes en la adquisición de las demandas de su cultura: los valores, las creencias y otros aspectos de la cultura van produciendo la emergencia de la personalidad del niño, que incluye no solamente los denominados rasgos o características particulares de cada individuo, sino también el desarrollo, utilización y ejecución de sus habilidades generales, la adquisición de motivaciones que lo impelen a actuar, el desarrollo de su vida afectiva, etcétera. De tal manera, la cultura es un mediador entre el niño y las relaciones con los miembros de su grupo, que en un principio es la familia, como elemento más importante; luego, otro tipo de interacciones, como las escolares, las de grupos de juego, las de los amigos, etcétera.

Al parecer, mientras una gran parte de las experiencias de la socialización de los niños sucede mediante el entrenamiento directo, otra más difusa, pero tan significativa como la anterior y a largo plazo, es adquirida por el proceso activo de imitación que el niño sigue de las formas de conducta de los modelos parentales, que en última instancia son los agentes primarios de socialización; por ejemplo, cuando a un niño se le castiga físicamente por haber golpeado a otro, la intención es que aprenda a controlar sus acciones futuras; pero al mismo tiempo está aprendiendo la conducta de hacer daño a otro niño, la cual expresará cuando la ocasión sea propicia.

La cultura, por supuesto, determina lo permitido y lo prohibido. Así, a los dos años de edad, el entrenamiento del niño para su sociabilización se reduce a la inhibición del comportamiento indeseable mediante la restricción, por parte de los padres, de las actividades intolerables para el “buen desarrollo” del niño, pero que para éste son bastante agradables. Es la época en que deberá irse alejando paulatinamente de la madre, lo cual lógicamente le producirá ansiedad y temor al sentir que repentinamente le falta la protección absoluta que antes tenía. Surge también la presencia de la agresión, misma que dependerá en su expresión de sus aprendizajes primarios.

La agresión, en sus diversos niveles, se expresa de diferentes maneras. Goodenough, en una de sus múltiples investigaciones, mostró la forma en que la ira y la agresión se van modificando, gracias a los aprendizajes en los primeros años de vida. Encontró las situaciones provocadoras y los factores del ambiente que producen esa conducta (como es natural, a los dos años de edad, esas expresiones se traducen en actividad motora, sin dirección). Conforme va desarrollándose, el niño va controlando estas expresiones, de tal manera que durante su vida puede ser capaz de eliminar las

respuestas originales, sustituyéndolas por otras más adecuadas y maduras, si fueron inhibidas a tiempo. Si esto no hubiera sido así, las respuestas se seguirán repitiendo al ser recompensadas por el medio.

En suma, como producto de la maduración y del aprendizaje –procesos que son continuos-, así como de la aplicación de ese aprendizaje en sus interacciones, el niño muestra habilidades motoras, de lenguaje y de cognición sorprendentes en su segundo año de vida. Con esos elementos, en adelante, al irlos desarrollando, irá siendo más capaz de enfrentarse, adaptarse y producir los estilos de comportamiento que se requieren para subsistir en el ambiente, logrando así el equilibrio organismo-ambiente indispensable para su supervivencia.

MEDICIÓN Y EVALUACIÓN DE LA EVOLUCIÓN DE LA CONDUCTA

Para poder detectar las características y formas que se van expresando a lo largo del desarrollo de la conducta del niño y para determinar si este desarrollo obedece a su crecimiento físico y fisiológico, es necesario establecer patrones o normas de comportamiento promedio en cada una de las etapas en que se ha dividido el desarrollo psicológico del niño para su estudio e investigación. La graduación de los patrones de maduración sirve como escala reguladora del curso promedio “normal” de la evolución conductual.

La maduración tiene su tiempo. No existe una edad absoluta en la cual se pueda decir que el sujeto debe conducirse estrictamente de determinada manera; sin embargo, se pueden jerarquizar los niveles y grados de desarrollo en función de los pasos ordenados y coherentes que sigue su madurez.

Como en todos los aspectos de la medición, cualquiera que sea el área o situación que se intente medir, si se sigue la metodología establecida para obtener mediciones objetivas válidas y confiables, en el caso de la medición del desarrollo psicológico del niño es necesario investigar en grandes grupos de niños de todas las edades, a fin de lograr la obtención de la media del comportamiento en cada uno de los niveles de desarrollo, teniendo en cuenta, desde luego, la existencia de la variabilidad individual que hace que la secuencia del desarrollo pueda retardarse en una amplitud más o menos pertinente, pero que de todas maneras es significativamente uniforme. El desarrollo conductual tiene una correlación altamente positiva con la edad, y está la tiene con la conducta. Dicho de otra manera, a través de las expresiones conductuales observables por uno mismo, se puede determinar la edad que un individuo manifiesta, a partir de la comparación con la conducta promedio establecida.

Aún cuando el desarrollo es un proceso continuo, como ya se ha indicado, se puede dividir en tres etapas globales fundamentalmente, a saber: *a)* la infancia, *b)* la etapa preescolar, y *c)* la etapa escolar. Esto es con el fin de poder indicarle al lector, con precisión y claridad, cómo es factible realizar mediciones y comprenderlas para cada situación del desarrollo, ya que presentan dificultades en relación con el tipo de instrumentación que ha de utilizarse en cada caso.

Cabe hacer notar nuevamente, como se manifestó en otros capítulos, que los instrumentos de medición que aquí se analizan son solamente aquellos que se están utilizando en México, muchos de ellos aún sin los estudios psicométricos debidos; además se hará una presentación de aquellos que

han sido contruidos o adaptados y estandarizados para México por el Instituto Nacional de Ciencias del Comportamiento y de la Actitud Pública (INCCAPAC), para el estudio del desarrollo psicológico del niño mexicano, especialmente del preescolar y de escolar. Esta investigación ha conducido a la estandarización y, por tanto, al establecimiento de normas de calificación para la evaluación del desarrollo, así como para estudios transversales del niño mexicano.

MEDICIÓN DEL DESARROLLO PSICOLÓGICO DEL INFANTE

La bibliografía mundial, en esta etapa del desarrollo psicológico, muestra un gran interés en el estudio del niño, desde el nacimiento hasta los 36 meses, momento en que, de acuerdo con los patrones tradicionalmente establecidos, se convierte en niño preescolar al ingresar al jardín de niños para su entrenamiento futuro que lo conducirá a su vida escolar.

En México, la psicología evolutiva ha sido poco estudiada como meta de investigación, y sólo se ha utilizado el conocimiento importado de otros países y aplicado al estudio del niño mexicano, muchas veces simplemente traduciendo al español las investigaciones realizadas, sin detenerse a analizar las implicaciones que pueda tener el describir el desarrollo del niño, valiéndose para ello de la medición y evaluación de éste a través de instrumentos producidos, validados y estandarizados en otras culturas, mismos que presentan un panorama, si no raquítrico (por tratarse de instrumentos procedentes de culturas supradesarrolladas), al menos distorsionado, puesto que tanto el crecimiento como el desarrollo de la conducta, como sistemáticamente se ha indicado en párrafos anteriores, variarán no sólo de cultura a cultura, sino también dentro de una misma cultura, dependiendo de diversos factores, tales como el ambiente urbano o rural, la localización geográfica, la situación económica de algunas zonas, etc., y presentan características peculiares (y, por tanto, variabilidad en el desarrollo del individuo en general). Esto no quiere decir que no existan factores comunes a la propia cultura, pero sí modalidades que obligan a los psicólogos que se dedican a este campo de trabajo a ser muy cuidadosos en la forma de evaluar los procesos evolutivos, a fin de no incurrir, como ha sucedido tantas veces, en los errores perjudiciales que distorsionan profundamente la imagen del mexicano, no sólo en el medio mexicano, sino también en el exterior.

Tal como se ha indicado en otros capítulos de este libro, se insiste en la recomendación a todo psicólogo para que en el campo profesional o en el de la investigación se prepare de tal manera que sea capaz de construir sus propios instrumentos o, cuando menos, que conozca tan profundamente como le sea posible la esencia de la medición psicológica, así como sus aplicaciones, sobre todo en el área específica de la evolución del individuo, ya que sus juicios, en un tiempo muy próximo, podrán ser útiles para la planificación de nuevas modalidades en el sistema educativo nacional

Al analizar el problema del uso y abuso de los tests psicológicos en la medición de los primeros años de vida del niño, cabe mencionar en primer lugar una prueba mundialmente difundida y magníficamente construida, cuyos fundamentos teóricos se comprueban o se validan al aplicarla al campo apropiado. Esta es la famosa prueba denominada *diagnóstico del desarrollo*, elaborada por Arnold Gesell y por C. Amatudra, la cual se divide en tres grupos de tests que corresponden: *a)* al desarrollo normal y anormal del niño; *b)* al desarrollo normal y anormal del preescolar, y *c)* al desarrollo normal y anormal del escolar. Es decir, este instrumento abarca un rango de edad desde el

nacimiento hasta los 15 años. En esta sección nos ocuparemos solamente de la primera parte de la prueba.

Durante mucho tiempo, Gesell y Amatudra se dedicaron a la investigación sobre el crecimiento normal en los cinco primeros años de vida. Durante más de veinte años estudiaron a una gran cantidad de niños normales y los observaron a partir del nacimiento, examinándolos cada mes y después a intervalos más amplios. En esas condiciones, manipulaban 25 niveles secuentes de edad, hasta llegar a los cinco años. Además se estudiaron, dentro de hospitales infantiles, muestras muy grandes de niños con desviaciones y problemas en su desarrollo, simultáneamente con el diagnóstico y tratamiento de sus alteraciones. Se compararon ambos grupos a fin de detectar cómo se efectuaba el desarrollo normal y cómo ocurría el anormal; qué diferencias existían en los grupos de niños normales, en cuanto a la velocidad con que se efectuaba su evolución, estableciendo el comportamiento promedio y sus límites inferior y superior de variabilidad; etcétera.

Gesell fundamenta su instrumento de medición, indicando que el crecimiento origina cambios progresivos en las estructuras y transformaciones que se encuentran estrechamente correlacionadas con las funciones psicológicas.

Brevemente descrita su postura teórica, Gesell postula que la conducta comienza a organizarse desde mucho antes del nacimiento y va de arriba hacia abajo (de la cabeza a los pies), y de los elementos más centrales hacia los periféricos. A groso modo, las pautas de desarrollo que este investigador considera son las siguientes:

- a)* En los tres primeros meses de vida, el niño desarrolla y controla sus doce músculos óculo-motrices.
- b)* En los tres meses siguientes controla los movimientos, hasta entonces difusos, de los músculos que mantienen firme la cabeza; mueve los brazos y además intenta alcanzar los objetos que percibe en su campo visual.
- c)* En los tres meses siguientes adquiere el dominio de sus manos, de su tronco, se sienta, toma los objetos, los escudriña y los cede a la persona que está cerca de él.
- d)* En los tres meses siguientes, es decir, en el cuarto estadio, su control muscular se extiende hasta las piernas y los pies, manejando con mayor firmeza los dedos índices y pulgar, con los cuales hurga y ase los objetos.
- e)* A los dos años camina, corre, articula palabras y frases, y controla los músculos de sus esfínteres rectal y urinario. Empieza a observarse en él un sentido primario de identidad personal y de posesión de cosas materiales.
- f)* A los tres años de edad habla, utilizando oraciones más o menos bien construidas y maneja las palabras para expresar sus pensamientos; asimismo, inicia su actitud de comprensión al ambiente, sometándose a las exigencias culturales de éste.

- g)* A los cuatro años de edad inicia la época de la continua pregunta acerca de todas aquellas cosas para las cuales requiere una explicación más clara y precisa. Ya es capaz de establecer semejanzas y, por consiguiente, de formular preconceptos y generalizaciones. Indica también sus primeros intentos de liberarse de la dependencia familiar.
- h)* En el quinto año de vida, su control motor ha madurado perfectamente. Se moviliza con gran facilidad y destreza; articula las palabras y las oraciones sin el matiz infantil que anteriormente utilizaba; muestra una retención de hechos, historias o cuentos suficientemente larga. Desea interactuar en actividades de todos tipos con sus compañeros; se preocupa por su indumentaria y ya es capaz de discriminar y exigir atenciones de cierto nivel social.

Esta descripción de los primeros cinco años del desarrollo del niño incluye, desde luego, las dos primeras etapas del desarrollo que se mencionaron al principio, es decir, la infantil y la preescolar.

El test del desarrollo infantil, de Gesell, cubre un rango de edades desde las cuatro semanas hasta los 36 meses, correspondientes a cuatro periodos de desarrollo o áreas de madurez, que son los siguientes: *a)* de las 4 a las 16 semanas corresponde el periodo supino; *b)* de las 28 a las 40 semanas es el periodo en el cual el niño es capaz de sentarse, realizando muchas de sus actividades en esa posición; *c)* de los 12 a los 18 meses es la etapa locomotriz, y *d)* de los 24 a los 36 meses es la etapa preparatoria para la entrada al jardín de niños, y el paso para convertirse de infante a preescolar.

El material que utiliza Gesell para extraer la información que requiere es muy variado: en las primeras semanas, el registro de la conducta es ejecutado directamente por el examinador, probando el control de la cabeza, la postura de los brazos y las manos, de los pies y de las piernas, la postura del cuerpo y el desarrollo progresivo de la prensión; el control postural de los ojos, la agudeza de la atención, las reacciones hacia los estímulos sociales, etcétera.

A pesar de que el recién nacido presenta conductas que recuerdan al feto, es tan rápida su organización que a las 16 semanas ya muestra conductas de adaptación al medio, por ejemplo, cuando el niño es levantado en brazos y expresa agrado, no solamente por la sensación en sí, sino también porque esto le permite observar el mundo en el cual ha de ir penetrando poco a poco.

Además de la actividad directa del examinador, que consiste en la manipulación del niño y en la observación de sus conductas, Gesell emplea una serie de objetos estímulo para provocar las conductas esperadas y para poder evaluar el nivel de desarrollo. Estos objetos son: un aro colgante que se coloca en un plano central del área visual del niño; una sonaja que, al igual que el aro, tendrá que ser movilizada del centro visual hacia los lados, a fin de detectar el incremento del desarrollo, y una campanilla; además, Gesell utilizó el mango de la sonaja para ejercitar la prensión.

Además de estos objetos, forman el resto del test cubos rojos de madera, de 2.5 cm cada uno, cuya función por su color, tamaño, forma, peso y textura provoca reacciones de prensión, fijación ocular, prensión palmar, prensión digital, etcétera; en una palabra, la forma geométrica del cubo permite diferenciar conductas que van desde la prensión gruesa hasta la más delicadas, o de la

palmar a la digital. Asimismo, van conduciendo a conductas más complicadas, tales como chupar los cubos, morderlos, observarlos, arrojarlos, tallarlos contra algún otro objeto o simplemente tirarlos o soltarlos.

Otro elemento de este instrumento es una esfera o pequeña pelota que representa un alto nivel estimulativo. Asimismo, hay una campana o campanilla con mango cilíndrico y cuerpo cónico; el mango es oscuro y el cuerpo es de metal plateado. Este estímulo producirá reacciones táctiles y auditivas, y por su especial construcción induce al niño a llevarlo de una mano a otra, a chupar el mango y a explorarlo con sus ojos y sus dedos. También hay una botella, en la cual el niño puede introducir fácilmente la pelota mencionada con anterioridad; además, hay una taza pequeña, una cinta, un espejo, una pelotita, una serie de tableros, unos de encastramiento, otros de similitudes, otros de completamiento, unos más de relación; láminas con dibujos, etcétera. En una palabra, este material es muy sencillo y cabe perfectamente en una pequeña caja.

Cada examen tiene un tiempo de duración de 10 a 20 minutos, explorando en forma ordenada las funciones que el niño es capaz de realizar a través de cada lapso de prueba. Realmente lo que aquí se ha explicado acerca de este instrumento es sumamente breve; habrá que consultar el libro y el manual del autor para conocer perfectamente la estructura, manipulación e interpretación de los resultados de la medición del desarrollo psicológico del niño en estas etapas.

Existen otras pruebas para la evaluación de esta etapa de desarrollo del niño, como es la Escala de Desarrollo de la Inteligencia, de Catell, la cual cubre un rango de edad desde los dos meses hasta los 30 meses. Esta escala se construyó como una extensión de la forma *L* del test de Stanford-Binet. Los reactivos del test están agrupados por niveles de edad que forman tres grupos, distribuidos de la siguiente manera: de los 2 a los 9 meses, de los 10 a los 20 y de los 21 a los 30.

El primer grupo contiene los siguientes ítems:

1. Atención a las voces.
2. Inspección del ambiente.
3. Seguir un arillo en movimiento horizontal.
4. Seguir los movimientos de una persona.
5. Balbuceos, o las siguientes alternativas:
 - a) seguir un arillo en movimiento vertical o
 - b) levantar la cabeza en posición prona (boca abajo).

En el segundo grupo están contenidos los siguientes elementos:

1. Descubrir juguetes.
2. Unir o juntar una taza y un cubo.
3. Intentar tomar el tercer cubo.
4. Atinar a colocar una cuchara dentro de una taza.
5. Empujar con los dedos unos pequeños palos sobre una tabla agujerada (alternativa: sacar la cuchara de la taza).

En el tercer grupo se piden las siguientes ejecuciones:

1. Diferenciar un puente de una torre.

2. Imitar dibujos de líneas y círculos.
3. Hacer girar una tabla con tres agujeros, tomada de la forma original de Binet.
4. Doblar papeles.
5. Identificar objetos de uso común, tomados también de la forma original de Stanford-Binet; o las siguientes alternativas:
 - a) identificar dibujos por su nombre, o
 - b) tener el concepto de alguno de ellos.

La forma de calificación de esta prueba es similar a la de Stanford-Binet (véase capítulo 3 de este libro) y los coeficientes de confiabilidad y validez son suficientemente altos en esta prueba, mismos que se obtuvieron de grupos de niños de las diversas edades que cubre esta etapa, para los más pequeños en sus propios domicilios, y para los demás en guarderías y jardines de niños de los llamados maternos.

Consideramos que, para los fines de este libro, es innecesario seguir describiendo este tipo de pruebas, que son casi desconocidas en México; sin embargo, cabe subrayar que en cualquier parte del mundo donde se tenga que evaluar esta etapa de desarrollo del niño o todas las demás etapas, las bases de construcción del instrumento no pueden variar significativamente. De este modo, se debe tener en consideración, para determinar la diferencia de instrumentación, la adaptación del contenido a la circunstancia particular de cada una de las culturas dentro de las cuales se intente llevar a cabo, mediante tests psicológicos, la evaluación de estas edades a fin de determinar el estado de desarrollo del niño con respecto a las funciones que se están midiendo, para predecir, hasta donde sea posible, su futuro desarrollo y su nivel intelectual. Esta segunda parte (es decir, el nivel intelectual) dependerá fundamentalmente de los datos que las mediciones futuras arrojen, ya que generalmente ocurre que en estas primeras etapas de la vida resulta un poco difícil determinar con exactitud la futura capacidad intelectual de un individuo, puesto que en su desarrollo psicológico se presentarán factores externos, y en ocasiones internos, que cambiarán de alguna manera nuestros pronósticos.

MEDICIÓN DEL DESARROLLO PSICOLÓGICO DEL PREESCOLAR

Conforme se va desarrollando la coordinación motora y el lenguaje del niño, su comportamiento se va modificando. Va adquiriendo nuevos hábitos; su riqueza perceptual y su ambiente social es cada vez mayor, lo cual determina que en esta época se presenten modificaciones importantes para su futuro, mismas que permiten observar la diversidad de conductas que en esta etapa ocurren. Es la época de la curiosidad por todo, desde la investigación –por parte del niño– de aspectos sexuales hasta la entonces dependencia, ahora en controversia con su necesidad de independencia; la expresión abierta de su agresión; sus deseos de lograr lo que se propone y dominarlo; sus temores, sus ansiedades, etcétera. Todas estas características y muchas otras más son factores importantes que han determinado que un gran número de estudiosos de este campo hayan consagrado sus obras teóricas y sus trabajos experimentales al estudio de estos problemas.

La combinación de la maduración, del aprendizaje en sus diversas formas y de la imitación de las conductas de los adultos como entrenamiento para llegar a ser “como ellos”, son factores que

determinan la complejidad de la conducta y las diferencias individuales durante los años preescolares.

Como se sabe, los rasgos de la personalidad pueden estudiarse desde las motivaciones fundamentales (conducta encubierta) o desde la conducta manifiesta u observable, pero ambos puntos de vista están íntimamente relacionados. Las diferencias individuales en todos los factores que se han mencionado, son notables entre los niños preescolares; no obstante, se ha observado que aquellos rasgos o características que permanecen constantes hasta los 5 años de edad, determinarán la base estructural de la conducta futura del individuo. Las influencias extra-familiares, como son la asistencia a guarderías o a jardines de niños, la interacción del niño con sus maestros, la relación e influencia con sus compañeros, sus aprendizajes previos del manejo de la competencia, su agresividad, etc., determinan el desarrollo de su personalidad en estas etapas, y si la situación en la cual se verifica no es propicia, se producirán adaptaciones que darán lugar a alteraciones de la conducta, con su consecuente problemática psicológica en los niños.

Al igual que en la etapa anterior, es necesario registrar cómo se van verificando los cambios en el comportamiento del niño, a fin de conocer y entender cómo, los factores que naturalmente influyen en este proceso, se comportan en el niño como individuo y como parte o elemento de un grupo. Este grupo puede ser visualizado en función del sexo, nivel socioeconómico, edad, grupos formados en la guardería o escuela a que concurre, etcétera.

A fin de poder evaluar el desarrollo en esta etapa, existen tests apropiados para ello. Hacemos hincapié en esta afirmación porque con mucha frecuencia se utilizan, en forma indiscriminada, instrumentos que han sido contruidos para otras edades, con lo cual los datos que se obtienen son perfectamente espurios.

Los instrumentos de medición psicológica para la evaluación de la etapa preescolar contienen más material verbal que los de las pruebas infantiles, pero, por supuesto, siguen conteniendo una carga suficientemente importante de material manipulable. Así, estos tests psicológicos son mixtos y al mismo tiempo valoran el desarrollo de la coordinación visomotora y visual, así como el desenvolvimiento de los procesos cognoscitivos cada vez más complejos y que, por tanto, obliga a realizar una medición más cuidadosa y analítica del fenómeno.

Desde luego, se mencionará la prueba de Gesell en su segunda etapa de medición, que es la de los niños preescolares y que, constituida en la misma forma en que lo hizo para la etapa infantil, representa un continuo en la detección del desarrollo del niño. Construida como se ha indicado anteriormente (es decir, con las características requeridas para examinar a niños preescolares), se ha empleado en México sin previos estudios metodológicos ni rigurosos, que den la confianza absoluta de que las mediciones hechas con este test son válidas, en la extensión en que deberían serlo. Sin embargo, si ha de seguir utilizándose, cabe hacer la recomendación de que el psicólogo dedicado a esta área de la psicología debe observar las reglas estrictas que la metodología psicométrica exige para emplear un instrumento de medición. No es deseo de la autora hacer aquí ninguna crítica específica a trabajos que se conocen, los cuales parecen no haber observado estas recomendaciones y a los que es menester que se les haga una revisión cuidadosa.

Otra prueba de uso común en México, que sólo se mencionará aquí brevemente, pues ya se ha analizado en el capítulo 3, es la parte que corresponde a la medición del test de Stanford-Binet

para niños de 3 a 5 años. A groso modo, esta parte de la prueba se divide de la siguiente manera: año III, año III-6, año IV, año IV-6, y año V; es decir, este test se aplica cada seis meses en los dos primeros años y después cada año. Como es lógico suponer, esto se debe a que en estas etapas del desarrollo los cambios son muy rápidos, lo cual no sucede en las etapas inferiores.

El contenido de los tests está dividido en subtests de la siguiente manera:

Año III

1. Ensartado de cuentas.
2. Vocabulario en imágenes.
3. Construcción de un puente con cubos.
4. Memoria de imágenes.
5. Copia de un círculo.
6. Repetición de tres cifras.
7. Prueba suplementaria: tablero excavado, tres figuras.

Año III-6

1. Obedecer órdenes sencillas.
2. Vocabulario en imágenes.
3. Comparar dos pequeñas barras de diferentes dimensiones.
4. Responder a una lámina que contiene un dibujo.
5. Identificación de objetos por su uso.
6. Prueba de comprensión I.
7. Prueba suplementaria: dibujar modelos X.

Año IV

1. Vocabulario en imágenes.
2. Nombrar objetos de memoria.
3. Completar el dibujo de un hombre.
4. Identificación de imágenes.
5. Discriminación de formas.
6. Prueba de comprensión II.
7. Prueba suplementaria: memoria de frases I.

Año IV-6

1. Comparaciones estéticas.
2. Repetición de cuatro cifras.
3. Semejanzas y diferencias en dibujos.
4. Identificación del material de que están hechos los objetos.
5. Ejecución de tres órdenes.
6. Analogías I.
7. Prueba suplementaria: identificación de imágenes.

Año V

1. Completar el dibujo de un hombre.
2. Doblar papel para hacer la figura de un triángulo.
3. Dar definiciones.

4. Copiar un cuadrado.
5. Memoria de frases II.
6. Contar cuatro objetos.
7. Prueba suplementaria: hacer un nudo con dos cordones y un lápiz.

Estos subtests corresponden a la forma *L* de la prueba Stanford-Binet. Como se sabe, esta prueba tiene dos formas, la *L* y la *M*, que son paralelas, lo que permite utilizarlas indistintamente, de acuerdo con las necesidades y objetivos de la evaluación. La forma *M* de los ítemes mencionados solamente difiere en su presentación formal más no en el contenido fundamental.

Existen múltiples tests de los denominados de habilidades especiales, contruidos específicamente para la evaluación de las funciones en proceso de desarrollo, a fin de poderlas analizar con mayor profundidad. Estas pueden ser pruebas de memoria –ya sea mediata o inmediata – o pruebas de capacidad perceptual – ya sea auditiva, visual, táctil, olfativa o gustativa -; ambas cubren diversos aspectos de cada una de estas funciones. Las más usuales son las de percepción auditiva y las de percepción visual, debido a que son las operaciones perceptuales más comunes en el individuo.

Como consecuencia de que en esta etapa del desarrollo hace su aparición el lenguaje, se ha llevado a cabo una amplia investigación sobre la adquisición, función y utilización del lenguaje en los niños, particularmente en el aspecto de la formación de conceptos que, en última instancia, es el proceso que va a dar la pauta de cómo ha sido, es y será el manejo de esta característica tan importante en la actividad humana.

Las teorías acerca del desarrollo del lenguaje han propiciado la construcción de instrumentos para su evaluación. A medida que la capacidad lingüística del niño avanza, sus capacidades cognoscitivas van modificándose ostensiblemente, convirtiéndose gradualmente en un medio de comunicación con los demás niños y con los adultos y en un instrumento que le permite organizar sus experiencias y establecer un orden del ambiente que lo rodea, al poder conocer y designar los objetos externos mediante la palabra, llegando de esta manera a formar conceptos. La formación de conceptos en el niño indica la manera como ha ido aprendiendo a manipular el mundo exterior.

Las pruebas que evalúan, en general, los procesos cognoscitivos son particularmente difíciles de manejar, sobre todo aquellas que se relacionan con la formación de conceptos, debido a que, por la naturaleza de la respuesta que se le pide al sujeto, establecer una cuantificación precisa implica la necesidad de manejar respuestas abiertas que obligan a la categorización de las mismas. En este sentido, resulta muy difícil llegar a obtener mediciones muy rigurosas y completas de lo que en realidad es la formación de conceptos, debido a que cada uno de los instrumentos que para estos fines se han construido, no abarcan en su totalidad los diversos aspectos sobre los cuales debe evaluarse este proceso.

Los tests psicológicos más usuales son los llamados de *clasificación de objetos*, mismos que pueden tener diversos contenidos, tales como objetos de uso común, figuras geométricas, materiales diversos, etc. Además, dentro del mismo contenido de la prueba y con los mismos objetos, se le puede pedir al sujeto que maneje diversos aspectos conceptuales, es decir, que clasifique a los

objetos por diferentes razones, características o atributos, lo cual indicará el conocimiento, ordenamiento y forma de utilizar los conceptos aprendidos.

Otra manera de abocarse a la investigación del proceso de formación de conceptos es la llamada de *estilos conceptuales*, que en cierta forma es similar a las anteriores, pero más específica y restringida en cuanto a sus objetivos.

Uno de los tests más utilizados en México, para ciertas situaciones, es el de Hanfmann-Kasanin, que está constituido por 22 bloques, los cuales contienen cinco colores, seis formas, dos alturas y dos anchuras. El problema que se le plantea al sujeto es el de agrupar en cuatro categorías los cuerpos geométricos: alto y ancho, plano y ancho, alto y angosto, y plano y angosto. El objetivo de este test es detectar con base en qué aspecto, el sujeto forma sus grupos conceptuales. La ejecución se analiza y se califica en función de la interpretación de la tarea, así como de la naturaleza de los intentos de solución y del descubrimiento de la solución correcta. En cada nivel de ejecución se distinguen tres jerarquías: primitiva, intermedia y conceptual, calificándose con 1, 2 y 3, respectivamente, con lo cual se trata de expresar cómo el sujeto se aproxima a la solución del problema, a su habilidad para conceptualizar y a su habilidad para verbalizar su ejecución. Este tipo de instrumento se ha empleado en la investigación cualitativa del problema, y los pesos asignados a la calificación no han sido experimentalmente determinados, sino que son valores escolares asignados arbitrariamente. Sin embargo, igual que otras pruebas de este tipo, proporciona al psicólogo-investigador una manera de observar la conducta en situaciones controladas, y con ella se obtiene información importante para adicionarla a otros datos psicológicos.

En México, este tipo de instrumento solamente se ha estudiado en forma metodológica y sistemática en el INCCAPAC, a través de algunos instrumentos que forman parte de la batería utilizada en la investigación sobre el desarrollo del escolar mexicano, pero sólo en forma colateral, ya que las muestras estudiadas no corresponden a sujetos preescolares. No obstante, se han llevado a cabo algunas investigaciones, especialmente con la prueba de estilo conceptual, de Jerome Kagan.

El Test de Estilo Conceptual, de Jerome Kagan, está basado en la teoría del mismo autor, quien durante más de treinta años ha realizado estudios sobre niños, a muchos de los cuales ha seguido estudiando hasta la edad adulta. Su propósito ha sido descubrir cuáles conductas del individuo permanecerán constantes durante su vida y cuáles irán desapareciendo, además de investigar cómo ocurre esto. Según Kagan, entre los 2 y los 5 años de edad, las diferencias conductuales en los niños son más bien aparentes, en tanto que de los 5 años en adelante irán apareciendo rasgos cuya persistencia será hasta la adultez.

Kagan afirma que en el periodo preescolar se presentan las siguientes características particulares:

- a) Rápido desarrollo de las capacidades verbales e intelectuales.
- b) Iniciación de la diferenciación de la curiosidad sexual.
- c) Identificación con los modelos paternos.
- d) Sumisión hacia las reglas o normas sociales, como consecuencia de su interacción social.
- e) Iniciación de las conductas defensivas, como respuesta a situaciones productoras de ansiedad.

En este periodo, el niño aprende muchas palabras y su uso en una forma más eficaz, produciendo un lenguaje más comprensible, mejor articulado y más apegado a las reglas sintácticas gramaticales, ya que empieza a emplear artículos, verbos, preposiciones y conjunciones, en vez de utilizar predominantemente sustantivos, como en etapas anteriores. Kagan considera que la motivación para que el niño use el lenguaje dependerá de la gratificación que éste le proporcione, lo cual redundará en una mayor verbalización.

A los 3 años de edad, el niño todavía utiliza una sola palabra para designar una serie de objetos que de algún modo tengan semejanza estructural para él. Heinz Werner describe la forma como usa el lenguaje el niño, como “indiferenciada y sincrética”. Así, cuando una palabra o frase expresa la combinación o fusión de varios objetos, la llama sincrética; por ejemplo, con la palabra “comer”, el niño puede dar a entender que quiere comida, la acción de comer o su necesidad de ser alimentado. Werner se refiere al lenguaje indiferenciado como aquel que con una sola palabra, el niño identifica varias cosas similares; por ejemplo, la palabra “perro” significa para el niño cualquier animal u objeto que se mueva o que tenga cuatro patas.

El desarrollo conceptual generalmente se mide mediante el número de conceptos abstractos de que es capaz de emitir el niño. De este modo, entre los 3 y los 4 años de edad, los conceptos del niño son expresados en función de sus acciones.

La teoría de Kagan es muy interesante, pero sumamente extensa; aquí solamente nos interesa fundamentar uno de los instrumentos que construyó, a fin de comprender las diferencias cognoscitivas entre los niños, en esta etapa y en las subsecuentes.

Para Kagan, el test de estilo conceptual es el instrumento adecuado para explorar el “modo preferido, estable e individual de categorizar y organizar las percepciones del ambiente”. Este investigador se ha interesado en encontrar los procesos fundamentales de un “estilo analítico” de conceptualización, porque supone que para ello es necesario llevar a cabo un análisis minucioso y propositivo, así como la discriminación de los atributos relevantes y los no relevantes de un estímulo, ya que la capacidad discriminativa más alta producirá conceptualizaciones de mayor calidad, que determinarán la eficiencia en diversas clases de actividades cognoscitivas. Por esta razón, Kagan se ha interesado en estudiar aquellos factores que predisponen al niño a desarrollar y mantener una conducta conceptual analítica.

El test de estilo conceptual califica tres categorías formales, a saber:

1. La analítica-descriptiva, que considera conceptos basados en elementos observables que forman parte del estímulo total.
2. La inferencial-categoría, cuyas bases radican en juicios inferenciales relacionados con los estímulos agrupados.
3. La relación que, como su nombre lo indica, se basa en una relación funcional entre los estímulos, misma que puede darse en el tiempo y en el espacio entre los objetos.

El test original consta de 44 ítems, cada uno de los cuales está constituido por una tríada de estímulos. Cada lámina consta de tres dibujos en blanco y negro, de objetos familiares para el niño. La forma actual con la cual no sólo ha trabajado Kagan, sino que también se ha utilizado en los estudios realizados en México, consta de 19 láminas, que fueron elaboradas de tal manera, para que

se produzcan las tres clases de conceptos fundamentales para Kagan: analíticos, relacionales e inferenciales.

La tarea que ha de realizar el sujeto con este instrumento es la de seleccionar dos de los tres estímulos “que se parezcan o que deban ir juntos por alguna razón”, indicándosele, desde el momento de darle las instrucciones, que al hacer su elección deberá dar la razón de la misma y que no hay respuestas incorrectas. En este test es importante registrar el tiempo de reacción, es decir, el tiempo transcurrido entre el estímulo y la respuesta del niño.

Entre los estudios realizados en México con este test en la etapa preescolar se puede mencionar el de la psicóloga Sara Rallo Llagostera, quien trabajó con una muestra de 128 niños asistentes a la guardería y jardín de niños de la Secretaría de Comunicaciones. Realizó este estudio en 1967, y los resultados reportados, en forma sumaria, son los siguientes: al comparar la prueba de estilo conceptual de Kagan con la escala de inteligencia de Stanford-Binet, revisión 1937, forma L, correspondiente a esta etapa de desarrollo, Rallo Llagostera encontró que:

- a) No existen relaciones estadísticamente significativas entre el número de respuestas relacionales y el cociente intelectual.
- b) No existen relaciones estadísticamente significativas entre el número de respuestas analíticas y el cociente intelectual.
- c) No existen relaciones estadísticamente significativas entre el número de respuestas inferenciales y el cociente intelectual.
- d) No existen relaciones estadísticamente significativas entre el número de respuestas de aquellas categorías que representan fallas conceptuales y el cociente intelectual.

En este estudio, Rallo Llagostera llega a las siguientes conclusiones:

- a) El instrumento es adecuado para detectar índices de desarrollo;
- b) La *atención selectiva*, productora de respuestas analíticas, no se incrementa en el rango de edades estudiadas en los hombres, pero sí en las mujeres.
- c) Las respuestas funcionales y las inferenciales presentan diferencias entre ambos grupos;
- d) Debido probablemente a la edad, se presentan respuestas de distorsión perceptual u omisiones de respuesta;
- e) El tiempo de reacción es ostensiblemente detectable, diferenciando el grupo de hombres y el de mujeres; y
- f) El cociente intelectual guarda cierto grado de relación con las variables investigadas.

El tiempo de reacción es un indicador valioso para el psicólogo, en relación con el control de impulsos. En México, Lara y Krinsky han estudiado este fenómeno, relacionándolo con el proceso de socialización.

Desde el punto de vista de la medición del desarrollo perceptual, puede mencionarse una prueba que se emplea, como muchas otras, indiscriminadamente, a la cual se le dan diversas interpretaciones, dependiendo del marco de referencia del cual se parta y sin tener en consideración

los fundamentos psicométricos sobre los que descansa la construcción de este instrumento. Se trata del test gestáltico visomotor, diseñado por L. Bender.

Este test está constituido por nueve figuras, presentadas una en cada lámina y que han de ser reproducidas por el sujeto, sin límite de tiempo. Tiene como objetivo principal la evaluación del desarrollo perceptual y del funcionamiento visomotor. Las figuras fueron elaboradas por Max Wertheimer, uno de los principales pilares de la teoría de la gestalt para sus experimentos sobre percepción.

El principio que sostiene el test de Bender es el de que las unidades estructurales como un total son las formas primarias de percepción en el ser humano.

La conducta perceptual es considerada como la recepción sensorial de las formas, la interpretación de los niveles del sistema nervioso central y la ejecución motora.

El proceso total de percepción y reproducción puede verse afectado por diversos factores, tales como alteraciones neuronales, desajustes emocionales y variaciones en los niveles de ejecución intelectual. Además, Bender exploró la capacidad de su instrumento para investigar las “funciones gestálticas” en algunos casos patológicos, como la afasia, la lesión orgánico-cerebral, la esquizofrenia, la psicosis maniacodepresiva, las deficiencias mentales y el desarrollo perceptual de niños normales.

Al analizar la conducta visomotora del niño desde el punto de vista de su maduración, Bender observó que sus primeros intentos al utilizar papel y lápiz eran simples garabatos, quizá en un deseo productor de satisfacción de emplear su actividad motora. Así, el producto de la prueba pone de manifiesto cómo el niño percibe y organiza el estímulo, qué significado tiene para él, dónde y por qué tiende a simplificarlo o qué elementos le agrega. En edades tempranas los preceptos no están organizados y el significado que el niño le da a cualquier ejecución es arbitraria.

En sus investigaciones, Bender demuestra que el niño en esta etapa, aún emplea su primitivo garabateo, que es la “espiral cerrada”. Al respecto, Bender afirma: “Los esquemas visomotores más primitivos dependen del principio del movimiento constante, que se acomoda generalmente a un movimiento en espiral, en el mismo sentido o en sentido contrario al de las manecillas del reloj, con un componente direccional constituido por radios y una tendencia a insistir en los planos horizontales. Los puntos fijos resultan difíciles de trazar y la línea recta no es tanto la distancia mínima entre dos puntos, como la expresión de las tendencias radiales.”

Esto es importante porque los aspectos del desarrollo de las reacciones visomotoras solo podrán ser calificados y valorados a través de los resultados del test, si se entiende perfectamente la base perceptual de la cual parte la autora.

La forma de resolver cualquier problema no depende solamente del grado de desarrollo biológico al cual ha llegado el individuo en su actividad visomotora, sino que también influyen esquemas o patrones de comportamiento aprendidos, sobre todo aquellos en los cuales ha sido facilitada o restringida la actividad motora.

Uno de los problemas que la autora ha encontrado a este test es la forma en que los resultados son interpretados por Bender, ya que a pesar de que sus estudios pudieron conducirla a establecer el comportamiento promedio en el desarrollo perceptual en el rango de edades estudiado, la interpretación que le da a la ejecución no permite tener seguridad de que exista una correlación positiva entre los rasgos que presentan los dibujos y la interpretación que, a manera de equivalencia, intentó darle.

Sin embargo, otros autores han llevado a cabo la tarea de cuantificar la ejecución de cada uno de los dibujos, a fin de poder establecer una medida más objetiva, válida y confiable; por ejemplo, Elizabeth M. Koppitz elaboró un sistema de calificación del desarrollo con la prueba de Bender, que cubre un rango de edades de los 5 a los 10 años, manipulando las variables inteligencia, funcionamiento neurológico y ajuste emocional (véase Koppitz, E. M., *The Bender Gestalt Test for Young Children*. Nueva York: Grune and Straton, 1964).

Existen otros intentos de codificación diferentes, pero solamente se menciona éste por ser el más actualizado y porque en su validación se emplearon los tests de Binet y de Wisc para establecer correlaciones que resultaron estadísticamente significativas a la inversa. Estas correlaciones negativas representan el número de errores que ocurren en la prueba de Bender al ser comparada con otros resultados.

Por su parte, Koppitz llega a la conclusión de que para niños de 5 a 10 años de edad, su sistema de calificación puede ser utilizado como una medida de lectura, como un predictor de logro escolar y como un detector de problemas de aprendizaje, de daño neurológico y de detección de retraso mental.

CONSIDERACIONES GENERALES

No es necesario insistir en mencionar y describir instrumentos para la evaluación del desarrollo preescolar en sujetos mexicanos que si, de hecho, no son inútiles, tampoco permiten establecer afirmaciones válidas, ya que han sido construidos para la medición del proceso de desarrollo en otras culturas.

Como se explicó en la introducción, es más importante para el psicólogo mexicano obtener información acerca de lo que en México se ha hecho al respecto. Hasta hace muy poco tiempo, los psicólogos mexicanos han adoptado una actitud receptiva y han aceptado como definitivos los descubrimientos y las aseveraciones que investigadores extranjeros han decidido generalizar, no solamente en la investigación sobre niños preescolares, sino también en todo lo que se refiere a los fenómenos psicológicos que, como ha podido observarse, han sido valorados a partir de sus propios parámetros.

En 1973, en el Instituto Nacional de Ciencias del Comportamiento y de la Actitud Pública (INCCAPAC) se llevaron a cabo una serie de investigaciones en el campo de la psicología evolutiva, educativa y social especialmente, que han determinado que en ellas se dedique especial atención a los procesos de desarrollo en la etapa preescolar, sobre la cual se habían realizado muy pocos estudios hasta entonces. Estas investigaciones han conducido a la construcción de tests especiales para medir tanto el desarrollo perceptual como la evolución de los procesos cognoscitivos y motores,

a partir de un marco de referencia sociocultural que permita detectar los “estilos” o “estrategias” que aprende el sujeto mexicano para su supervivencia, dado su ambiente general.

Construir instrumentos de medición psicológica teniendo en consideración las particulares circunstancias del medio mexicano, o adaptar y estandarizar instrumentos dentro de las reglas rigurosamente establecidas por la psicometría, que después de analizarlos cuidadosamente permitirán obtener un conocimiento –y, por tanto, una mayor comprensión del cómo y el por qué de nuestro particular comportamiento –, es la razón por la cual el grupo de investigadores del INCCAPAC se ha trazado esta tarea.

Hasta el momento se han construido 17 tests psicológicos que miden el rendimiento intelectual, la percepción, las habilidades específicas y el desarrollo del lenguaje, en los cuales se enfatiza fundamentalmente la formación de conceptos. Estos tests psicológicos para preescolares ya han sido evaluados mediante “estudios piloto”, que incluyen análisis de ítemes, consistencia interna, coeficiente de estabilidad temporal (confiabilidad), etc.

En primer lugar se mencionarán los nombres con los cuales se conocen esos instrumentos de medición, y a continuación se hará una breve descripción de sus objetivos.

Para la investigación sobre rendimiento intelectual se han construido los siguientes tests:

1. Test de conocimientos generales.
2. Test de números.
3. Test de relaciones.

Para la investigación sobre percepción:

1. Test de habilidad para clasificar.
2. Test de clasificación de habilidades.
3. Test de comparación.
4. Test de eliminación.
5. Test de figuras ocultas.

Para la evaluación del desarrollo del lenguaje:

1. Test de comprensión oral.
2. Test de designación de objetos.
3. Test de reconocimiento.
4. Test de memorización de oraciones.
5. Test de comprensión de oraciones.
6. Test de letras.
7. Test de clasificación de objetos para niños preescolares.
8. Test de memorización de objetos.
9. Test de creatividad verbal.

Test de conocimientos generales

Esta prueba fue traducida y adaptada de una similar, desarrollada por el Educational Testing Service para la primera evaluación del programa educativo *Sesame Street* y empleada en México para los mismos fines durante el desarrollo del mismo programa en México, que lleva el nombre de

“Plaza Sésamo”. El test está constituido por 37 ítemes que cubren un rango de calificaciones de 0 a 37. Estos ítemes forman una serie de preguntas sencillas, tales como: “¿Cuál es el nombre de diferentes partes del cuerpo?”, “¿con qué piensas?”, “¿con qué vemos?”, etcétera.

Esta prueba tiene la peculiaridad, como la mayoría de las que se enunciarán, de ir cambiando el tipo de respuesta del niño, con el fin de evitar el cansancio y la monotonía que puedan alterar los resultados del test. A veces se le pide al sujeto que indique “qué es esto”; otras “con qué”, etc.

Test de números

Traducida y adaptada igual que la anterior y con los mismos objetivos, contiene 60 ítemes, distribuidos en las siguientes áreas: contar, reconocer números, nombrar números y encontrar un número específico entre cuatro que se presentan. El rango de calificación es de 0 a 60.

Test de relaciones

Traducido y adaptado para México, igual que los dos anteriores, está formado de 11 ítemes que tienen como fin evaluar la capacidad del niño para relacionar objetos por diversas razones y, además, registrar el manejo que el niño hace de las preposiciones. Un ejemplo de este test se muestra en la figura 4 y figura 4-a.

Como puede observarse en el ítem uno, varios estímulos están en posición y se le pide al sujeto que solamente emita dos respuestas por tetrada. En este ejemplo tiene que decir cuándo el perro está sobre la barda y cuándo está atravesándola.

En el siguiente ejemplo (véase figura 5) se le pide al niño que diga cuál es el oso más grande y cuál el más pequeño, y así sucesivamente en los 11 ítemes. El rango de calificación es de 0 a 22.

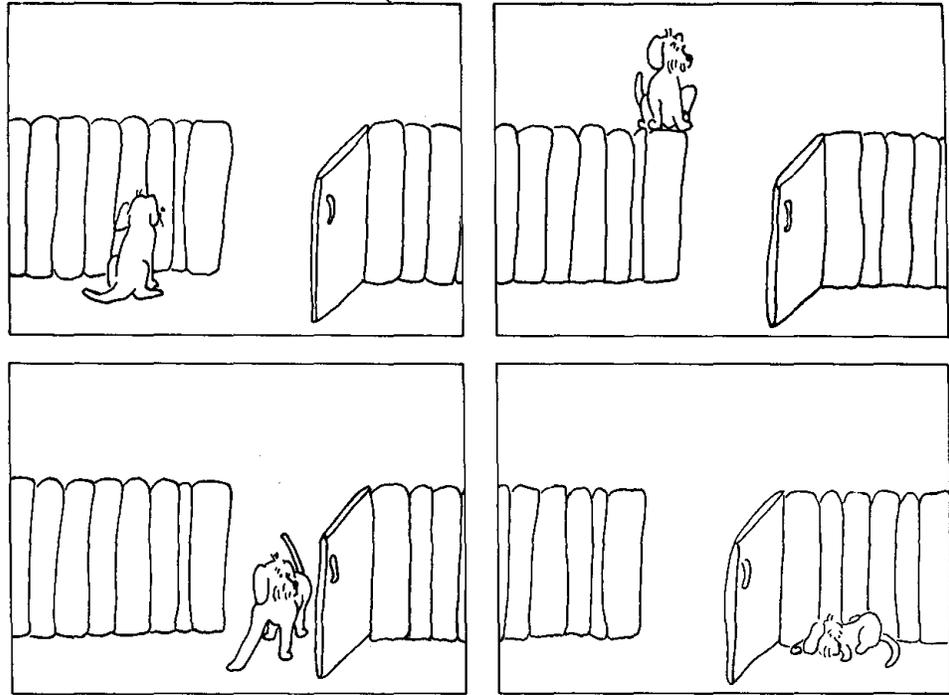


Figura 4. Ejemplo del test de relaciones.

D

B C D A

Figura 4-A.

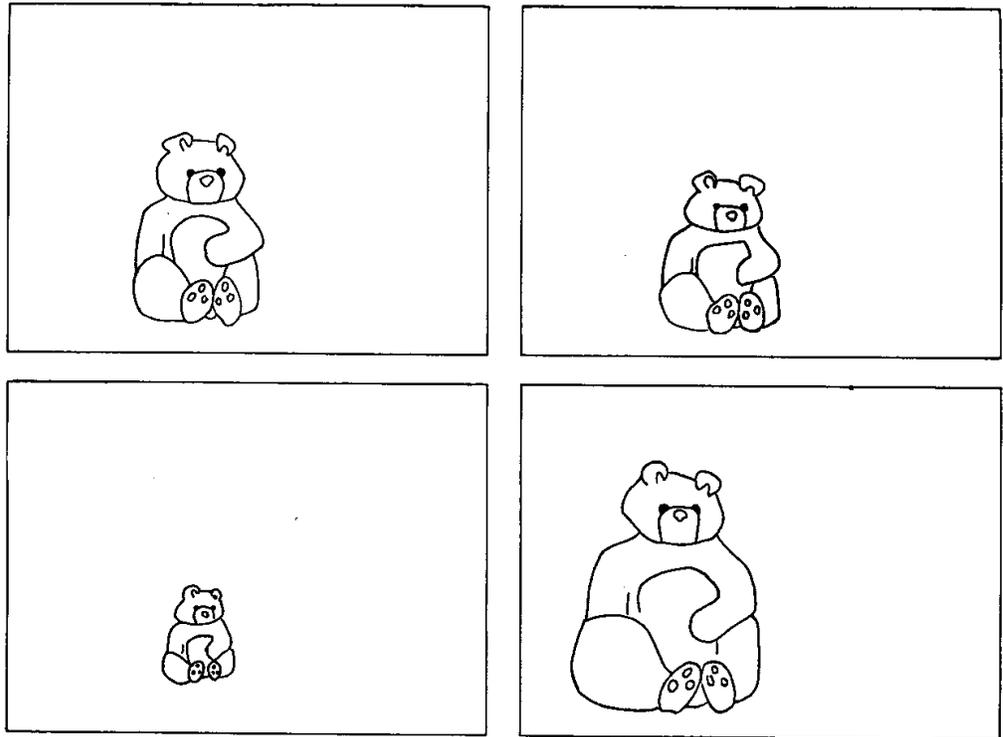


Figura 5. Ejemplo del test de relaciones.

Test de figuras ocultas

Este test, construido para escolares por M. Cates y adaptado sobre bases teóricas para niños preescolares de la prueba de German Witkin, no requiere traducción, por la naturaleza de los ítems. Consiste de 24 láminas que contienen diseños geométricos complejos, en los cuales el niño deberá buscar una figura simple que se le muestra previamente y que deberá encontrar en la misma posición, forma y tamaño de aquélla. Esta prueba mide la habilidad del sujeto para identificar preceptos específicos dentro de un contexto.

Las bases teóricas de este instrumento se derivan de la teoría del doctor Witkin, sobre la forma en que se realiza el desarrollo perceptual discriminativo en el individuo. Esta teoría es la denominada *teoría de la dependencia-independencia del campo*, que se relaciona fundamentalmente con el problema de la diferenciación psicológica, estudiado de manera exhaustiva por Witkin.

En este test de figuras ocultas se le permite al sujeto hacer dos ensayos para encontrar la figura y dispone de un tiempo de ejecución de 30 segundos, teniendo a la vista tanto la figura simple como la compleja. El rango de calificaciones va de 0 a 24. (Véase figura 6).

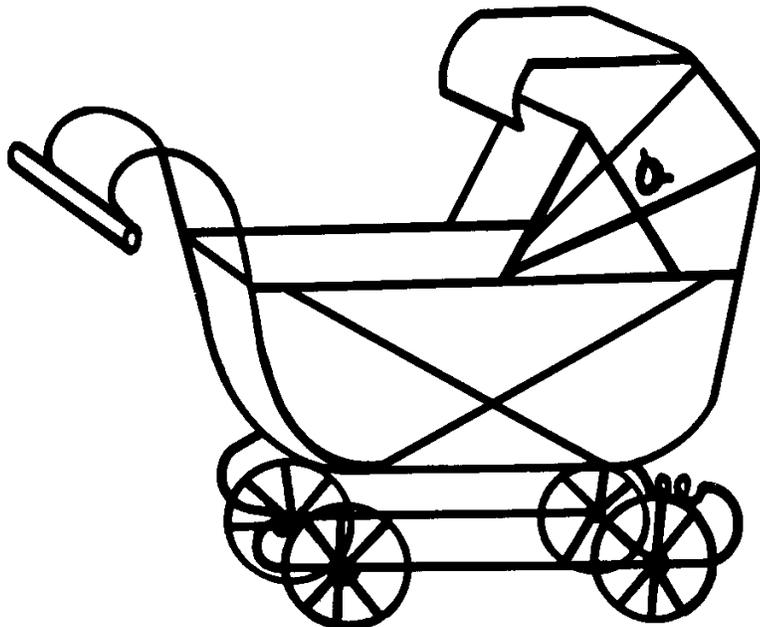


Figura 6. Ejemplo del test de figuras ocultas.

Test de habilidad para clasificar

También traducida y adaptada para utilizarla en el programa de “Plaza Sésamo”, contiene 16 ítems, y el rango de calificación es de 0 a 16. El contenido de este test se le presenta al niño en tarjetas que comprenden cuatro figuras, de las cuales se le pide que indique aquella que sea diferente de las otras tres. En la figura 7 se puede advertir cómo se presentan las tarjetas y cuál será la respuesta esperada por parte del sujeto al recibir las instrucciones.

Test de clasificación de habilidades

También adaptado y traducido para “Plaza Sésamo” este test está constituido por 24 ítems. Cada ítem consta de dos tarjetas, una de las cuales contiene tres objetos que definen una clase específica, y la otra cuatro objetos, de los cuales sólo uno puede ser clasificado correctamente con los que contiene la otra tarjeta. En este test se le pide al niño que diga cuál de los cuatro objetos de la segunda tarjeta puede incluirse con los tres de la primera tarjeta. (Véase figura 8).

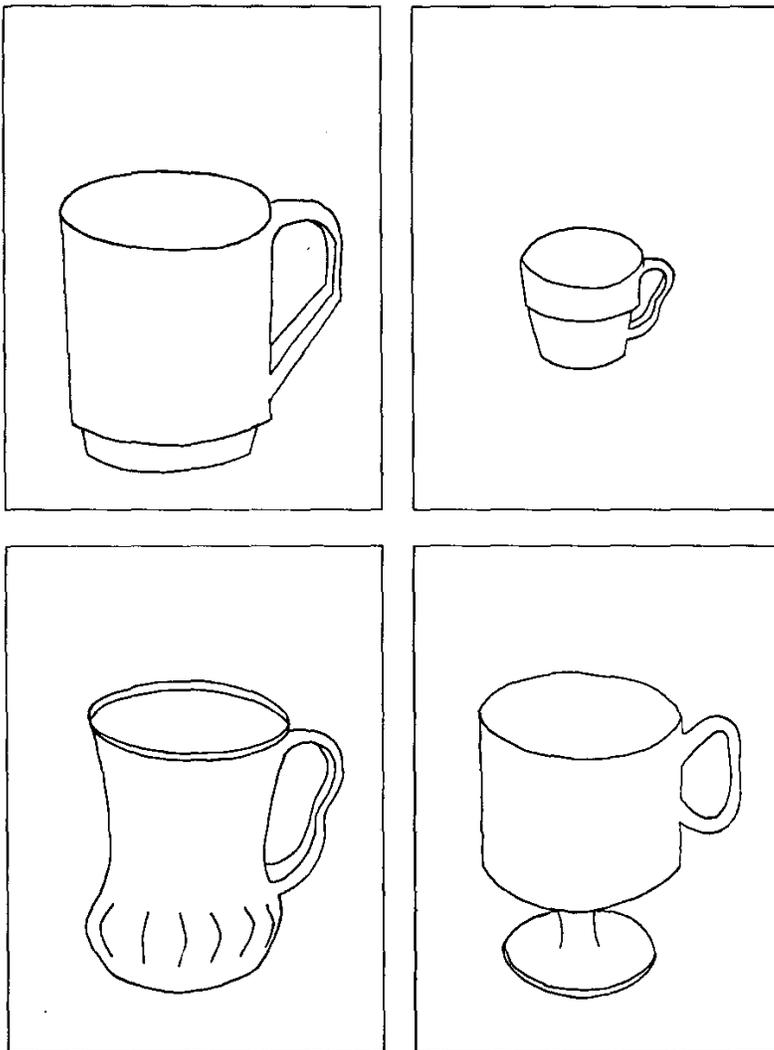


Figura 7. Ejemplo del test de habilidad.

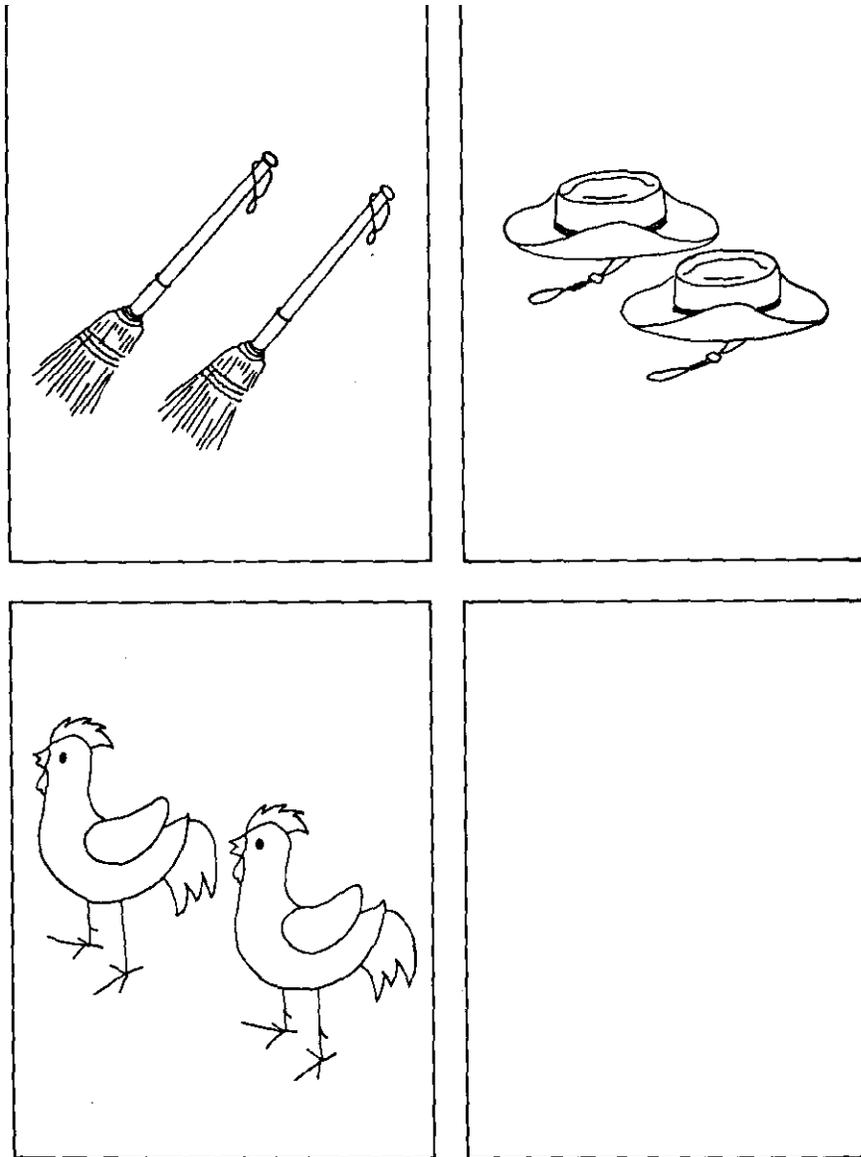


Figura 8. Ejemplo del test de clasificación de habilidades.

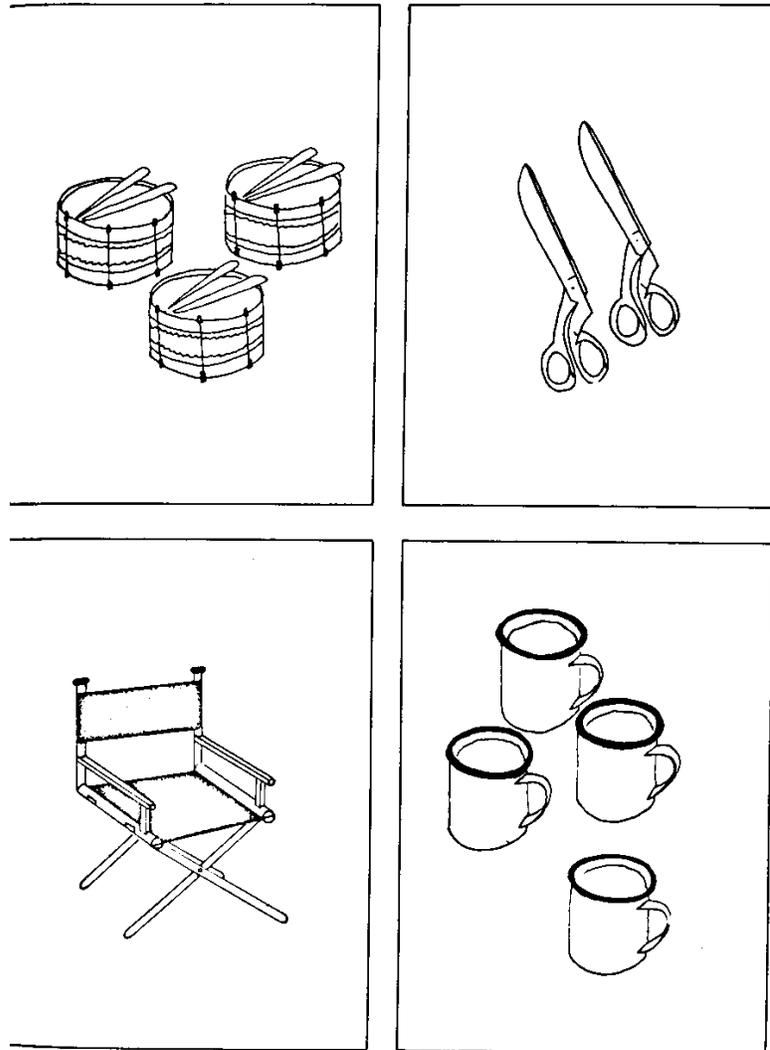


Figura 8. (Continuación).

Test de comparación

Este test fue construido por Robert Klein y sus colaboradores, para aplicarlo a niños preescolares en Guatemala. Contiene dos ejemplos y 18 ítemes. En esta prueba, la tarea consiste en encontrar, de entre cuatro figuras, aquella que sea exactamente igual al estímulo que de manera simultánea se le presenta al niño, para que éste establezca la comparación mirando ambas partes de la prueba. La confiabilidad de este test para Guatemala y México es de 0.60. En el INCCAPAC, y con la ayuda del doctor Choynowsky, se han desarrollado nuevos ítemes con el fin de aumentar la longitud de esta prueba. (Véase figura 9).

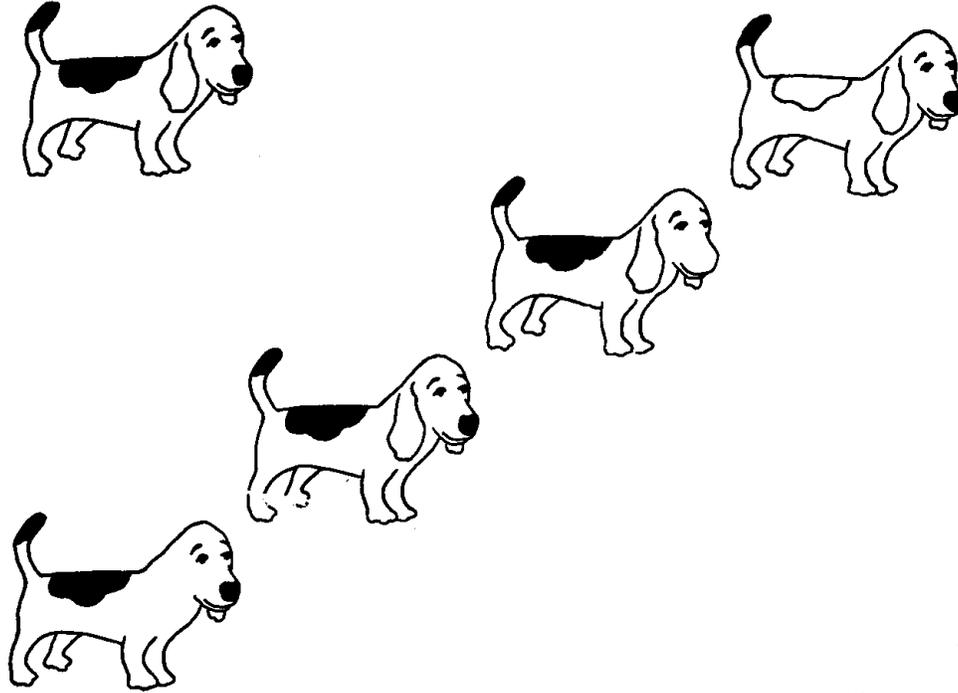


Figura 9. Ejemplo del test de comparación.

Test de eliminación

Este test consta de 25 ítems, cada uno de los cuales contiene cinco estímulos diferentes; aquí la tarea del niño consiste en encontrar cuál de estos estímulos no pertenece a ese grupo de objetos, sin que el examinador le pida ninguna explicación para eliminarlo del grupo.

Esta prueba explora la formación de conceptos; pero está considerado dentro del grupo de pruebas de percepción, porque es a través de ésta, en cuanto a color, tamaño y clase, como se le piden al niño las respuestas. El coeficiente de estabilidad temporal para México es de 0.68, y la intención de su empleo en esta investigación es la de encontrar la medida de algunos aspectos que el programa “Plaza Sésamo” enseña (Véase figura 10).

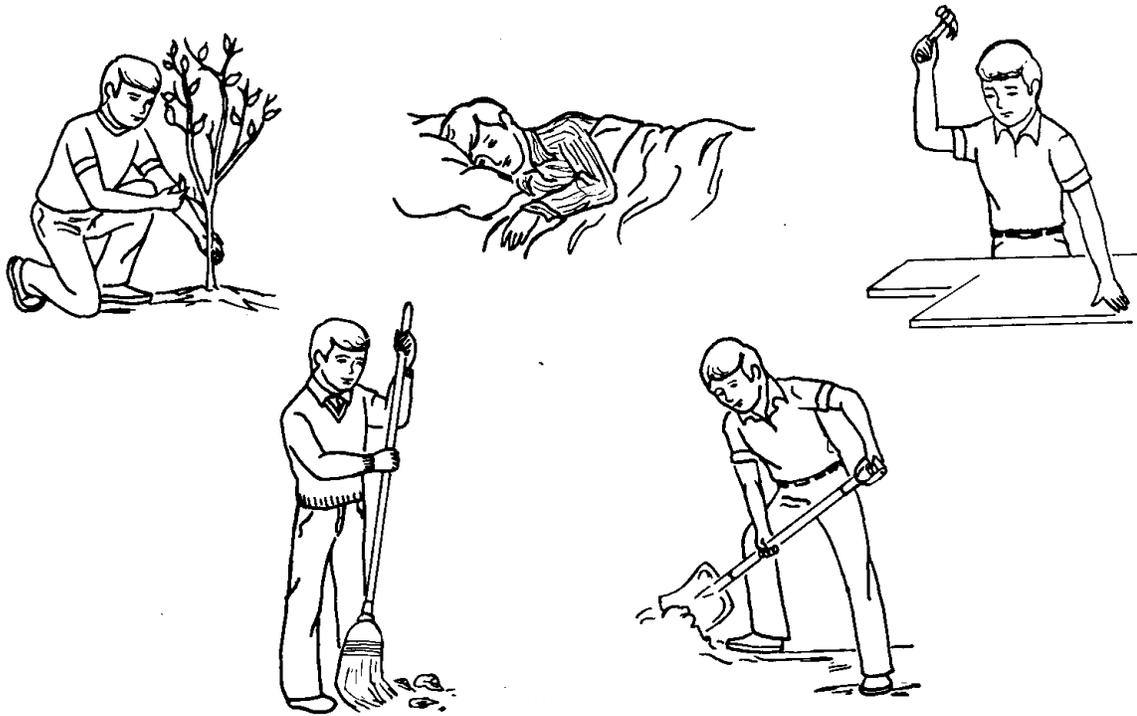


Figura 10. Ejemplo del test de eliminación.

Test de comprensión oral

Este test fue estandarizado por Herschel T. Manuel para niños preescolares, presentando una confiabilidad promedio de 0.51, obtenida de la siguiente manera: a los niños de 4 años de edad, después de seis meses de haberseles administrado el test, se les aplicó el retest y mostraron diferencias a lo largo de las edades en donde la correlación va de 0.22 hasta 0.74, a través de los tres niveles de edad.

En el CIPAC, se llevó a cabo el mismo trabajo, encontrándose diferencias significativas durante el pretest, el retest y el postest. También se hallaron importantes correlaciones entre este instrumento y otros de la batería utilizada en “Plaza Sésamo”, por ejemplo, con el test de letras y palabras (que fue de 0.34), con el test de habilidades para clasificar (que fue de 0.28), etc.

En el INCCAPAC se hizo un pequeño estudio con niños de 5 años de edad, a los cuales se les aplicó el retest un mes después del test, y se obtuvieron los siguientes resultados: el coeficiente de consistencia interna de 0.80 y el coeficiente de estabilidad temporal 0.52.

Dentro de las pruebas relativas al desarrollo del lenguaje se encuentran las utilizadas en el INCCAPAC para la evaluación del programa “Plaza Sésamo”.

Test de reconocimiento y designación de objetos

Este test implica dos tareas diferentes para el niño: la primera es la de reconocer los objetos que constituyen el material de la prueba, y la segunda es la de identificar aquellos objetos cuando el examinador se lo pida. Es decir, en el ejemplo de la ilustración (véase figura 11), si el niño no

nombró la hamaca, entonces el examinador le preguntará: “¿Cuál es la hamaca?” Esto se hará al finalizar el test, regresando desde los primeros estímulos donde el sujeto falló.

Esta prueba, desarrollada por Klein en Guatemala, fue utilizada en México, obteniéndose un coeficiente de estabilidad temporal de 0.90, con un intervalo de un mes.

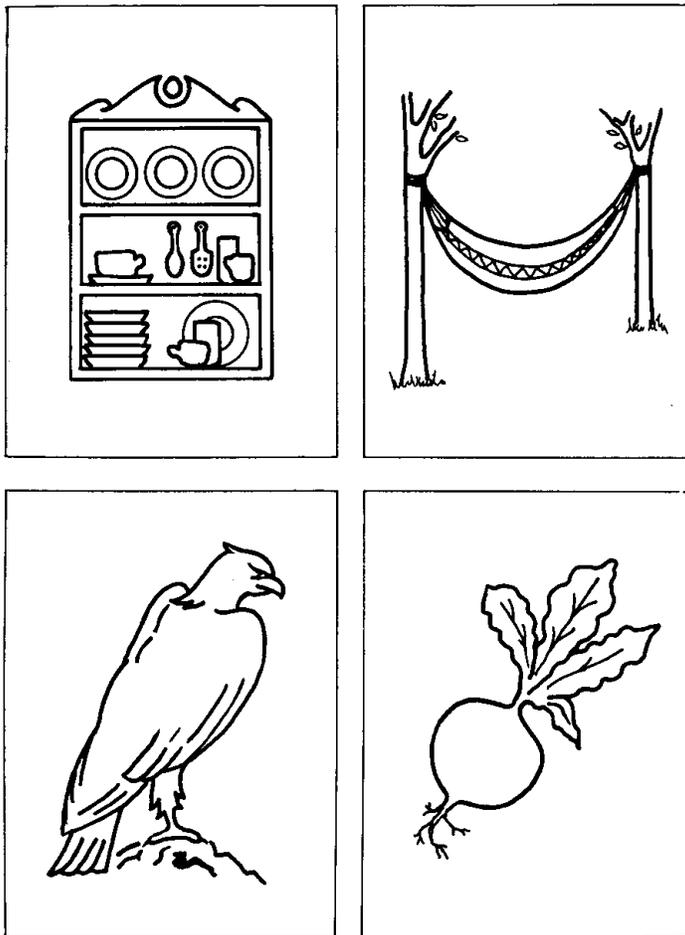


Figura 11.

Test de preposiciones

El objetivo de este test es investigar el uso y conocimiento que el sujeto tiene de las preposiciones, como parte importante del desarrollo de su lenguaje. Es un magnífico instrumento diseñado para investigar *relaciones* instrumentales como arriba, abajo, detrás, adelante, entre, etc., mediante la manipulación de objetos que el propio niño hace en las posiciones que las preposiciones se lo piden. Este test fue desarrollado especialmente en el INCCAPAC, obteniéndose un coeficiente de confiabilidad de 0.55 en el lapso de un mes.

Test de funciones de los objetos

Mediante este test se investiga la utilización de los verbos reflexivos. También fue desarrollado en el INCCAPAC. Los estímulos de este test son dibujos de objetos de uso común que se le muestran al niño con la pregunta: “¿Para qué te sirve esto?” (por ejemplo, un vestido). Al plantearle la pregunta, el niño deberá contestar: “Para vestirme”. La confiabilidad del test-retest arrojó un coeficiente de 0.76.

Test de memoria de oraciones

Este test fue desarrollado por el doctor Klein en Guatemala, en el subtest similar del WPPSL. A sugerencia del doctor Bernardo Vallejo, asesor del INCCAPAC, consta ahora de dos partes: en la primera, como su nombre lo indica, se le pide al sujeto que repita automáticamente la frase estímulo. La segunda de ellas tiene como fin la comprensión de oraciones, de tal manera que se vuelven a encontrar dos diferentes objetivos en una sola prueba: por un lado, la medición de la retención de las oraciones (es decir, la cuantificación de la memorización del sujeto) y por otro, la comprensión de estas oraciones. El coeficiente de estabilidad temporal para la parte de memoria fue de 0.52 y el de comprensión de 0.71. Además, en el INCCAPAC se obtuvo la consistencia interna del test con un coeficiente de 0.41.

Test de letras y palabras

Esta prueba fue originalmente desarrollada por el Educational Testing Service (ETS), adaptada posteriormente por el CIPAC (Centro de Investigaciones Pedagógicas, Asociación Civil) y modificada por el INCCAPAC. La confiabilidad del test-retest dio un coeficiente de 0.80 al mes de aplicación.

Esta prueba consiste en lo siguiente: un grupo de letras-estímulo se le presentan simultáneamente al sujeto y se le pide que identifique una de ellas; después, que identifique sílabas, luego palabras, y posteriormente se le muestran una serie de dibujos de objetos, los cuales empiezan con la misma letra, y se le pide al sujeto que señale con qué letra empieza el nombre de cada uno de estos objetos (citar un ejemplo de cada uno). Esta es una prueba de rendimiento. (Véase figura 12).



BAJA PAN
CAMA NIÑO

Figura 12.

Test de lenguaje creativo

Este test fue construido especialmente para el programa “Plaza Sésamo”, con el fin de medir el aprendizaje de los niños de la muestra de investigación. Consta de seis láminas, cada una de las cuales contiene una fotografía de personajes, tres que realmente pertenecen a dicho programa y tres que no lo son, todas ellas destinadas a que el niño desarrolle su creatividad, mostrando cómo puede construir una historia observando cada una de las fotografías.

SOL

SOL LES LOS COL

Figura 13.

Test de clasificación de objetos

Este test fue desarrollado en el INCCAPAC. Contiene 24 objetos, mismos que el niño deberá agrupar, dando la razón por la cual forma sus grupos. La calificación aquí utilizada es la misma que emplearon McGaughan y Moran en 1956, para evaluar dimensiones conceptuales.

Test de memoria de objetos

Este test, también desarrollado por el doctor Klein, tiene por objeto llevar a cabo la medición de la capacidad de “recuerdo”. Contiene doce juguetes, los cuales se le presentan al sujeto para que los memorice; esto es, se le pide que los vea y que los recuerde. Después explicará al examinador la forma en que ha compuesto sus categorías, clasificando los objetos recordados. Esta prueba muestra claras diferencias a través de la edad. En Guatemala tuvo un coeficiente de confiabilidad de 0.70. En México apenas se encuentra en proceso.

En este experimento, diseñado para trabajar con niños preescolares como parte de los estudios evaluativos de “Plaza Sésamo”, se manejó un grupo experimental y uno control muestreados al azar.

Como se ha indicado en la descripción de cada una de las pruebas, algunas de ellas fueron tomadas del estudio experimental del *Sesame Street* y esta investigación permitió la construcción de algunos tests diseñados especialmente para “Plaza Sésamo” en México. Además, algunos de los tests que miden habilidades cognoscitivas fueron adaptados para ser utilizados en niños preescolares de habla hispana (Latinoamérica).

La muestra de niños se obtuvo de ocho guarderías dependientes de la Secretaría de Salubridad y Asistencia Pública. Tales niños eran representativos de la clase socioeconómica, edad y sexo que el diseño de investigación requería. Se hizo hincapié en la clase socioeconómicamente baja, a fin de establecer los programas de televisión educativos para niños (en esta importante parte de la población se ha encontrado un alto grado de analfabetismo). Se utilizó un cuestionario demográfico para la obtención de la muestra de niños cuyos padres no habían terminado la escuela primaria y tenían una ocupación de trabajadores no especializados. Para el primer estudio piloto se seleccionaron 221 niños, la mitad formando el grupo control y la otra mitad el grupo experimental.

Los niños del grupo experimental fueron concentrados en un salón especialmente preparado para ver el programa de “Plaza Sésamo”, durante 50 minutos cinco días a la semana. La serie de “Plaza Sésamo” estuvo constituida por 130 programas transmitidos durante seis meses. Entretanto, a los niños del grupo control se les permitía ver otro tipo de programas no educacionales, o bien dedicarse a otras actividades.

Las pruebas utilizadas para evaluar el aprendizaje obtenido por los niños de ambos grupos fueron administradas antes de iniciarse el programa pretest; luego, a las siete semanas de la transmisión, y al terminarse la serie del programa postest.

El impacto que causó el programa de “Plaza Sésamo” en el grupo experimental, fue evaluado mediante la administración individual de los tests. Además, un grupo de observadores

registró el grado de atención mostrado por los niños del grupo experimental sobre el programa, utilizando una escala de cuatro puntos para medir la atención.

Otra variable que se controló fue el número de ausencias de los niños en cada grupo, a fin de mantener en el último análisis el diseño experimental tan puro como fuera posible.

Los psicólogos examinadores fueron cuidadosamente entrenados para que la administración, calificación y resultados finales de las pruebas no se vieran alterados o se invalidaran por errores en el manejo de los instrumentos. Cada una de estas pruebas fue sometida a estudios piloto en muestras pequeñas, antes de aplicarse a las muestras definitivas.

A todos los datos recabados en los estudios mencionados se les aplicaron los siguientes tratamientos psicométricos:

1. Test-retest, para la obtención del coeficiente de estabilidad temporal.
2. Análisis de contenido, para comprobar si las pruebas utilizadas realmente midieron aquello para lo que habían sido construidas.
3. Análisis de ítemes, para detectar el peso de cada uno de ellos y para establecer el orden de dificultad para México y el poder discriminativo de cada ítem.
4. Análisis correlacionales entre las variables de todas y cada una de las pruebas.
5. Análisis factoriales para cada una de ellas.
6. Una diversidad de análisis de varianza.

El siguiente estudio, que ya se encuentra en desarrollo, contiene una muestra de 1080 niños de 4 a 5 años de edad, de ambos sexos, con un número igual de niños y de niñas en cada grupo, pertenecientes a tres niveles socioeconómicos: clase media, clase baja y clase baja rural.

Como en el estudio anterior, la muestra total se dividió equitativamente en dos grupos, el experimental y el control, y se está administrando en la misma forma conocida (es decir, pretest, retest y postest). En este momento la investigación se encuentra en el periodo de muestreo y pretest, para que, una vez terminado este primer paso, se inicie la transmisión del programa al grupo experimental.

Resulta importante mencionar aquí, que es la primera vez que en México, así como en todos aquellos países que transmiten por televisión el programa “Plaza Sésamo”, se ha diseñado esta investigación que, bajo la dirección del doctor Rogelio Díaz-Guerrero y la subdirección de la maestra Isabel Reyes Lagunes, además de un equipo de psicólogos asistentes de investigación, se está realizando en el INCCAPAC, con el fin de hacer un estudio longitudinal en un periodo mínimo de dos años y máximo de cuatro, para probar los efectos de dicho programa a largo plazo. Es decir, se investigará si los objetivos para los cuales ha sido elaborado cumplen sus propósitos, al hacer más fácil la adquisición de conocimientos del niño en el primer año escolar. En este momento de la investigación se les administrarán a los sujetos de ambos grupos una serie de pruebas de rendimiento escolar que están en proceso de construcción en el propio INCCAPAC, para medir lo aprendido a través del programa, cuando ambos grupos se encuentren cursando el primer año de primaria. Además, durante el primer semestre de ese año se llevará a cabo un programa de análisis conductual en el medio escolar, a fin de validar por una parte los instrumentos, y por otra, para poder llegar a conocer más profundamente los efectos de “Plaza Sésamo” en los escolares mexicanos.

Para mayor información acerca de estos instrumentos de medición, puede consultarse el proyecto llamado “Televisión Educativa para Niños Preescolares en México: Un Estudio Sumario, Sistemático y Experimental de Plaza Sésamo”, que se encuentra en la biblioteca de la Facultad de Psicología de la UNAM.

UTILIDAD DEL DIBUJO EN LA MEDICIÓN DEL DESARROLLO PSICOLÓGICO DEL NIÑO

Es bastante antigua la idea de que el dibujo espontáneo del niño puede arrojar algún tipo de información acerca de su desarrollo.

En 1885, Ebenezer Cooke publicó un artículo sobre el dibujo de los niños, en el cual describió estadios sucesivos del desarrollo conforme los fue observando.

En 1867, Conrado Ricci publicó un estudio de dibujos de un grupo de niños italianos, quizá la primera colección que se ha registrado en la historia de la investigación de estos aspectos.

Este movimiento, iniciado en la última década del siglo XIX, ha continuado hasta la actualidad. Aunque se piensa que en algunos aspectos, particularmente en ciertos periodos, los dibujos de los niños no pueden ser considerados con un alto grado de validez, de todas formas constituyen material importante que, aunado a otro tipo de expresiones conductuales del niño, son de gran ayuda para la evaluación de su desarrollo.

Podrían mencionarse, a partir de Cooke, un sinnúmero de investigadores que entre 1885 y 1920 se dedicaron a trabajar profundamente con dibujos de niños. Sin embargo, en la mayoría de ellos hubo poca seriedad en la fundamentación de una teoría del dibujo como una forma de comportamiento factible de medición.

Así pues, puede mencionarse primeramente a Florencia Goodenough, quien en 1926 demostró con gran éxito la fuerte carga intelectual que existe como componente en el desarrollo de los dibujos de un hombre en el niño. El estudio de este interesante problema tomó entonces una nueva dirección: la investigación psicométrica de la inteligencia, que a lo largo de muchas décadas, aún en el presente, especialmente en América, Inglaterra y Japón ha tenido su auge.

El test de Goodenough ha sido estudiado por múltiples investigadores y en muchos lugares del mundo. Además, no sólo se ha investigado psicométricamente, sino que también se le ha empleado en los llamados “métodos proyectivos de la personalidad”, que en términos de investigación ha dejado muy confusos resultados.

Desde un principio, Goodenough indicó que los dibujos podrían ser utilizados para el estudio de la personalidad mediante su interpretación. Esto originó una serie de investigaciones poco metodológicas y rigurosas que solamente arrojaron aquellos datos que deseaban encontrar quienes las trazaron.

Este test tiene por objeto estudiar los factores intelectuales que intervienen en el dibujo espontáneo del niño pequeño. Al respecto, Goodenough dice: “Cuando el niño traza la figura humana sobre un papel, dibuja lo que sabe acerca de ella y, por tanto, está ejecutando un trabajo

intelectual y expresando su repertorio conceptual, el cual crece conforme su edad mental se desarrolla, observación que puede hacerse a partir de los dibujos que el niño ejecuta desde, aproximadamente, los 5 hasta los 10 años de edad”.

Las funciones que fundamentalmente intervienen para la ejecución de esas tareas, según Goodenough, son las siguientes: asociación, observación analítica, discriminación, memoria de detalles, sentido espacial, juicio, abstracción, coordinación visomanual y adaptabilidad.

Su escala, cuyo mérito tiene el haber estudiado a miles de niños de 2 a 15 años de edad, de ambos sexos y de diferente nacionalidad, raza, grado y rendimiento escolar, fue sometida a estudios minuciosos y repetidos de validez y confiabilidad de análisis de ítems y de procedimientos estadísticos muy cuidadosos, a fin de llegar a establecer el nivel de desarrollo del niño a través del grado de perfeccionamiento y completamiento de un dibujo, calificando la presencia o ausencia de los ítems establecidos a partir de los siguientes criterios:

- a) Cantidad de detalles representados.
- b) Proporcionalidad.
- c) Bidimensionalidad.
- d) No-transparencias.
- e) Congruencias.
- f) Plasticidad.
- g) Coordinación visomotora.
- h) Perfil.

Al sujeto simplemente se le pide que dibuje la figura de un hombre. Los resultados obtenidos por la doctora Goodenough a través de todos sus estudios psicométricos fueron los siguientes:

1. Respecto a la validez, y con el fin de precisar si esta prueba puede servir como predictora de éxito escolar, se llevó a cabo un experimento con un grupo de 286 niños, pertenecientes al cuarto y quinto año de primaria, a quienes se les aplicó la prueba de Goodenough y dos tests de inteligencia (el Army Alfa y la forma B del test de completamiento Trabue). También mediante una escala de 5 puntos se obtuvieron los juicios de los maestros acerca de la inteligencia de los examinados. La correlación encontrada entre el Goodenough y el Army Alfa fue en relación al progreso-grado, arrojando un coeficiente de 0.69, en tanto que los demás instrumentos utilizados dieron un promedio de correlación de 0.59. Más adelante se encontró que la prueba de Goodenough y la proporcionalidad del progreso-grado era de 0.37. Esto indica que el test mide un aspecto del intelecto en forma distinta que el resto de las pruebas, el cual es desatendido por los maestros cuando estiman la inteligencia de los alumnos.
2. También se estableció la correlación entre la puntuación del dibujo y las edades mentales correspondientes a la prueba de Stanford-Binet, la cual se aplicó a 334 niños clasificados en grupos de edad. Se encontró que, en general, las correlaciones fueron bastante significativas, siendo las más bajas las de los niños de 8 años, con un coeficiente de 0.55, y las más altas las de los niños de 4 años, con un coeficiente de 0.86.

3. En cuanto a la confiabilidad, se obtuvo un coeficiente de 0.77 mediante el método de división por mitades, utilizando la fórmula de corrección de Spearman-Brown.
4. El error probable (EP) en la estimación de un cociente intelectual verdadero fue aproximadamente de 5.4 para las edades de entre 5 y 10 años.
5. En los análisis de varianza se encontró que, de los siete grupos de edades estudiados (4 a 10 años), el coeficiente de variabilidad permaneció casi constante. No se tuvieron en cuenta a las edades de 11 y 12 años porque se encontró que el test no discrimina a esos niveles, en relación a capacidades superiores.

En México, como parte del programa de investigación del IDPEM, se utilizó el test de Goodenough dentro de la batería diseñada para esta investigación a lo largo de los cinco primeros años del estudio longitudinal.

Cabe mencionar el hecho de que desde 1964, año en que se inició este proyecto de investigación, hasta la fecha se han realizado importantes estudios respecto a la estabilidad temporal del *Test del dibujo de la figura humana*. Obviamente, existen discrepancias en cuanto a criterios. Así pues, mientras en Inglaterra, Kelmer y Pick Up, investigadores de la Universidad de Birmingham, no hallaron estabilidad en la medición de la inteligencia a través de esta prueba, J. Vane, de los Estados Unidos, encontró coeficientes de estabilidad temporal de 0.65 y 0.80 en test-retests, con intervalos de 12 meses cada uno, a una muestra de 280 niños preescolares, blancos y negros.

En México se llevó a cabo un estudio con una muestra de 384 niños, distribuida en tres diferentes grupos de edades, siguiendo el diseño de la muestra en el orden siguiente:

- a) Grupo I, 6.7 años de edad (N = 133).
- b) Grupo II, 9.7 años de edad (N = 118).
- c) Grupo III, 12.7 años de edad (N = 133).

Con un intervalo de doce meses y dada la complejidad de los factores que influyeron en el desarrollo psicológico del niño a lo largo de un año, los resultados obtenidos en estos estudios pueden considerarse significativos. Así, puede afirmarse, por una parte, que la media aritmética (0.50, 0.51 y 0.58, respectivamente) para cada grupo no sufre cambios ostensibles de un año a otro entre los tres grupos. Esto quiere decir que la prueba no muestra claramente los cambios que ocurren a través de un año; en contraposición, su poder discriminativo es evidente cuando se estudian a intervalos de tres años, en donde ya se puede observar una curva de desarrollo cuando la calificación aumenta con la edad. Es importante hacer notar que la curva de desarrollo de la muestra mexicana no es similar a la que reporta Goodenough en su estudio de estandarización, en el cual a los seis años no se presentan diferencias significativas, mientras que en la muestra mexicana, de 9 a 12 años, la media fue inferior, sobre todo a los 12 años.

A fin de poder dar un resumen más detallado acerca de la utilidad de esta prueba en México, tiene que hacerse mención del test del doctor Dale B. Harris, llamado *Test de dibujos de niños como media de la maduración intelectual*. Este trabajo es una revisión y extensión de la prueba de

Flores Goodenough, en la cual el doctor Harris fundamenta, teórica y prácticamente, la utilidad del dibujo en la medición del desarrollo psicológico y contribuye ampliamente en la revisión de este test, primero en la extensión de la prueba para ser aplicada a adolescentes, y luego en la elaboración de una escala de una forma “alternada” a la escala para el hombre que se aplique también a la mujer, empleando los mismos requerimientos. El primer intento (es decir, la aplicación a adolescentes) parece que no ha tenido éxito; en cambio, el segundo ha sido de gran utilidad.

Las escalas de puntuación para el hombre y la mujer, de Harris, han sido estandarizadas en muestras más representativas que las de Goodenough y, entre otras cosas, confirman parcialmente la idea sustentada por esta investigadora, en el sentido de que la figura femenina no discrimina tan bien como la masculina. Harris encuentra que la figura femenina presenta progresos definitivamente evidentes.

MEDICIÓN DEL DESARROLLO PSICOLÓGICO EN LA ETAPA ESCOLAR.

Esta etapa abarca desde los 6 hasta los 15 años de edad; es decir, nos referimos a aquellos niños que supuestamente tienen una escolaridad que va desde el primer año de primaria hasta el tercero de secundaria.

Esta etapa es de fundamental importancia, ya que está caracterizada por un notable desarrollo intelectual cognoscitivo.

De los 6 años en adelante, el crecimiento físico, anteriormente muy rápido, empieza a observarse con un poco de lentitud. De los 6 a los 10 años, la estatura de los niños tiene una tasa de entre 6 y 7% anual. En promedio, los niños son ligeramente más altos que las niñas y, al igual que esta información, se pueden encontrar múltiples cambios físicos relacionados incluso con el sexo.

Los procesos psicológicos implicados en esta etapa son significativos. Si se observan los procesos cognoscitivos (es decir, la información sensorial, la comprensión de palabras y signos, la habilidad para manipular símbolos que permitan solucionar problemas, etcétera), se verá que son procesos activos de percepción, memoria, ideación, evaluación, razonamiento, y el establecimiento de correlatos libremente.

Se pueden distinguir tres unidades cognoscitivas fundamentales: las imágenes, los símbolos y los conceptos y reglas de conceptualización. Así, en la manipulación de todos los elementos mencionados se va desarrollando el fenómeno psicológico denominado *inteligencia*.

Los conceptos son formas simbólicas que denotan la similitud entre situaciones y objetos y que permiten al niño establecer ciertos órdenes de experiencia. En todo proceso cognoscitivo desempeña un papel primordial la percepción y, luego, la codificación del ambiente, que implica operaciones de atención selectiva hacia ciertos eventos y la estipulación de un membrete o de un nombre determinado para localizarlo. El niño ejecuta estos procesos en forma espontánea, y lo importante de esto es que sus interpretaciones del ambiente se van modificando a través del desarrollo.

En la edad escolar, el niño aprende a usar las palabras y los conceptos para la evaluación de las situaciones que se le presentan. Mientras más completo sea su repertorio conceptual, más apegado a la realidad será su funcionamiento.

Para poder darse cuenta de los progresos psicológicos que el niño va adquiriendo, es necesario considerar forzosamente, por una parte, la diversidad de escenarios en los cuales actúa el niño, y por otra, cómo se puede evaluar la calidad de la información adquirida de ese ambiente y la forma de utilizarla. Esto lleva de inmediato a plantear las siguientes preguntas: ¿Cuánto ha adquirido de información?, ¿qué clase de información ha adquirido?, ¿cómo la utiliza? Dicho en otras palabras, ¿cuánto ha aprendido?, ¿de qué calidad es ese aprendizaje y cómo lo maneja para solucionar los problemas a los cuales diariamente se enfrenta en su ambiente específico?

Aquí entran en acción los conocimientos del psicólogo acerca del desarrollo psicológico y sus formas de medirlo. La medición de las múltiples operaciones que ejecuta el individuo en esta etapa tiene que considerar no sólo el medio familiar, escolar y, en general, social que permita interpretar en forma correcta las mediciones efectuadas sobre aquellos elementos o funciones que nos interesa medir, a fin de determinar el proceso evolutivo del niño, sino también, en forma particular, el tipo de instrumentos de medición psicológica que habrán de emplearse para obtener tales resultados.

Por cuanto toca a la medición del proceso de desarrollo psicológico, se ha puesto especial atención, como ha ocurrido en general en la psicología, en los aspectos intelectuales, perceptuales y en aquellos aspectos conductuales denominados rasgos de la personalidad.

Desde luego, la necesidad de conocer más acerca de estos fenómenos tan importantes ha producido una gran cantidad de investigaciones sobre los diversos aspectos que componen el comportamiento del individuo, en este caso el niño. En países muy desarrollados se lleva, incluso, un registro de tales aspectos desde que el niño nace, de tal manera que cuando éste llega a la escuela primaria, se tienen ya los antecedentes suficientes para darle una orientación particular al proceso de enseñanza-aprendizaje.

El gran número de tests psicológicos contruidos para esas mediciones, que en algunos países, en donde, por la cultura y la educación, se consideran de gran importancia, ha determinado el establecimiento de instituciones especializadas en ejecutar ese trabajo.

Así como en las situaciones anteriores se ha indicado la intención que este texto tiene de presentarle al estudiante de psicología lo que en su propio país es de utilidad para el desempeño de su futura profesión, aquí mencionaremos simplemente lo que sabemos que se ha realizado en México en este campo, sin dejar de tener en consideración la importancia que tanto las investigaciones psicológicas como el trabajo profesional del psicólogo en otros países han aportado a nuestro conocimiento, para poder aplicarlo al medio mexicano.

MEDICIÓN DE LA INTELIGENCIA EN LA ETAPA ESCOLAR

En ese sentido, cabe mencionar primero el hecho de que en México, los psicólogos, los maestros, los padres de familia y todas aquellas personas que están en contacto con el niño, se han preocupado poco y le han dado menor atención a la medición de las etapas del desarrollo psicológico de aquél. Esto ha producido efectos hasta cierto punto negativos, ya que el mexicano, en general, ve

en los tests psicológicos una tremenda “amenaza”. Como es lógico suponer, esto se debe al desconocimiento de su utilidad y a la desconfianza en que conceptos equivocados han colocado a la psicología.

Aquí no se hablará sobre el desarrollo teórico de la medición de la inteligencia, pues ya se ha hecho en el capítulo 3. Nos limitaremos simplemente, hasta donde la información alcance, a mencionar aquellos instrumentos de medición psicológica más utilizados en México, a pesar de que su uso no esté dentro de las reglas estrictamente psicométricas que se requieren. En este caso se encuentran numerosos tests que, importados, sólo se han traducido al español y se han empleado sin la precaución debida; es decir, muchos tests no han sido adaptados para México e, incluso, ni siquiera para los grupos específicos a los cuales se les ha aplicado; en consecuencia, tampoco han sido estandarizados, ni existen tablas de normas de calificación especiales que permitan confiar plenamente en los resultados de esas mediciones.

Durante mucho tiempo se ha venido utilizando el test de Stanford-Binet, revisión de 1936 y 1960, sin que exista estandarización alguna, ni siquiera adaptaciones de este test a los diversos grupos de niños a quienes se ha aplicado y que necesariamente difieren en aspectos tan fundamentales como el nivel socioeconómico, el nivel educacional de la familia, los sistemas escolares, las zonas geográficas de localización, el sexo, la edad, etcétera. Esto es válido para cualquier tipo de test que se emplee en México que no haya sido tratado previamente a través de los requisitos psicométricos y metodológicos necesarios para que puedan obtenerse resultados válidos y confiables.

Se sabe que, además el test de Stanford-Binet, se utilizan otros, como el de Goodenough, el gestáltico visomotor de Bender, el de Merrill-Palmer, las escalas de inteligencia de Wechsler y una gran cantidad de tests de habilidades específicas, cuyos nombres aún no se conocen porque algunos han sido cambiados al traducirlos al español, y otros porque están constituidos por partes de diversos tests y llevan el nombre que se le ha ocurrido al investigador que ha realizado ese trabajo.

Es importante hacer hincapié en el hecho de que los estudiantes de psicología deben entender que si no se lleva a cabo una evaluación cuidadosa de los instrumentos que se emplean para medir los atributos del ser humano que se requieran en un momento dado, nunca se podrá estar seguro de que la ejecución de las tareas de los psicólogos es realmente la correcta.

Anteriormente se mencionó a las escalas de Wechsler como algunos de los tests utilizados indiscriminadamente en México para la medición de la inteligencia. En este caso, como en todos los demás, ocurre que cada psicólogo que emplea los tests hace sus propias traducciones, establece sus propias instrucciones y utiliza normas de calificación originadas de los grupos extranjeros para los cuales se construyó el test. El resultado lógico es que cada psicólogo tiene su propia versión del test y sus conclusiones no son legítimas, pues ha utilizado un parámetro de comparación que hipotéticamente es diferente del que debería usar con sus propios grupos o individuos. Decimos que hipotéticamente, porque en algunos casos tal vez no existan diferencias significativas entre los grupos extranjeros y los de México; sin embargo, esto no se puede probar, sino hasta después de haber efectuado los estudios correspondientes con el test de que se trate.

Existen algunas estandarizaciones para grupos específicos hechas por algunos psicólogos mexicanos para los grupos pertenecientes a su área de aplicación profesional. Entre éstos, se pueden

mencionar los trabajos realizados por la maestra Olga Loredo y por sus colaboradores, con el *Test de matrices progresivas* de Raven en estudiantes de preparatoria de la ciudad de México, y algunas pequeñas estandarizaciones de este mismo test realizadas por psicólogos que trabajan en diversas escuelas primarias, principalmente privadas, con el fin de clasificar niños para colocarlos homogéneamente en sus respectivos grados escolares; pero, como se indicó con anterioridad, no existe la comunicación entre estos psicólogos y, por tanto, cada uno tiene sus propios estudios.

A través del trabajo realizado en tests profesionales para obtener el grado de psicólogo, se han podido llevar a cabo algunas estandarizaciones y adaptaciones de ciertos tests, como el de dominós, de Anstey, para diversos grupos; o el de Barranquilla (BARSIT), construido por el doctor Del Olmo en Colombia y que ha sido adaptado y estandarizado en México, tanto por el Departamento de Orientación Vocacional de la Preparatoria del Estado de México, como en algunos grupos de obreros especializados, tales como los de Ferrocarriles de México, o bien para los trabajadores de Petróleos Mexicanos y para algunos grupos de empleados del Seguro Social.

Podrían mencionarse muchos otros tests; pero, para no incurrir en errores de información ya que ésta no es muy clara, por el momento sólo cabe recordar la minuciosidad que el psicólogo debe tener al evaluar un instrumento que ha de emplear para tomar decisiones importantes en la vida de los individuos.

Así como existen numerosas pruebas de medición de la inteligencia, utilizadas en los diversos departamentos de psicología, de orientación vocacional, de selección de personal, de selección escolar, etcétera, diseminados por todo el país, también hay muchos tests de personalidad que se aplican a los niños aún con más peligro, ya que tales tests, en su mayoría, conducen a la libre interpretación de su contenido por parte de quien los aplica, pues, como se sabe, este tipo de instrumentos son cualitativos o tipológicos y poco estudiados para su utilización en estas edades.

Debido a esta situación tan ambigua en el campo de la medición en México, y habiendo conciencia de la necesidad de tener instrumentos psicológicos confiables para los requerimientos del medio mexicano, el doctor Rogelio Díaz-Guerrero y sus colaboradores iniciaron en 1964, en el Centro de Ciencias del Comportamiento (ahora INCCAPAC, Instituto Nacional de Ciencias del Comportamiento y de la Actitud Pública, Asociación Civil, del cual es director el doctor Díaz-Guerrero), una investigación denominada IDPEM, “Investigación sobre el desarrollo de la personalidad del escolar mexicano”, cuyos objetivos eran, por una parte, entrenar estandarizaciones y normas de calificación para sujetos mexicanos al utilizar algunos instrumentos psicológicos.

Este estudio longitudinal fue realizado a lo largo de seis años de investigación sucesiva en sujetos que cubrían un rango de edad de 6 a 18 años, divididos en tres grupos en función del sistema escolar, comprendiendo ambos sexos. El sistema escolar se refiere al de escuelas privadas y al de escuelas oficiales que, para fines analíticos, se dividió en dos: el sistema propiamente oficial y el de las escuelas correspondientes a las unidades habitacionales, que en este caso fue la Unidad Independencia.

Al iniciarse la investigación, después de haberse obtenido una muestra al azar y representativa de los escolares de la ciudad de México, también se formaron tres grupos por edad y grado escolar. El grupo I correspondió a aquellos niños que tenían 6 años 8 meses al iniciarse el estudio; el grupo II, a aquellos con una edad de 9 años 8 meses, y el grupo III, a aquellos que tenían

una edad de 12 años 8 meses, todos ellos correspondientes al primero y cuarto años de primaria y al primero de secundaria, respectivamente.

Este diseño experimental cubre, así, doce años de investigación en escolares mexicanos a través de seis años de estudio. Esto ocurrió así, debido a que la muestra fue una de las denominadas muestras superpuestas, ya que los sujetos que se encontraban en el primer año de primaria al iniciarse la investigación, estaban ya en cuarto grado al cuarto año de aplicación, obteniéndose de este modo dos muestras comparables de sujetos en el mismo grado y con la misma edad, ya que el grupo II de niños fue investigado inicialmente cuando se encontraba en cuarto año de primaria, y al llegar al primer año de secundaria se superponía a la muestra de aquellos niños que se hallaban en este grado al iniciarse la investigación, y así sucesivamente. Para una mejor comprensión de esto, obsérvese la siguiente tabla:

Tabla 5. Distribución de la muestra del IDPEM, por grado escolar y secuencia de aplicación.

Grados Escolares

I	1°			4°			7°					
II		2°			5°			8°				
III			3°			6°			9°			
IV				4°			7°			10°		
V					5°			8°			11°	
VI						6°			9°			12°

Año de
Aplic.

La batería para el estudio estuvo constituida por los siguientes instrumentos:

1. HIT, Holzman Inkblot Technique.
2. WISC, Wechsler Intelligence Scale for Children.
3. TASC, Text Anxiety Scale for Children.
4. DSC, Defensive Scale for Children.
5. WAIS, Wechsler Adult Intelligence Scale.
6. Test de Witkin.
7. Test de clasificación de objetos, de R. W. Gardner.
8. Test de fraccionamiento visual, de J. Kagan.
9. Test de estilo conceptual, de J. Kagan.
10. Test de Goodenough.
11. Test de Goodenough-Harris.
12. Test de estimación de tiempo.
13. Escala de conducta.
14. Test de stress.
15. Test de asociación de palabras.
16. Test de maduración perceptual.

A continuación, en forma muy breve se explicará qué aspectos mide cada uno de estos tests y qué variables manejan.

De la *Técnica de manchas de tinta* de Holtzman, sólo se indicarán las variables, ya que en el capítulo 6 se hablará más ampliamente de ella. Tales variables son las siguientes:

1. Tiempo de reacción	TR
2. Rechazo	R
3. Localización	L
4. Espacio	E
5. Forma definida	FD
6. Forma apropiada	FA
7. Color	C
8. Sombreado	S
9. Movimiento	M
10. Verbalización patognómica	VP
11. Integración	I
12. Humano	H
13. Animal	A
14. Anatomía	At
15. Sexo	Sx
16. Abstracto	Ab
17. Ansiedad	As
18. Hostilidad	Hs
19. Barrera	Br
20. Penetración	Pn
21. Balance	B
22. Popular	P

Por lo que se refiere a las escalas de Wechsler, éstas también han sido explicadas ampliamente en el capítulo 3. Aquí sólo se expondrán las variables que se manejan, las cuales son:

Escala verbal

1. Subtest de información.
2. Subtest de comprensión.
3. Subtest de aritmética.
4. Subtest de analogías y semejanza.
5. Subtest de retención de dígitos.
6. Subtest de vocabulario.

Escala de ejecución

7. Completamiento de figuras.
8. Diseños con bloques.
9. Ordenamiento de imágenes.
10. Ensamble de objetos.
11. Símbolos de dígitos.

12. Laberintos.
13. Suma de las calificaciones de la escala verbal.
14. Suma de las calificaciones de la escala de ejecución.
15. Suma de calificaciones de la escala total.

Además, para los análisis pertinentes, se tienen en consideración las puntuaciones transformadas o “pesadas” de cada uno de los subtests mencionados. Esta descripción corresponde a la escala de inteligencia Wechsler para niño, es decir, el WISC. (Como todas las escalas de Wechsler se fundamentan en los mismos principios, se le sugiere al lector que consulte el capítulo 3 para ampliar su información.)

Otro de los instrumentos estandarizados en esta investigación es el del *Dibujo de la figura humana*, construido por Florencia Goodenough, y la revisión y extensión de este test, realizadas por el doctor Dale B. Harris, que se denomina *Los dibujos de los niños como medida de la maduración intelectual*. En la sección anterior se expusieron ambos tests; aquí solo se mencionan las variables que se manejan:

1. Calificación de los elementos presentes en el dibujo de un hombre, del test de Goodenough.
2. Calificación de la figura de una mujer, de acuerdo a los criterios de Goodenough.
3. Calificación de la primera figura del test de Goodenough-Harris.
4. Calificación de la segunda figura del test de Goodenough-Harris.
5. Escala de masculinidad y femineidad (figura masculina).
6. Escala de masculinidad y femineidad (figura femenina).

Como dato curioso, se ha podido observar que la confiabilidad de la prueba disminuye conforme el niño crece.

Test de estilo conceptual, de Kagan

Esta prueba se ha adquirido para analizar los procesos cognoscitivos y, en general, aquellos procesos que se consideran dentro del campo del pensamiento. El material de este test es muy sencillo, su administración fácil, su tiempo de ejecución corto, y tiene la gran ventaja de ser una prueba objetiva, tanto en su aplicación como en su calificación y, por tanto, en sus resultados. Mediante esta técnica se ha intentado encontrar, a lo largo de los años de investigación, los estilos estándar de conceptualización en los estudiantes mexicanos para cada edad, nivel socioeconómico, sistema escolar y sexo. Este test también se estudió en la sección anterior; en consecuencia, aquí solamente se mencionarán las variables que en él se manejan, que son las siguientes:

1. Tiempo total de reacción.
2. Número de respuestas analíticas.
3. Número de respuestas relacionales.
4. Número de respuestas inferenciales.
5. Media del tiempo de reacción a respuestas analíticas.
6. Media del tiempo de reacción a respuestas relacionales.
7. Media del tiempo de reacción a respuestas inferenciales.

Test de fraccionamiento visual, de Kagan

Este test se ha construido para evaluar los procesos de aprendizaje perceptual. Consta de dos partes: la primera consiste de una tarea de aprendizaje de pares asociados; en ella, el sujeto debe aprender cuatro sílabas sin sentido (WON, FAM, PUF y SEP) y debe relacionarlas con cuatro estímulos que contienen las tres características implicadas en los procesos perceptuales, que son: figura, fondo y elemento. El estímulo es presentado tres veces por serie, haciendo que el sujeto repita la palabra asociada a cada dibujo, durante los cuatro primeros estímulos de la serie 1; después sólo se le pide que diga la palabra, y el examinador reforzará la respuesta presentando la palabra asociada con el dibujo, independientemente de que haya acertado o no la respuesta. Esta primera parte consta de cinco series, con doce estímulos cada una para la presentación de los pares asociados. Esta prueba puede suspenderse cuando el sujeto ha sido capaz de contestar correctamente doce estímulos consecutivos. En caso contrario, de todas maneras se aplicarán las cinco series y se pasará a la administración de la segunda parte del test. Esta consta de 24 estímulos, contenidos en tres variables, que son: figura (f), fondo (fo) y elemento (e); es decir, en cada estímulo se encuentran partes del estímulo total, mismas que pueden ser cualquiera de las tres variables mencionadas. Estas representan una fracción del estímulo total.

En esta parte del test, la tarea que el sujeto ha de llevar a cabo es la de relacionar la frase sin sentido que aprendió en la tarea 1, con las partes del estímulo que se le van presentando en cada ítem. Para esta segunda parte, los 24 estímulos se distribuyen de ocho en ocho para cada variable de las mencionadas y con un rango de calificación de 0 a 24. La distribución de los estímulos se hizo al azar. La primera parte del test se denomina *Tarea de aprendizaje* y la segunda *Transferencia de aprendizaje*.

Las variables que se manejan en este test son las siguientes:

1. Número de ensayos para aprender la tarea 1.
2. Tiempo de reacción del elemento.
3. Tiempo de reacción de la figura.
4. Tiempo de reacción del fondo.
5. Tiempo de reacción total.
6. Número de elementos correcto.
7. Número de figuras correcto.
8. Número de fondos correcto.
9. Número total de variables correctas.
10. Número de *wons* correcto.
11. Número de *fams* correcto.
12. Número de *pufs* correcto.
13. Número de *seps* correcto.
14. Número de errores.

Test de figuras ocultas, de Witkin

Este test se ha elaborado para medir la capacidad que tiene el niño para descubrir una figura sencilla que se encuentra oculta dentro de un contexto complejo. German Witkin, autor de este test, lo fundamenta en su teoría de la dependencia-independencia del campo, a través de los procesos perceptuales que permiten que el sujeto perciba, por ejemplo, un estímulo vertical dado, independientemente de que su campo visual externo sea vertical o esté inclinado hacia la izquierda o

hacia la derecha. Para Witkin, una persona es dependiente del campo cuando es incapaz de encontrar la figura sencilla que se encuentra dentro del contexto complejo; en caso contrario, será una persona independiente del campo.

La hipótesis de Witkin consiste en que el niño inicialmente se forma el concepto de su cuerpo de una manera global, y conforme va desarrollándose, va teniendo límites cada vez más precisos –y, por tanto, más diferenciado – del medio que lo rodea.

Este test está constituido por 12 figuras complejas, a colores, teniendo cada una su figura simple respectiva. Las figuras se presentan en tarjetas diferentes, para que el sujeto no pueda ver simultáneamente la figura compleja y la simple.

El tiempo que se le permite al sujeto como límite máximo para encontrar la figura simple es de 180 segundos. Al aplicarle el test, se le mostrará primero, durante 15 segundos, la figura compleja, y luego, durante 10 segundos, la figura simple. Inmediatamente después empieza el conteo de los 180 segundos. Durante este tiempo, el sujeto puede pedir volver a observar la figura sencilla tantas veces como lo necesite. Cada reexamen durará 10 segundos, en los cuales deberá detenerse el cronómetro. Al sujeto se le debe indicar claramente que tiene que encontrar cada figura en la misma posición y del mismo tamaño que la figura simple que se le ha mostrado. Cuando el sujeto diga que ya ha encontrado la figura simple, se le pedirá que la delinee solamente, sobre la figura compleja, con un artefacto de punta fina para no manchar ni marcar el estímulo. En el protocolo de calificación se anotará el número de errores y de reexámenes, así como el tiempo utilizado en localizar cada figura sencilla. Este test solamente puede aplicarse a sujetos que tengan un rango de edad mínimo de 9 años.

Las variables que este test maneja son las siguientes:

1. Número de errores.
2. Número de reexámenes.
3. Tiempo de ejecución de la tarea.
4. Número de respuestas correctas.
5. Media del tiempo correcto.
6. Media del tiempo total.

Test de clasificación de objetos

Este test, elaborado por el doctor R. W. Gardner, tiene como objetivo principal evaluar la categorización conceptual que los sujetos ejecutan para ordenar los objetos que los rodean. Debido a la libertad de que goza el sujeto para emitir sus respuestas, éste puede producir un gran número de ellas y tiene muchas posibilidades de expresión conceptual. El proceso de calificación e interpretación de este test es muy complejo y requiere de un minucioso estudio para su dominio. Así pues, considera dos dimensiones fundamentales: la público-privada y la abierto-cerrada, las cuales al intersectarse forman cuatro categorías o áreas conceptuales, que son: cerrado-pública, abierto-pública, cerrado-privada y abierto-privada. (Véase figura 15).

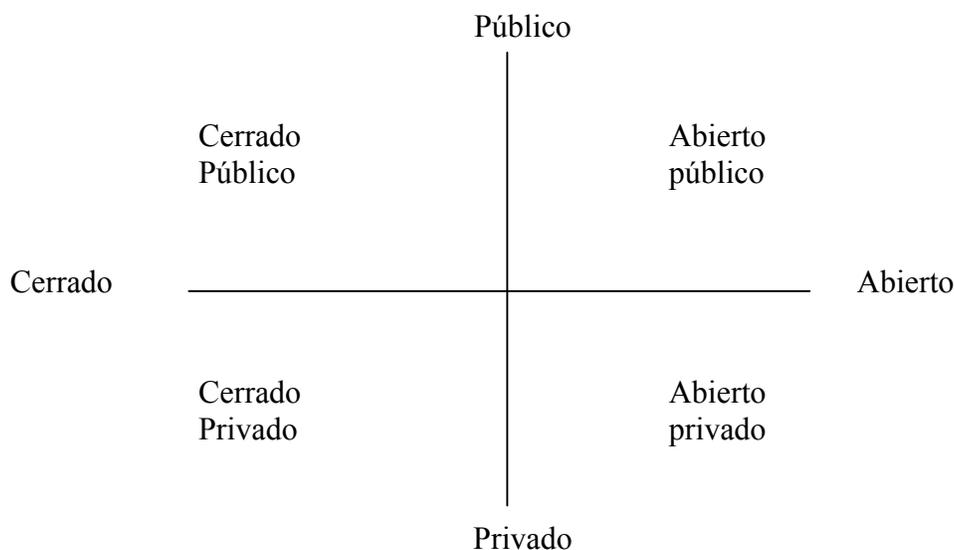


Figura 15.

Estas cuatro categorías se dividen en subcategorías, las cuales no se estudiarán aquí, por no ser tema de este libro. (Aquellos que estén interesados en el estudio de esta prueba, pueden consultar el “Manual de psicometría aplicada”, de la misma autora.)

Este test maneja las siguientes variables:

1. Número de grupos.
2. Número de ítems por grupo.
3. Número de objetos no utilizados.
4. Porcentaje de dimensiones cerrado-públicas.
5. Porcentaje de dimensiones cerrado-privadas.
6. Porcentaje de dimensiones abierto-públicas.
7. Porcentaje de dimensiones abierto-privadas.
8. Porcentaje de dimensiones cerradas.
9. Porcentaje de dimensiones abiertas.
10. Porcentaje de dimensiones públicas.
11. Porcentaje de dimensiones privadas.

Escalas de Ansiedad (TASC) y de Defensa para Niños (DSC)

La primera de estas escalas, publicada en 1960 por Sarason y sus colaboradores, como instrumento válido para la medición de los efectos de la ansiedad en la ejecución del rendimiento escolar de los niños, así como los estudios de Rebusch en el mismo año respecto a la escala de defensa para niños, cuyo objetivo es medir la tendencia a negar las experiencias desagradables en diversidad de situaciones, fueron de fundamental importancia en el diseño de la investigación en escolares mexicanos.

El TASC, como se conoce comúnmente a la escala de ansiedad, está constituido por 30 ítems, divididos en dos secciones: la primera (del ítem 1 al 18) mide la ansiedad en relación con situaciones escolares específicas perfectamente definidas, y la segunda (del ítem 19 al 30) registra el nivel relativo de ansiedad, expresado ante estímulos relacionados con situaciones de “prueba”. La

escala total considera una serie de mediciones para situaciones en la escuela y en el hogar, tanto en la vigilia como en los sueños y fantasías del niño.

Por otra parte, el DSC consta de 39 ítemes, también divididos en dos secciones: la primera, destinada a la medición de respuestas defensivas, y la segunda, a una evaluación del grado de mentira revelado por el sujeto durante la aplicación del test.

En general, su contenido se refiere a la explicación de experiencias de ansiedad comunes, tales como hostilidad, miedo al rechazo, temor de sufrir daño, etc. Este test, a diferencia del TASC, no plantea situaciones escolares. Las variables que este test maneja son las siguientes:

- a) Para el TASC: la calificación total.
- b) Para el DSC: las calificaciones de la escala de mentira y de la escala de defensa.

Test de estimación del tiempo

La importancia de este test radica en las formas características de percibir el tiempo que pueden observarse no solamente en diferentes culturas, sino incluso dentro de una misma cultura.

En la vida diaria constantemente se oyen expresiones muy variadas acerca de “cómo vive” el tiempo. Tal vez podría decirse que, desde el punto de vista motivacional, los intervalos de gratificación y privación determinan el grado de estimación automática del tiempo.

Para determinar el concepto de tiempo lo más objetivamente posible, se necesitan utilizar conceptos operacionales y métodos de investigación estandarizados.

Así pues, mediante este test se intenta medir el desarrollo de la formación del concepto de tiempo en el niño. Esta prueba es muy fácil de aplicar. Consta de tres ítemes o estimaciones iguales, en cada una de las cuales se le pide al niño que calcule la duración de un minuto a partir de una señal que le dará el examinador.

Este instrumento fue utilizado en el IDPEM del siguiente modo: como la batería de tests de esta investigación fue administrada en dos sesiones de dos horas y media de duración cada una, al principio de la primera y al final de la misma se le pedía al niño que calculara cuánto tiempo dura un minuto; en la segunda sesión solamente se le aplicó una vez.

Las variables que esta prueba maneja son las siguientes:

1. Calificación de la duración, consistente en sumar las tres estimaciones de tiempo, considerando que las mejores serán aquellas que se aproximen más a los 180 segundos.
2. Calificación de exactitud, la cual se obtiene restando a los 180 segundos la suma de las tres estimaciones del niño; por tanto, mientras menor sea la diferencia obtenida, mayor será la exactitud de la estimación.

3. Consistencia, que es la desviación estándar de la suma de las tres estimaciones para obtener el grado de semejanza que existe entre ellas.
4. Primer ensayo.
5. Segundo ensayo.
6. Tercer ensayo.

Escala de conducta

Esta escala tiene por objeto registrar la conducta que el examinador observa en el sujeto mediante la situación de prueba, y puede utilizarse como medida de control. Está dividida en cinco áreas, que son las siguientes:

1. De confianza en la situación social.
2. De confianza en la propia habilidad.
3. De cooperación.
4. Nivel de ansiedad.
5. Atención a las tareas.

De acuerdo a estas cinco áreas, en cada una de ellas se calificará con un rango de cinco posibilidades; por ejemplo, en el área de confianza serían las siguientes:

- a) Perfectamente seguro en el trato social.
- b) Más bien confiado; buena compostura.
- c) Moderadamente confiado: ni tenso, ni con buena compostura.
- d) Más bien tímido; poco desenvuelto.
- e) Extremadamente tímido; torpe en el trato social.

Todos los instrumentos mencionados con anterioridad y de los cuales solamente se ha hecho una síntesis, han sido estudiados a través de una gran cantidad de investigaciones psicométricas, a fin de poder tener la seguridad de que funcionen debidamente para los sujetos mexicanos. En primer lugar, han sido estandarizados, obteniéndose normas de calificación que cubren un rango de edad de los 6 a los 15 años para casi todos los tests. Además, se han llevado a cabo estudios de confiabilidad en la gran mayoría de los instrumentos, obteniéndose lo siguiente:

1. Coeficiente de estabilidad temporal.
2. Coeficiente de consistencia interna.
3. Coeficiente de confiabilidad por jueces (consistencia entre calificadores).
4. Análisis de varianza, para encontrar la dispersión del comportamiento a través de las siguientes variables: edad, sexo y sistema escolar.
5. Se han verificado estudios evolutivos, utilizando las calificaciones obtenidas por cada uno de los sujetos, a través de los seis años de investigación. (Debido al diseño de la muestra, se tomaron para estos estudios solamente los tres primeros años de la investigación, a fin de comparar año con año los progresos y los cambios conductuales que se van operando en los niños, conforme éstos se desarrollan. Para estas investigaciones se emplearon pruebas de significancia, tanto paramétricas como no paramétricas, dependiendo de las variables que se manejaron).
6. Se hicieron intercorrelaciones de cada una de las variables contra todas las demás variables de todos los tests utilizados.

7. Algunos de los instrumentos (por ejemplo, el HIT, el WISC y otros) fueron estudiados mediante análisis factoriales.
8. En algunos de ellos se llevaron a cabo estudios de validación cruzada.

Finalmente, cabe hacer notar que los tests mencionados en este capítulo constituyen sólo una parte de los que se han utilizado en diversas investigaciones realizadas en el INCCAPAC. Aquellos que estén interesados en estas investigaciones pueden consultar las memorias de los diversos congresos interamericanos de psicología; un gran número de tesis profesionales (más de 50) que han empleado instrumentos y datos de las investigaciones mencionadas, y diversas obras y manuales que serán publicados próximamente.

6

LA MEDICIÓN EN EL CAMPO DE LA PSICOLOGÍA CLÍNICA

Para hablar de medición en psicología clínica tenemos, por historia, que mencionar una serie de eventos que han ocurrido y siguen ocurriendo desde el momento en que se inició la medición de lo que es el contenido de este campo de la psicología.

Para iniciar este capítulo debemos explicar qué se ha entendido tradicionalmente por psicología clínica. Desde luego, cuando hablamos de ella nos estamos refiriendo a la psicología patológica. Este concepto no sólo implica el comportamiento anormal en todos sus niveles desde el punto de vista de las alteraciones provocadas por daño, funcional u orgánico, del sistema nervioso, sino también toda alteración conductual provocada por cualquier tipo de alteración que el organismo sufra, como consecuencia de algún daño o disfunción fisiológica, física o producida por factores ambientales.

Para poder entender los diversos enfoques, lenguajes y técnicas que en el transcurso del tiempo han venido teniendo lugar en el contexto de la psicología clínica, es menester hacer un poco de historia.

Así pues, podemos afirmar que el iniciador de todas las controversias modernas acerca de si debe o no desaparecer el término “psicología clínica” del léxico del psicólogo e, incluso, de su currículo como estudiante y futuro profesional de la psicología fue Ribot, quien en 1870, en una publicación importante para la historia, enfatizó, primero que la psicología tenía que separarse definitivamente de la filosofía y aún de la metafísica; y segundo, sugirió que, en vez de perder el tiempo en disertaciones poco o nada funcionales ni solucionadoras de problemas, se enfocara la energía y la investigación al aspecto científico de lo “observable”, más que a la interpretación de “observaciones introspectivas”. Partiendo de la fisiología, Ribot vislumbró en la conducta caótica, patológica, una manera fructífera e importante de realizar experimentación científica. Siguiendo estas afirmaciones, se dedicó a la investigación en diversos campos de la psicofisiología, produciendo trabajos y obras que lo hicieran famoso, tales como: *Las enfermedades de la memoria*, *Las enfermedades de la voluntad*, *Las enfermedades de la personalidad*, etc. Profesor de La Sorbona, en París, creó una cátedra de psicología experimental y comparada. En 1896, siguiendo a Freud, centró su atención en la importancia de la *vida afectiva*, en donde puso de manifiesto la existencia de tendencias parcialmente inconscientes que, al expresarse en conducta observable, adquieren diversas formas y contenidos, dependiendo de los niveles de desarrollo en donde se localicen. Siendo filósofo, llama la atención el enfoque que le dio a la problemática psicológica: a sus discípulos les exigía no sólo la formación filosófica, sino también la formación médica. (En algunos países de Europa aún se exige el título de médico para ser psicólogo).

Obviamente, Ribot no descubrió esto por mera inspiración. En aquel tiempo existían ya algunos profesionistas, principalmente psiquiatras franceses, que hablaban de una escuela médico-

psicológica. En Alemania, Weber, Fechner y Wundt planteaban la necesidad de utilizar métodos más efectivos para hacer investigaciones cada vez más rigurosamente científicas. Por otra parte, Claudio Bernard, siendo maestro de fisiología en La Sorbona, cátedra que él mismo fundó, afirmó que “los fenómenos intelectuales podían explicarse mediante la fisiología”. En Inglaterra, H. Jackson, H. Spencer y otros investigadores interesados en estos problemas plantearon formas susceptibles para investigar estos fenómenos. Por su parte, Pierre Janet primero y luego G. Dumas, discípulos de Teódulo Ribot, siguiendo a éste, tanto en el patrón de su preparación médico-filosófico-psicológica como en sus teorías, desarrollaron nuevas ideas, algunas en oposición a las de Ribot. Las aportaciones más importantes fueron las de Janet, quien, entre otras conclusiones, llegó a la de que la neurosis no produce una afección, sino una supresión de las funciones que implican un alto despliegue de energía, y por tanto de tensión, mientras que las funciones simples son automáticas, persistentes y a veces maximizadas. Además, encontró una relación de sus hipótesis con los psicólogos estadísticos, como Spearman, quien consideró la existencia de un factor *g* como la energía intelectual por excelencia. Por otro lado, Dumas aportó su estudio sobre las emociones, a partir de todos los elementos que podía proporcionarle la fisiología. Escribió, además, gran cantidad de obras y, en 1930, publicó su *Tratado de psicología*, que es uno de los trabajos que lo han hecho que forme parte de la historia.

Debe mencionarse también a Charcot, quien, además de sus investigaciones y su desempeño como catedrático en la Facultad de Medicina de París, es el iniciador de las investigaciones sobre hipnotismo, en 1878. Este investigador era neuropatólogo. Sus primeras investigaciones estuvieron dirigidas a la detección de los estados hipnóticos por medio de índices fisiológicos, difícilmente simulables. Anteriormente, en 1778, Mesmer había puesto en práctica el tratamiento de ciertas enfermedades mentales mediante la hipnosis. Janet, detractor de la hipnosis, insinuó que los rasgos para detectarla eran fácilmente simulados. Sin embargo, Charcot puso en tela de juicio estas afirmaciones e intentó darles una objetividad científica.

En 1895, Breuer y Freud, en Viena, publicaron sus *Estudios sobre la histeria*. Breuer era fisiólogo y se dedicaba a estudiar la respiración y el sentido del equilibrio. Por mero accidente se vio en la necesidad de utilizar el hipnotismo, el cual abandonó porque su paciente, la primera que intentaba someter, se había enamorado de él.

Por su parte, Sigmund Freud, alumno de Charcot y quien habría de ser considerado más tarde como una de las figuras más sobresalientes en el desarrollo de la psicología, después de doctorarse inició su larga serie de trabajos sobre una gran cantidad de fenómenos psicológicos y formuló sus primeros estudios sobre hipnosis y más tarde su práctica clínica, en donde, mediante la utilización del método de libre asociación y a través del análisis de los sueños (tanto propios como de sus pacientes), elaboró las primeras teorías sobre el comportamiento humano, con el fin de explicar por qué sucedían ciertas cosas en el hombre que aparentemente no tenían una razón “visible”. Sus teorías revolucionarían al mundo y particularmente a sus discípulos. Fue él quien, al plantear la teoría de la naturaleza del inconsciente, dio originalidad y enriqueció con un gran número de aportaciones la idea de la existencia de tales fenómenos, que ya otros estudiosos de estos problemas habían mencionado como la “mente inconsciente”. Todos los trabajos que desarrolló Freud, además de la teoría del inconsciente, la teoría de la libido, los mecanismos de defensa, los niveles topográficos y funcionales de la personalidad (ello, yo y superyo los segundos; consciente, preconsciente o subconsciente e inconsciente los primeros), es decir, lo que él denominó *aparato psíquico*; la teoría de la transferencia, etc., para 1909 iniciaron su evolución no sólo en el maestro,

sino también en sus discípulos, aunque algunos de ellos siguieron direcciones opuestas. Para 1926 el psicoanálisis, en sus diversas modificaciones, sobre todo en aspectos teóricos, había alcanzado gran desarrollo en el mundo, sobre todo en Europa.

Siendo el psicoanálisis, en principio, un método terapéutico, propició el desarrollo de teorías de la personalidad “normal”. El método clínico, tradicionalmente aplicado a la actitud del médico frente a los pacientes que sufrieran cualquier enfermedad, fue adoptado por la psicología patológica. Parece que en su origen este método fue una reacción en contra de lo que realizaba Wundt en su laboratorio de Leipzig y Galton en el suyo.

La actitud clínica consiste en la observación paciente y profunda, a largo tiempo, de la problemática de un sujeto, a fin de desmembrarla para poder llegar al origen, estudiando cada hecho o comportamiento del individuo para luego reunirlos, con el fin de establecer la dinámica, es decir, indagar cómo se mueve o interactúa el individuo para producir determinada conducta. Todo esto, en contraste con el estudio del individuo en el ambiente del laboratorio, intenta poner de relieve la inutilidad del trabajo realizado con un ser humano en un medio de “aislamiento de la realidad”.

L. Witmer, psicólogo norteamericano, destaca dos situaciones en el campo de la psicología: la psicología clínica y el método clínico. Influido por Wundt, su maestro, este investigador se dedicó al estudio de niños anormales con alteraciones conductuales un tanto severas o difíciles de manejar. Por otra parte, Catell, también alumno de Wundt pero influido por Galton, se dedicó a estudiar el campo experimental-estadístico de la psicología en la misma época de Witmer. Este último, en 1896, en Pensilvania, fundó la Clínica Psicológica destinada al tratamiento de niños retardados y anormales. Utilizó por primera vez la frase “el método clínico en psicología”, implicando con ello un método de investigación que consistía en estudiar un gran número de individuos para poder llegar a establecer la generalidad en el comportamiento mediante sus aptitudes, ya fueran anormales o deficitarias. Con esto, determinó que las clínicas psicológicas serían los sitios apropiados donde se podría emplear este método y donde los psicólogos aprenderían a utilizar la psicología, aplicándola debidamente a los campos educativo, social, industrial y correctivo, así como al de higiene mental del individuo en general.

A partir de aquí, se hablará de psicología clínica, en su más extenso sentido, cada vez que un psicólogo se centre en la solución de los problemas conductuales que presente un sujeto en lo particular, ya sea niño, adolescente, joven, adulto o anciano, en dificultades de distintos tipos que los coloquen a la mayoría como “sujetos anormales”. (Realmente, una persona normal puede interesarle al psicólogo clínico; pero se han establecido divisiones dentro del campo general de la psicología que hacen que estas personas sean estudiadas por otras especializaciones del psicólogo, como puede ser el educativo, el orientador propiamente dicho, etc.)

El psicólogo clínico no sólo ha recibido influencia considerable del psicoanálisis. Al respecto, en 1927 aparece la obra de Healy y Bronner titulada *Manual de tests individuales y su administración*. En ella el psicólogo clínico encuentra ayuda no sólo en el método de investigación profunda de la historia del paciente, sino también en los exámenes mediante tests estandarizados que, a través de datos numéricos, y “perfiles” lo capacitan para pronosticar determinada conducta.

El método clínico suele estar influido por el psicoanálisis y por la tradición clínica; así, cuando a él nos referimos, lo hacemos de acuerdo a la concepción de Witmer, en tanto que cuando se habla de tests en psicología clínica, se les vincula con la psicometría, originada en Wundt y Galton.

Parece que actualmente ya se pueden combinar ambos métodos de “auscultación” en el campo de la psicología clínica. Puede decirse que los métodos de medición en esta área empezaron a utilizarse en 1930, pero como una actitud opuesta a la psicométrica. Los tests “proyectivos” que suelen usarse dentro del contexto del psicoanálisis, tienen un contenido muy ambiguo y se han elaborado fuera de la metodología rígida que marca la teoría de la medición en psicología. Así pues, se arguye que las manchas de tinta o las láminas con escenas que representan actividades o comportamientos hasta cierto punto cotidianos, así como otro tipo de material inestructurado le permiten al examinado tener una mayor flexibilidad para expresar el contenido profundo de su personalidad.

LOS TESTS APLICADOS AL CAMPO DE LA PSICOLOGÍA CLÍNICA

Problemas de validez y confiabilidad de los Tests psicológicos de uso clínico

Una de las múltiples funciones de todo psicólogo clínico es hacer el llamado *psicodiagnóstico*. Aún en la actualidad se cree que esta función es la única que lleva a cabo el psicólogo, en cuanto a la clínica, quien debe subordinarse a las indicaciones del psiquiatra. Los propios psicólogos desconocen que poseen una metodología y una instrumentación, de la cual carecen tanto la psiquiatría, rama de la medicina, como el psicoanálisis, método clínico para el tratamiento de algunas alteraciones conductuales, que hacen de la psicología una ciencia totalmente independiente de las demás, aún cuando tiene grandes nexos con todas ellas, dada la naturaleza del universo que estudia e investiga: el ser humano.

De tal suerte, el psicólogo clínico restringe su campo de trabajo y se concreta a hacer lo que sólo es una mínima parte de su desempeño profesional, que es el instrumental por medio del cual puede llegar a la comprensión, predicción y planeación de la reorganización de la conducta de la persona o personas bajo su responsabilidad como especialista, de la manera más rápida.

Emplear tests para advertir y comprobar los problemas que tiene una persona que acude a tratamiento psicológico, es más práctico que esperar solamente a que, a través del paso del tiempo, mediante la observación un tanto pasiva de la conducta, se pueda determinar qué es lo que ocurre, con el peligro, además, de que los juicios estén influidos por la particular manera de ver o de manejar determinado tipo de problemas por parte de quien evalúa.

Así pues, suele suceder que el psicólogo clínico vaya abandonando si no el campo, sí el uso de tests psicológicos, debido a que va advirtiendo que tiene dificultades para manejar el instrumental con que cuenta, ya que, por una parte, tal instrumental revela una baja capacidad de consistencia, y por otra, un contenido que por su ambigüedad, tanto en el dominio de ítemes como en todos aquellos aspectos que deben ser cubiertos como requisitos indispensables para considerar a un test confiable y válido, no permite dar resultados que realmente resuelvan problemas que lleven al psicólogo clínico

a establecer la instrumentación y las técnicas apropiadas para cambiar, reestructurar, entrenar, etc., al sujeto en nuevos comportamientos que le ayuden a adaptarse mejor a su situación real.

Las técnicas proyectivas se originan a partir del concepto dinámico de la personalidad. Este concepto implica que la personalidad es un proceso de organización de las experiencias y de la estructuración del espacio vital, lo cual lleva a plantear la pregunta acerca de cómo se podrá descubrir, y en qué forma, la personalidad de un individuo que organiza esas experiencias con el fin de comprender, o al menos intentar penetrar en el mundo de sus sentimientos, deseos, pautas, etc.

El psicólogo clínico se ha aproximado al estudio de la personalidad, induciendo al individuo a que revele su forma de organizar sus experiencias, proporcionándole un campo en el cual encontrará objetos o situaciones poco estructuradas y ciertos patrones culturales, para que su personalidad pueda proyectar, sobre ese campo dúctil, su manera de ver la vida, sus deseos, sus patrones conductuales y, muy especialmente, sus sentimientos. Se espera del sujeto que, al manipular el material que se le presenta, exprese su mundo privado al interpretar el campo vital que lo rodea en ese momento dado y que reaccione afectivamente hacia él.

Los métodos de proyección para el estudio de la personalidad consisten en la presentación de estímulos seleccionados y diseñados específicamente para ello, porque serán significativos para el sujeto que ha de expresar su idiosincrasia y su organización privadas.

Cuando se revisan los procedimientos para obtener la información acerca del mundo privado de un individuo, se encuentran numerosas técnicas planeadas con ese objetivo. Para estos fines se han desarrollado, así como para otros tópicos acerca de la medición, una gran cantidad de teorías sobre el significado de la personalidad, y muchas de ellas han dado lugar a la construcción de instrumentos que permitan detectar y probar sus conceptos y construcciones teóricas. Algunas definiciones han sido muy amplias, pero suficientemente ambiguas para no permitir una clara visualización de lo que se trata de entender por personalidad.

De tal suerte, ha habido muchas discusiones de la medición de esos atributos que componen lo que se ha entendido por personalidad. Para fundamentar tales discusiones, deben tenerse en consideración los siguientes aspectos: a) el hombre es un organismo que existe en el mundo común de la naturaleza; b) además, pertenece a un grupo, y como miembro de él, desarrolla el curso de su vida en el mundo social compuesto de patrones y prácticas culturalmente preescritas, y c) el hombre vive en su mundo privado, desarrollado a partir de sus experiencias personales. Así pues, el hombre es un organismo que reacciona en forma abierta y fisiológicamente a los impactos del ambiente. La actividad humana presenta problemas de observación y medición similares a los de cualquier otro organismo o evento. El átomo, por ejemplo, tiene procesos internos de los cuales sólo se pueden hacer inferencias a través de sus manifestaciones externas, como son las transformaciones de actividad y energía que se pueden ver y frecuentemente medir.

El principal problema que se presenta en la medición de esos procesos del ser humano es la existencia de las diferencias individuales. El inicio de los primeros estudios para medir algunos aspectos del individuo, desde hace muchas décadas, puso de manifiesto la variabilidad que existía entre cada individuo en relación con ciertos atributos o características de diversas índoles: anatómicas, fisiológicas, antropológicas, psicológicas, etc., que producen en la emisión de una

conducta cualquiera hacia una misma situación, diferentes formas de comportamiento de uno a otro individuo.

Desde el punto de vista de las diferencias individuales de la personalidad y dependiendo de los diversos significados que se le han adjudicado, ya que se interpreta en distintas formas y circunstancias, los psicólogos han intentado ponerse de acuerdo acerca de lo que miden, aún cuando no hayan podido decir con exactitud qué es la personalidad.

Como resultado de los diversos enfoques que se han dado a las definiciones de personalidad, ha perdurado una que la define como el conjunto de rasgos o cualidades. (Desde luego, el concepto de rasgo se ha conservado desde los días precientíficos). En primer lugar, las variables para el estudio de la personalidad a veces son poco explícitas y muy estrechas conceptualmente, y otras demasiado amplias. Al respecto, Allport y Odberg, en 1936, demostraron que existen más de 5000 palabras para describir los rasgos o cualidades de una persona, en cualquier idioma más o menos desarrollado.

Sobre estas ideas se han elaborado cientos de tests para detectar estos rasgos; sin embargo, parece que éstos han constituido medidas gruesas de tal contenido que pueden abarcar varios rasgos medidos. Además, existe el problema de que los rasgos no son totalmente independientes uno de otro, y que algunos son mucho más importantes que otros en la estructura de la personalidad dada. Debido a las dificultades que presenta el concepto de rasgo, algunos psicólogos han eliminado de su preocupación científica esos aspectos y han intentado elaborar otras formas de medir la experiencia única de cada persona. Esto, desde luego, limita la intención de construir instrumentos de medición psicológica, porque no permite la comparación entre sujetos.

La formación de los rasgos particulares de cada individuo depende de situaciones muy complejas, a las cuales debe ponerse la debida atención. Por una parte, los rasgos son, en cierta medida, productos que hay que comprender desde el marco de referencia de la genética, la herencia y las predisposiciones genéticas que tienen lugar en la formación de los rasgos. Por otra parte, en vista de que nuestro mayor conocimiento lo tenemos sobre el aprendizaje social y teniendo en cuenta que éste es aún de mayor importancia que la composición genética de los individuos para ciertos fines, tendremos que recurrir a las teorías del aprendizaje. Además, no debe olvidarse que la medida de los rasgos de la personalidad está estrechamente relacionada con las diferencias individuales.

En estas condiciones, la medida de los rasgos de la personalidad, para su comprensión, se clasifica en tres grandes grupos:

- a) Rasgos de personalidad sociales, que se explican a través de la “capa” que cubre la personalidad, es decir, la superficie de ésta con la cual el individuo se enfrenta a la sociedad.
- b) Rasgos motivacionales, que pueden traducirse en la dinámica de la personalidad.
- c) Rasgos de categorías de ajuste y de desajuste de la personalidad que, en otras palabras, son la “normalidad” y la “enfermedad”.

Se ha indicado anteriormente que, a pesar de las apariencias, estas tres categorías de rasgos no son independientes entre sí, sino que unos influyen en los otros, y viceversa.

Como los rasgos sociales son los de más fácil observación, la mayoría de los instrumentos de medición de la personalidad que han sido estandarizados tienen un contenido que denota rasgos sociales. Los otros dos sólo pueden inferirse de los resultados de la medición, si ésta se ha realizado correctamente.

Es menester aclarar que para los psicólogos, el término rasgo significa lo mismo que atributo, elemento o característica. Sencillamente, es algo que puede medirse porque está presente, en alguna cantidad, en la estructura humana.

MÉTODOS PARA MEDIR LOS RASGOS DE LA PERSONALIDAD

Se ha discutido mucho acerca de si existen o no rasgos generales de la personalidad. Esto ha sido planteado por aquellos psicólogos que sostienen las teorías ideográficas y nomotéticas de la personalidad, para lo cual es necesario hallar correlaciones entre estos rasgos y los rasgos específicos; en otras palabras, deben encontrarse correlaciones entre factores y hábitos; sin embargo, la experiencia diaria demuestra que esto no es así: no existen tales correlaciones o éstas son muy bajas. Para los teóricos que sostienen las teorías nomotéticas, existe un rasgo general de la personalidad, el cual corroboran en las correlaciones que encuentran entre muchos rasgos específicos, y esto sólo lo pueden hacer mediante la ayuda de sus análisis factoriales.

Se considera que, en última instancia, las teorías ideográficas son las más razonables, puesto que los rasgos de la personalidad se encuentran distribuidos entre los sujetos, de tal manera que la única forma de entender a un individuo es mediante la observación retrospectiva de su historia, para ver cómo es que llegó a ser como es ahora. De todas maneras, los psicólogos se hallan en desacuerdo respecto de los esfuerzos que se han hecho para tratar de medir los rasgos de la personalidad.

Contrariamente a lo que se ha visto acerca de la medición de las habilidades humanas, existen múltiples intentos de acercarse a la medición más exacta de los rasgos de la personalidad. De este modo, las formas más desarrolladas de medición son las de las funciones intelectuales. Así, se dice que los rasgos de personalidad no tienen una relación, ni siquiera mencionable, con el “qué tan bien” hace las cosas una persona dada. Las funciones intelectuales son actividades de la vida diaria que implican una conducta común y corriente, en la cual no siempre se pone de relieve abiertamente la o las habilidades intelectuales, aunque, desde nuestro punto de vista, definitivamente no se puede hacer una separación tan drástica de las conductas llamadas rasgos; por ejemplo, la hostilidad que muestra un individuo hacia las personas y la negligencia en su trabajo, evidentemente alteran en cierto grado, a veces muy elevado, su eficiencia y su éxito en el mismo. Como este ejemplo, podrían citarse muchísimos y demostrarse experimentalmente con resultados positivos.

Para medir los rasgos de la personalidad se han elaborado diversas formas de instrumentación. Estas formas dependen de cuál sea la parte de la personalidad que se desee investigar. Tradicionalmente, las más importantes se dividen en dos grandes grupos: a) tests de estructuración de la personalidad, y b) tests o técnicas proyectivas de la personalidad.

Dentro de los primeros se encuentran todos aquellos tests cuyas tareas implican, de algún modo, actividades de percepción y de organización de esos preceptos, a partir de la manipulación de materiales poco estructurados (por ejemplo, los tests de manchas de tinta o los de mosaicos).

En el segundo grupo pueden incluirse desde los tests de apercepción temática, hasta las pruebas de frases incompletas, las historias incompletas, los autoinventarios, los inventarios de personalidad, las entrevistas (como elementos complementarios, muchas veces, del material obtenido por los tests), etc. Todas estas técnicas están basadas en situaciones que se le presentan al sujeto para que abiertamente éste dé su interpretación. La proyección se refiere a la necesidad de atribuir a otros individuos comportamientos motivacionales, sociales, etc., de uno mismo, que no son agradables o aceptables –por diversas razones – por quien “proyecta”, en un intento de sentirse mejor adaptado a su ambiente.

Mucho se ha hablado acerca de la estructura de las pruebas o técnicas proyectivas. Por una parte, hay quienes han sugerido que se cambie, el término “proyectivas” por otro que denote más la interpretación de la proyección. La autora está parcialmente de acuerdo en cuanto al cambio, pero no en cuanto a la sugerencia de membrete, pues debe aceptarse que, a estas alturas, resulta difícil llevar a cabo un cambio tan drástico. La autora, al utilizar los nuevos métodos de medición de los rasgos de la personalidad, ha dado un nombre que va más de acuerdo con los procedimientos de construcción y manejo de los instrumentos, como es el de *métrica de la personalidad*, es decir, tests contruidos sobre las bases metodológicas y sobre las reglas estrictas de la medición psicológica.

Por otra parte, en la forma de clasificar las técnicas proyectivas, tradicionalmente se ha insistido en indicar que éstas están constituidas de material inestructurado, para diferenciarlas de los inventarios, autoinventarios y otros métodos, que sí son estructurados. Sin embargo, en esto no está de acuerdo la autora. Para ello arguye que las técnicas proyectivas no pueden constituir material inestructurado porque lo que se le presenta al sujeto para interpretar son situaciones hechas (por ejemplo, una tarjeta que contiene una escena en donde aparece una persona ante un escritorio, en posición de escribir). Si desde el punto de vista conceptual hay un consenso social –por tanto, público- del significado de un estímulo configurado, se dice que está estructurado; en cambio, si no hay convenio de significado público, entonces será inestructurado. Así, mientras más estructurado sea un estímulo, menos se prestará para una interpretación abierta y libre del percepto. También cabe advertir que una manera de hacer menos estructurado a un material, aunque aparentemente tenga cierto grado de estructuración, es en cuanto a lo que se le pide al sujeto que haga con él. Así, por ejemplo, si se le pide, en el caso anterior, que diga qué es lo que está pensando o escribiendo la persona que está detrás del escritorio o que diga cómo es esa persona, entonces se le está permitiendo al sujeto interpretar con más libertad lo que ve. Aquí son mucho más fácilmente detectables las diferencias individuales que cuando se emplea material altamente estructurado.

TESTS PARA MEDIR RASGOS DE LA PERSONALIDAD, MÁS UTILIZADOS EN MÉXICO

Como se sabe, el uso de este tipo de tests, en México, data de hace ya muchos años. Este uso ha sido muy difundido, pero carente de coherencia, en cuanto a las formas y razones para su empleo. Esto tiene una explicación histórica: en México, la psicología clínica era la única rama de la psicología, como profesión, todavía hace aproximadamente 15 años. Fue hasta 1960 cuando en la Escuela de Psicología, gracias a las aportaciones de nuevas técnicas y conocimientos proporcionados

por el doctor Rogelio Díaz-Guerrero y sus colaboradores, se inició una nueva era de la psicología en México al incluirse la enseñanza del método experimental y la teoría de la medición psicológica en el currículo académico. Hasta entonces sólo se conocía y se practicaba una psicología con una marcada tendencia psicoanalítica y psiquiátrica que, además, carecía de la información de todos los nuevos hallazgos en el campo de la investigación psicofisiológica, psicosocial, psicoeducativa, etc. De ahí, el uso indiscriminado de las pruebas de personalidad y de algunas pruebas de medición de las habilidades humanas que conducían a errores muy graves.

A continuación se mencionarán los tests más usuales, mismos que hoy día van dejando su lugar a otras técnicas más efectivas y más válidas. En primer lugar, como ejemplo de material altamente inestructurado, está el *Test de manchas de tinta*, de Hermann Rorschach, quien la construyó como resultado de sus investigaciones sobre la estimulación de la imaginación. Desde luego, este investigador no fue el primero ni el único que descubrió que las manchas de tinta eran un buen medio de estimulación de este tipo, pero sí fue el primero que desarrolló una técnica para emplearla en el diagnóstico de la personalidad. Rorschach desarrolló su instrumento y su método como una técnica práctica utilizable en casos clínicos para el estudio de “factores inconscientes” en la percepción y para descubrir los factores dinámicos de la conducta y de la personalidad. (No se pretende aquí analizar la construcción del Rorschach, sino simplemente establecer que no fue muy ortodoxo en su metodología). Este investigador partió de la idea de que la ejecución de una persona, en ciertas tareas, es una expresión de su personalidad total, sobre todo si su conducta está relacionada con situaciones estímulo no convencionales, libres del contexto social. Las manchas de tinta permiten, expresa Rorschach, una respuesta del individuo que revela, mediante lo que él informa que ve, lo que dice que percibe, lo cual constituye una especie de retrato o delineación de su personalidad.

El test de Rorschach está compuesto de diez láminas simétricas en lo que se refiere a las manchas de tinta. Estas se hallan distribuidas de la siguiente manera: en cinco de las tarjetas son blancas y negras, en otras dos son blancas, negras y de diversos colores, y en las tres restantes son sólo de color. Las láminas o tarjetas están numeradas del 1 al 10 y deben ser presentadas al examinado en una misma ocasión y en el orden que los números indiquen. En esta prueba se le pide al sujeto que diga o explique qué le parecen que son o qué figura representan dichas manchas. Estas instrucciones de la prueba algunas veces han sido modificadas por quienes la aplican. Así, por ejemplo, Klopfer y Kelly han dicho a sus sujetos: “Diferentes personas han visto una serie de cosas en estas manchas. Ahora dígame usted qué ve en ellas; qué podrían representar; ¿qué es lo que hace que usted piense así?”. Como puede advertirse, la primera objeción que no puede hacer a este tipo de tests es que las instrucciones pueden modificarse por todo aquel que aplique la prueba, contrariamente a lo que ocurre en la aplicación de tests de inteligencia o de aptitudes.

La prueba original de Rorschach no tiene tiempo límite, ni número específico de respuestas para cada tarjeta. El examinador debe anotar el comportamiento del sujeto cuando éste es enfrentado a cada estímulo: expresiones, registro (si es posible, exacto) de la verbalización de la respuesta, tiempo de reacción (es decir, el tiempo transcurrido desde la presentación de la tarjeta, hasta el principio de la respuesta del sujeto), longitud del tiempo entre las respuestas, total de tiempo para responder a cada lámina; posición en que colocó la lámina para dar la respuesta, indicando la extensión o minuciosidad con que el sujeto exploró la situación estímulo; movimientos extraños del sujeto, y otras conductas significativas.

Según Rorschach, los tres registros del tiempo sirven para analizar los bloqueos emocionales o la resistencia a responder de cada individuo, frente a la percepción de cada mancha en particular. Todo esto constituye la primera parte de la aplicación. La segunda parte la constituye la “encuesta”, es decir, preguntarle al sujeto en qué parte de la mancha centró su respuesta; qué parte de la mancha le hizo pensar o imaginar que podría ser lo que respondió (o sea, qué fue lo que le produjo esos procesos de asociación, si fueron aspectos totales o parciales de la mancha); su tamaño, color, sombreado, movimiento, etc., todo aquello que pueda servir para dar una calificación. En la encuesta, el sujeto suele hacer agregados a sus primeras respuestas; esto debe ser espontáneo de su parte, sin sugerencias del examinador. Este sólo debe concretarse a aclarar aquello que sea oscuro o ambiguo, a fin de facilitar la calificación. La única pregunta que debe hacer el examinador es la de la localización de los perceptos, para determinar el área exacta de la respuesta.

La calificación se basa en cuatro categorías: localización, determinantes, contenido y respuestas originales o populares.

La *localización* se refiere a las áreas que pueden ser marcadas por el sujeto para dar sus respuestas: pueden obtenerse respuestas totales, detalles grandes, detalles pequeños, detalles raros o inusuales y la utilización de los espacios blancos, que pueden ser parte de los patrones de respuesta de cada persona. De acuerdo con Rorschach, la localización revela la amplitud de la organización o desorganización perceptual del sujeto, medida en términos de una norma de percepción y habilidad para analizar el total y sintetizar las partes.

Los *determinantes* se refieren a las características de las manchas que han sido percibidas por el sujeto, y son las cualidades o aspectos de las manchas que han producido la respuesta, tales como forma, sombreado, color, movimiento, perspectiva o combinaciones de éstos. En el uso de este test, los principios de Rorschach han sido modificados y sustituidos por otros, como lo han hecho Klopfer y sus colaboradores, S. J. Beck y otros. Esto vuelve a poner de relieve la falta de objetividad del instrumento, ya que tanto Rorschach como sus modificadores siguen sus propias ideas al manejar este test.

El *contenido* es la categorización de respuestas. Este se clasifica de acuerdo con grupos de pertenencia común: plantas, animales, personas, objetos naturales físicos, objetos hechos por el hombre, fenómenos naturales, respuestas sobre aspectos anatómicos, sexuales, etc. El contenido no sólo se incluye en grupos, sino que también de ahí se toman para explicar los significados, actitudes, intereses y algunos complejos de la persona. A veces, éstos se interpretan psiquiátricamente, otras psicoanalíticamente, y otras más de acuerdo a quien examina. (Aquí cabe volver a hacer la misma crítica: al calificar hay una gran tendencia a subjetivizar.)

Las *respuestas originales o populares* han sido discutidas fuertemente por aquellos que utilizan el test de Rorschach: no hay un total acuerdo acerca de cuáles respuestas deben calificarse como populares y cuáles como originales. Esto ya podría haberse resuelto estadísticamente.

Los datos obtenidos de las respuestas del sujeto (es decir, la calificación) tienen como fin principal conocer la adaptación general del mismo: si es la adecuada o si existen dificultades psicológicas. Dicen los especialistas en el Rorschach que si bien es difícil la administración del test, más difícil es su interpretación. Evidentemente, esto es cierto, puesto que en la utilización del instrumento no existe una regla o método específico que permita objetivizar los resultados, lo cual no sólo obedece a la negligencia de quienes lo emplean, sino también a algunos aspectos de

confiabilidad del instrumento. (Aquí la confiabilidad se refiere a la “longitud” del test y a la posibilidad de control de las respuestas).

Un tanto en contraposición con el test de manchas de tinta de Rorschach, se puede hablar de la técnica de manchas de tinta de Wayne H. Holtzman. En 1954 se inició en la Universidad de Texas una amplia investigación a fin de detectar, con la mayor exactitud, las limitaciones psicométricas del Rorschach. Con este fin, se inició la producción de una serie de láminas que contenían manchas de tinta. Estas se hacían colocando combinaciones de tinta de colores en una hoja, al centro y doblándola por la mitad, para que al presionarla se extendiera la tinta; así se producían formas ambiguas e inestructurales que, una vez secas, iban almacenándose para formar parte de los estímulos que posteriormente habrían de ser probados para determinar su poder discriminativo, en relación con la percepción y organización que expresaba cada sujeto examinado. Las 135 manchas de tinta que finalmente se consideraron como las mejores, después de los análisis pertinentes y requeridos para elaborar un instrumento de este tipo, fueron retocadas por un dibujante profesional para que intencionalmente variaran en color, sombreado, forma y simetría. Además, esto permitió elaborar dos formas paralelas (es decir, sustituibles) que no sólo tuvieran el mismo número de manchas, sino también que fueran correlativas unas de otras. De este modo, quedó elaborado un test con formas paralelas A y B, conteniendo cada una 45 láminas, más otras dos de prueba o entrenamiento; es decir, un total de 47 láminas que el sujeto tendría que examinar.

En este test, el sujeto, a diferencia del Rorschach, sólo puede dar una respuesta; inmediatamente después se hace la encuesta sobre la localización y sobre las razones que el sujeto tuvo para emitir su respuesta. Por último, se le pregunta al sujeto si quiere agregar algo más acerca del precepto. (La respuesta sólo será considerada si tiene relación con la respuesta primaria).

En esta nueva forma de investigación sobre rasgos de la personalidad se tienen en consideración, para cada una de las respuestas, no sólo el calificarlas a través de 22 variables que constituyen los aspectos que se intentan medir con ella, sino también el desarrollar una nueva forma de calificación *objetiva*, es decir, dimensionalizando o cuantificando cada variable, para superar las deficiencias del Rorschach y de otros sistemas anteriores y para obtener resultados válidos y confiables.

El test de manchas de tinta, de Holzman (H.I.T.), maneja variables que fueron tomadas parcialmente por su autor, de las categorías generales del Rorschach, y las desarrolló con mayor extensión, más visión y más analíticamente. Estas variables son las siguientes:

1. Tiempo de reacción.
2. Rechazo.
3. Localización.
4. Espacio.
5. Forma definida.
6. Forma apropiada.
7. Color.
8. Sombreado.
9. Movimiento.
10. Verbalización patognomónica, que se clasifica en:

- a) Fabulación;
- b) Fabulación combinada;
- c) Respuestas raras;
- d) Incoherencias;
- e) Lógica autística;
- f) Contaminación;
- g) Autorreferencias;
- h) Deterioro del color, e
- i) Respuestas absurdas.

- 11. Integración.
- 12. Humanas.
- 13. Animales.
- 14. Anatómicas.
- 15. Sexuales.
- 16. Abstractas.
- 17. Ansiedad.
- 18. Hostilidad.
- 19. Barrera.
- 20. Penetración.
- 21. Balance.
- 22. Populares.

Como se indicó anteriormente, cada respuesta deberá pasar por este matiz de 22 variables, cada una de las cuales ha sido dimensionalizada asignándosele números, a fin de indicar las cantidades en que pueden aparecer en cada individuo, las cuales van del 0 al 4. Por supuesto, cada uno de los números asignados implica la cantidad o carga que en cada sujeto existe de estas variables o rasgos que estructuran su personalidad. Así, la puntuación variará de 0 a 45 en peso unitario, es decir, sin la transformación a calificaciones derivadas. Estas calificaciones crudas se obtienen de la simple suma de las calificaciones que ha logrado el sujeto a través de las 45 láminas; sin embargo, es necesario efectuar la transformación a calificaciones derivadas, con el fin de obtener una información más válida acerca del sujeto. También se advierte que pueden presentarse calificaciones combinadas, lo cual constituye una de las reglas que deben observarse cuidadosamente, o sea, la de que haya dos o más variables al mismo tiempo.

Se han llevado a cabo numerosos estudios con este test, a fin de obtener la confiabilidad y objetividad de los resultados, así como la validez de índices importantes de la personalidad del individuo, de su pensamiento y de su forma de percibir el mundo que lo rodea.

Para obtener la efectividad del instrumento en cuanto a su calificación, se han hecho múltiples estudios de los denominados *confiabilidad de jueces*. Esta confiabilidad sólo puede lograrse a partir de la formulación rigurosa de los criterios de calificación para cada una de las variables; es la única forma de obtener un acuerdo legítimo, confiable y válido de los datos extraídos del examen.

Por otra parte, para esta prueba se han obtenido los coeficientes de validez de construcción, de contenido y de predicción, lo que nos hace pensar, a diferencia de los rudimentarios estudios

realizados con la prueba de Rorschach que, independientemente de las múltiples deficiencias que este instrumento pueda tener, es menester construir una prueba más completa o al menos modificar las que existen, a partir de futuros desarrollos en la medición psicológica.

Debe hacerse hincapié en que todas las pruebas estadísticas, psicométricas y psicológicas a que se ha sometido el test de Holtzman no sólo se han realizado, como en la mayor parte de los demás instrumentos de medición, en los Estados Unidos, sino que también en México se han hecho estudios profundos para demostrar que aquí pueden utilizarse, con plena seguridad, como instrumentos confiables y válidos en una parte considerable de la población.

En México, los primeros estudios con la prueba de manchas de tinta de Holtzman se iniciaron en 1964, en lo que originalmente fue el Centro de ciencias del Comportamiento de la UNAM. Desde luego, como es usual, se realizaron los estudios pertinentes con el fin, primero, de adaptar la prueba a la situación particular, para posteriormente iniciar estudios de estandarización, confiabilidad y validación del instrumento, a fin de hacerlo útil en la aplicación clínica y en la de desarrollo de la personalidad de sujetos mexicanos.

Mediante la prueba de Holtzman se llevaron a cabo estudios longitudinales para observar el desarrollo de la personalidad del escolar mexicano, enriquecidos con la utilización de un gran número de instrumentos de medición, entre los cuales figuran, a groso modo, los tests de desarrollo perceptual, de desarrollo conceptual, de desarrollo intelectual y, en general, de desarrollo cognoscitivo que han permitido investigar, mediante intercorrelaciones, análisis factoriales y análisis de varianza, tanto el desarrollo de los rasgos de la personalidad que mide el instrumento aludido, como las interacciones entre las diversas variables medidas por éste y por los demás instrumentos.

Por otra parte, tal como lo hizo el autor en los Estados Unidos, en México se han establecido normas de calificación para grupos específicos, porque no sólo se ha trabajado con las muestras preestablecidas sino que también ha habido derivaciones para estudiar otro tipo de muestras muy importantes.

Las muestras originales sobre las cuales se llevaron a cabo los primeros estudios quedan circunscritas, en primer lugar, para escolares mexicanos del Distrito Federal. Además se han hecho estudios con muestras de sujetos que presentan diversas patologías. La muestra de escolares del Distrito Federal cubre 12 años de desarrollo de la personalidad del escolar mexicano.

Estas muestras fueron obtenidas acuciosamente, a fin de lograr la representatividad de la población. La investigación longitudinal se inició de la siguiente manera:

Las muestras estuvieron constituidas inicialmente por sujetos que se encontraban cursando el primero y cuarto años de primaria y el primero de secundaria. Estos sujetos fueron estudiados a lo largo de seis años; de tal suerte, al finalizar este tiempo, los sujetos que en un principio se hallaban cursando el primer año de primaria, estaban ya en sexto de primaria; los que estaban en cuarto de primaria, iban terminando el tercero de secundaria, y los que se encontraban en primero de secundaria estaban ya en tercero de preparatoria, obteniéndose al final una muestra "superpuesta". Como es lógico en todo estudio longitudinal, para cada grupo los sujetos eran los mismos, a los cuales se les aplicaba la batería de tests constituida por la prueba de manchas de tinta y otros instrumentos entre doce y quince más, cada año en la fecha en que cumplían la edad original en que

por primera vez fueron estudiados; es decir, para los de primer año, siete años ocho meses; para los de cuarto año, 9 años 8 meses, y para los de primero de secundaria 12 años 8 meses, con una flexibilidad en el tiempo de aplicación no mayor de 15 días antes, hasta 15 días después de cumplir la edad más ocho meses. En aquellos casos en que por razones imprevisibles estas fechas variaban, el sujeto era eliminado de la muestra. Afortunadamente, estas situaciones se presentaron rara vez.

La descripción anterior sólo se consideró para el estudio longitudinal. Para la obtención de normas de calificación (es decir, para la estandarización de este instrumento y de todos los considerados) se obtuvieron muestras similares de los años intermedios, o sea, segundo, tercero, quinto y sexto de primaria, segundo y tercero de secundaria, y segundo y tercero de preparatoria. Esto tuvo que hacerse así, ya que, como se sabe, desde el punto de vista metodológico psicométrico no es válida, para efectos de estandarización, la aplicación de instrumentos de medición a una misma muestra de sujetos (como lo es una muestra para estudios longitudinales), ya que existen variables que interfieren en la validez de los resultados como puede ser, fundamentalmente, el aprendizaje, a pesar de que existan formas paralelas de los tests y tiempo suficiente entre cada aplicación (un año, por ejemplo) para producir cierto grado de olvido.

Las pruebas a que se sometió el test de Holtzman fueron las siguientes:

1. Obtención del coeficiente de consistencia interna.
2. Coeficientes de confiabilidad entre calificadores.
3. Análisis factorial, considerando el estatus socioeconómico y utilizando el método Varimax de Kaiser.
4. Obtención del coeficiente de estabilidad temporal (test-retest).

Con estas pruebas se obtuvieron, en general, resultados menos altos en México que en los Estados Unidos respecto a los grupos por edad, sobre todo en los grupos menores, incrementándose la estabilidad temporal conforme transcurrió el tiempo. Esto ocurrió en la mayor parte de las variables, lo cual implica la variabilidad de los rasgos conforme el sujeto iba aumentando de edad.

Las normas de calificación para esta prueba fueron manejadas en forma de rangos-percentiles en los Estados Unidos. En México, en el transcurso de este año, podrán estar al alcance de los interesados en la medición de la organización de los rasgos de la personalidad, a través de este instrumento, las normas ya en elaboración.

INVENTARIOS DE PERSONALIDAD

Cabe hacer un poco de historia acerca de los inventarios, como instrumentos de medición de la personalidad: tuvo su origen, atribuido a R. S. Woodworth, a principios de este siglo y empezó a desarrollarse durante la Primera Guerra Mundial. El fin primordial de su construcción fue el poder seleccionar a aquellos soldados norteamericanos “mentalmente hábiles”, para enviarlos a Europa.

La estructura de los inventarios le exige al sujeto que responda a aquellas preguntas o afirmaciones directas relacionadas con el concepto que él tenga de sí mismo, y sobre lo que opina de otros tópicos. Los inventarios pueden tener diversos contenidos: pueden ser de hábitos, de sentimientos, de inquietudes, de preferencias, etcétera. Los ítems deben ser presentados generalmente en forma de afirmaciones y pueden referirse a la primera, a la segunda y hasta la

tercera persona del singular; sin embargo, debe tomarse la decisión acerca de cuál de las tres ha de utilizarse. No es recomendable hacer combinaciones en un solo instrumento o, al menos, en una misma escala. Las posibilidades de respuesta son limitadas: pueden ser “sí o no”, “falso o verdadero”, “acuerdo o desacuerdo”, etcétera.

Se considera que este tipo de instrumento de medición es tan completo como cualquier otro, siempre y cuando, primero, sea manejado objetivamente, es decir, que se base en criterios objetivos, claros y precisos que impidan la distorsión de interpretaciones subjetivas o personales del psicólogo que lo maneje, y segundo, el psicólogo que utilice estos instrumentos realmente debe saber psicología, a fin de que pueda entender el instrumento y sus resultados.

A continuación se describirán brevemente algunos de los inventarios más útiles y conocidos, sobre todo en el medio mexicano.

Inventario multifásico de la personalidad de Minnesota

Los inventarios de personalidad resultan útiles si son manejados dentro de sus límites por psicólogos calificados. Así pues, el MMPI, Inventario Multifásico de la Personalidad, de Minnesota, elaborado por Stark R. Hathaway y J. C. McKinley, es sumamente ambicioso en cuanto a sus resultados. Los autores intentaron proporcionar en una sola prueba información suficiente sobre los aspectos más importantes de la personalidad. El psicólogo clínico que desee obtener información sobre aquellos rasgos comunes en la anormalidad psicológica, puede hacerlo mediante este instrumento.

El rango de aplicación de este inventario va desde los 16 años de edad hasta la adultez, en sujetos que pueden ser capaces de leer. Está constituido por 550 afirmaciones, a las que hay que responder, mediante el método de “elección forzada”, en una hoja especialmente diseñada para ello. Sin embargo, si el sujeto insiste en responder: “No puedo decir”, se le dejará en libertad de hacerlo, aún cuando antes se agotarán todos los medios para que esto no ocurra. El contenido de las afirmaciones incluye situaciones físicas y morales, así como actitudes sociales.

Los ítemes se han clasificado en 26 áreas, por ejemplo, salud general, sistema gastrointestinal, sistema marital y familiar, religión, afectos, fobias, intereses masculinos, intereses femeninos, etcétera. Los reactivos han sido agrupados en escalas separadas para calificar nueve rasgos de personalidad, a saber: hipocondriasis, depresión, histeria, desviaciones psicopáticas, intereses masculino-femeninos, paranoia, psicastenia, esquizofrenia e hipomanía. Además, contiene una escala denominada *escala social*, la cual se desarrolla posteriormente para detectar las relaciones sociales del sujeto.

Este inventario se construyó con el fin, casi exclusivo, de establecer la diagnosis diferencial de problemas patológicos. Además, contiene la escala “K”, cuya función es la de validar las demás escalas: es un “factor de corrección”. Esta escala contiene 30 ítemes, que permiten diferenciar a los casos “clínicos” que aparecen como normales en las otras escalas, de los realmente normales.

Otra escala de igual importancia es la “F”, constituida por 64 ítemes, cuyo contenido es una combinación de reactivos que revelan síntomas que difícilmente pueden ser detectados en un enfermo mental. Estos reactivos más bien denotan a una persona poco cuidadosa o que no ha entendido las instrucciones.

Hay también la escala “L”, cuyo contenido está relacionado con las acciones o conductas socialmente aceptables y a cuyos ítemes las personas difícilmente responden con la verdad. Aquí, la persona que marca un número considerable de ítemes anula, de hecho, el inventario, ya que se supone que, así como falseó la escala “L”, también ha falseado las demás escalas.

Las calificaciones del MMPI pueden graficarse produciendo un perfil que permita, rápidamente, observar cómo han quedado distribuidas las calificaciones de las nueve escalas de enfermedad y las de las cuatro escalas de validación. Finalmente, las normas se han obtenido en calificaciones *T* (recuérdese que son calificaciones estándar derivadas).

Para interpretar el perfil, se han elaborado diversos métodos, aunque cabe hacer notar que es requisito indispensable ser psicólogo clínico experimentado, para poder interpretar los resultados, cualquiera que sea el método empleado.

La validez del MMPI se ha estudiado profundamente, no sólo en los Estados Unidos, sino también en un gran número de países, ya que ha tenido una considerable difusión.

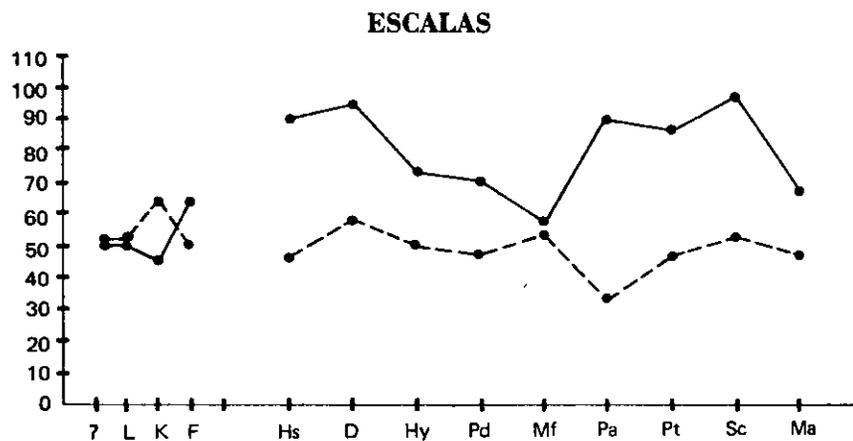


FIGURA 16. Perfil de una persona adulta normal (punteado) y de una persona típicamente psicótica (líneas negras). (Tomado de Nunnally: *Tests and measurement*, pág. 327.)

En general, después de múltiples investigaciones tanto con el MMPI como con todo tipo de inventarios, se ha llegado a la conclusión de que lo que determina la respuesta del sujeto son disposiciones o estilos particulares mediante los cuales aquél analiza y evalúa el contenido del ítem, a fin de elegir su respuesta.

Algunos autores distinguen a la disposición, del estilo. Así, afirman que las disposiciones funcionan en relación con el contenido, y que los estilos lo hacen en su ausencia (Leonard G. Rorer, 1965). Parece que en los sujetos que se examinan con inventarios existe una disposición a “estar de

acuerdo” (es decir, de condescendencia a aceptar las instrucciones), pero con una tendencia muy marcada hacia cierto tipo de respuestas.

Si se revisa la bibliografía específica acerca del MMPI, se encontrará que la gran mayoría de los estudios realizados con ese inventario han llevado a obtener mayor seguridad en su uso, ya que, a pesar de las diferencias que se han podido encontrar en cuanto a su construcción, validez y confiabilidad al ser conocidas, se ha tratado de corregirlas o de minimizarlas en las formas pertinentes y adecuadas. Cuando esto no ha sido posible, el psicólogo experto ha tomado con reserva los resultados de la aplicación del test.

En el Distrito Federal, desde hace más de seis años, el doctor Rafael Núñez Ovando ha estado trabajando con sus colaboradores, en la estandarización del MMPI, y en estudios particulares sobre aspectos importantes de este inventario (parece que aún no se han terminado). Se espera que salgan a la luz próximamente para que se pueda tener al alcance un instrumento que ha sido, y seguirá siendo, utilizado ampliamente en México por reunir múltiples ventajas. Lo ideal sería tener baremos propios, para que los resultados fueran más válidos, dentro de las limitaciones del test. Puede considerarse a este instrumento como uno de los que forman parte de la métrica de la personalidad.

Inventario de 16 factores de la personalidad De Raymond Catell

Igual que el anterior inventario, éste pertenece al grupo de pruebas consideradas dentro de la métrica de la personalidad. Está constituido por 15 factores “temperamentales y dinámicos” y un factor de inteligencia general. Desde hace 17 años se empezó a utilizar. Se ha traducido a varios idiomas y su uso es muy común en la investigación fundamental de la personalidad.

Raymond Catell es un investigador internacionalmente conocido, sobre todo por sus estudios realizados en los últimos veinte años, por el número de pruebas que ha elaborado y por las proyecciones que, en su momento, tendrán sus disertaciones teóricas; sin embargo, ha sido poco entendido por quienes se han dedicado a revisar sus trabajos.

Para Catell, el objetivo primordial de la psicología es poder establecer leyes que regulen la conducta, a fin de lograr las metas finales: predecir y controlar el comportamiento humano. A lo largo de sus investigaciones, ha cambiado, sustituido o eliminado muchas de sus afirmaciones, en un intento de depurar sus ideas presentes en relación a las anteriores. Como su teoría sobre la predicción es la más relevante en sus estudios, su definición de la personalidad se ve influida por ella. Al respecto dice: “La personalidad es aquello que permite predecir lo que una persona hará en determinada situación” Enfatiza que deben considerarse cuidadosamente las variables motivacionales, y subraya que la personalidad es un producto biológico y social, que si no puede medirse, cuantificarse y demostrarse, ha de cambiársele el nombre.

Para este investigador, lo importante es, dentro del campo de la psicología, ir de lo analítico a lo global y que, mientras en la psicología clínica no se incluya el control de la medición ni los estudios estadísticos, no podrán formularse hipótesis a verificar, sobre los cambios conductuales. De este modo, emplea dos tipos de análisis factorial en sus estudios: uno lo lleva a establecer las

diferencias entre sujetos, es decir, sus variaciones (es la técnica *normativa* o “R”), y la otra *ipsativa* o “P”). Asimismo, Catell se ha destacado por emplear la técnica multivariada, en vez de la bivariada. Así, prefiere obtener tantos factores como sea posible, a fin de rotarlos después para obtener una estructura sumamente sencilla.

Este investigador ha construido diversos tipos de tests, con el fin de recopilar la información que necesita para seguir desarrollando su teoría y sus instrumentos principales. De tal suerte, por una parte utiliza una historia del individuo que lo sitúe en lo que es realmente su vida diaria; emplea también cuestionarios en busca de la información, como ocurre con el 16 P.F. Además, ha utilizado pruebas objetivas, a fin de evitar distorsiones, sensibilidad del examinador o del examinado, actitudes de simulación por parte de quien resuelve la prueba, etcétera. Incluso, ha llegado a obtener medidas de la personalidad mediante procesos fisiológicos con productos emocionales, es decir, mediación del ritmo cardiaco, de la presión arterial, de la presión del puño al ejecutar ciertas tareas, análisis de orina, etcétera, y ha comparado los resultados con una prueba simple de papel y lápiz que se les pide a los sujetos sometidos a la investigación.

Para Catell, el rasgo es una estructura mental, una inferencia obtenida a partir de la conducta explícita, que permite la estimación de la constancia y equilibrio de tal conducta. De tal suerte, afirma que hay rasgos únicos y rasgos comunes. Los primeros, como su nombre lo indica, son particulares para cada sujeto y no pueden detectarse en otro; los segundos son aquellos que todo sujeto posee por el solo hecho de pertenecer a un medio social similar. Además, este autor afirma que hay rasgos únicos propiamente dichos y rasgos intrínsecamente únicos: los primeros son aquellos que están determinados por la influencia de elementos externos e internos del organismo, y los segundos son genuinamente diferentes y exclusivos de un individuo.

Otro tipo de rasgos considerados por Catell son los superficiales, o sea, aquellos factores que pueden ser agrupados y a los cuales debe asignárseles un nombre cuando aparecen más de dos, mismos que al establecer correlaciones entre ellos, exponen un coeficiente superior a 0.60, lo que implica que en este grado son semejantes, pero superficiales, porque tal semejanza se encuentra en la superficie, o sea, es una semejanza de *facie*.

Asimismo, Catell considera entre los tipos de rasgos a los llamados rasgos “fuente”, los cuales requieren del análisis factorial para poder detectarlos. Son más profundos que los superficiales. Dice este autor que son “los promotores de la influencia estructural real, que forma la base de la personalidad”. (Los rasgos superficiales se derivan de los rasgos fuente). En forma general, los rasgos fuente pueden ser, según Catell, rasgos ambientales o rasgos constitucionales o hereditarios. Del mismo modo, este investigador hace otras clasificaciones de rasgos: de habilidad, temperamentales, dinámicos, etcétera. Sin embargo, entre todos ellos, Catell considera muy importantes a los rasgos fuente, como posibilidades para la medición de la personalidad. Así, los obtiene de la siguiente manera: a) mediante registros de vida; b) a través de cuestionarios, y c) por medio de pruebas objetivas.

Del registro de vida, Catell deriva el inventario de los 16 factores de la personalidad, o sea, el 16 P. F., que para conocerlo y entenderlo se necesita estudiar más de un libro y tener una considerable experiencia en medición. El 16 P. F. abarca, según Catell, la esfera de la personalidad, factorializada, con una carga específica de tipo factorial para cada uno de los ítemes. Este inventario se originó de un dominio de muestras de ítemes extraídos del análisis de encuesta de opinión y de

escalas de valores e intereses, agregando nuevos ítemes que había en reserva para incrementar el valor de los 14 factores hallados en la factorialización de la conducta estudiada. La primera factorialización ratificó el número y tipo de factores, que fueron 20; sin embargo, fue necesario eliminar tres o cuatro, por la ambigüedad de su presencia. Así quedó construido el 16 P.F. con un factor de inteligencia y con 15 factores de la esfera total de la personalidad más válidos. Estos factores son:

1. Factor A:ciclotimia o esquizofrenia.
2. Factor B: inteligencia.
3. Factor C:fuerza del yo.
4. Factor D:excitabilidad o inseguridad.
5. Factor E:dominancia o sumisión.
6. Factor F:impetuosidad o decaimiento.
7. Factor G:fuerza del superyó.
8. Factor H:atrevimiento o timidez.
9. Factor I: sensibilidad o realismo.
10. Factor J:depresión (astenia) o acción entusiasta.
11. Factor K:conformismo o reacción (relacionados con la cultura).
12. Factor L:tendencias paranoicas y tensión, o relajamiento interior.
13. Factor M:decisión o indecisión (limitación de intereses).
14. Factor N:agudeza o ingenuidad
15. Factor O:confianza o desconfianza (o culpabilidad infundada).

Desde luego, esta es una presentación muy general de este instrumento, el cual, como puede fácilmente observarse, está complejamente estructurado. Naturalmente, quien se interese en su conocimiento y manejo deberá recurrir a las fuentes originales, para mayor seguridad.

En México ya se han llevado a cabo varios estudios preliminares de estandarización; por esta razón, se ha mencionado en este texto. Se resumirán brevemente las investigaciones y sus resultados; no obstante, cabe reiterar que éstas sólo son investigaciones preliminares y que los datos aún están sujetos a análisis más profundos y a procedimientos más completos acerca de la información que producen. Sin embargo, debe subrayarse la importancia de estos estudios, pues han aportado valiosos datos que pueden establecer su uso y sus limitaciones.

En 1969, Lucy Reidl realizó el primer estudio de estandarización en estudiantes adolescentes, pertenecientes a una escuela bilingüe y a otra mixta, y de nivel socioeconómico medio-alto y alto. El rango de edades fue de 12 a 18 años, y la muestra fue de 365 sujetos (204 del sexo masculino y 161 del femenino) que cursaban entre el séptimo y el doceavo grado escolar.

En resumen, los resultados obtenidos fueron los siguientes:

- a) La prueba mostró capacidad discriminativa por sexo.
- b) Las diferencias sexuales fueron evidentes en 11 de 14 factores posibles.
- c) Las diferencias sexuales, a pesar de haber establecido claramente el nivel socioeconómico, mostraron una orientación cultural, más que de otro tipo, debido a las funciones y actividades que la cultura permite a cada sexo, aunque

pudo observarse que se trataba de un grupo con una orientación muy definida al cambio social y cultural.

Reidl estableció las diferencias entre la muestra que utilizó en México y la similar (aparejada) empleada por Catell en los Estados Unidos. Se controlaron las variables edad y sexo, y la cultura se dejó como variable independiente. Esta investigadora encontró diferencias culturales definitivamente marcadas, tanto en los grupos femeninos como en los masculinos.

Deberá revisarse el trabajo de Reidl, a fin de obtener una información más completa. Aquí sólo cabe decir que en México, las diferencias entre hombres y mujeres fueron mucho más significativas que en los grupos norteamericanos; por tanto, se puede inferir que tales diferencias están determinadas por los papeles femenino y masculino en cada cultura.

En 1970, la psicóloga María Eugenia Juárez llevó a cabo una investigación con estudiantes de la carrera de psicología de la UNAM. Al respecto, esta psicóloga dijo: “Aunque este grupo fue de naturaleza muy específica, en relación con el de Catell, se intentó aparear las muestras en cuanto a escolaridad y edad promedio”. Se estudiaron 280 alumnos del primer año de psicología, que representaban el 38.34% de la población total, con una edad promedio de 21 años. El 95.74% de la muestra fueron solteros: 152 del sexo femenino y 58 del masculino.

Los resultados fueron los siguientes: en dos de los factores, C y E, hubo diferencias significativas. El C se vio más elevado en las mujeres (la fuerza del yo, estabilidad emocional y tranquilidad) y el factor E, en el sexo masculino, se colocó en el punto medio; es decir, hubo un participación de ambos extremos del factor sumisión contra el de dominancia, en tanto que el grupo femenino mostró un decremento que implicó tendencia a la sumisión, a la docilidad y al conformismo. Se compararon los grupos de Juárez y Catell, encontrándose diferencias sexuales perfectamente definidas tanto en los hombres como en las mujeres; pero en diversos factores se ratificaron en gran medida los hallazgos de Reidl.

También en 1970, la psicóloga María D. Flores Reyes, utilizando una muestra que variaba en nivel socioeconómico, educación y religión, analizó diversas variables de la personalidad en sujetos de tercer grado de secundaria y sexto de bachillerato, en cinco escuelas del Distrito Federal. El rango socioeconómico fue de clase media baja a clase media alta, del sexo femenino entre 14 y 18 años de edad, pertenecientes a escuelas de educación laica, católica y judaica.

El resultado obtenido fue que no hubo diferencias significativas entre los cinco grupos, tal vez por tratarse de personas del mismo sexo, cultura y nivel escolar. En sus conclusiones, este estudio vuelve a ratificar tanto la tesis de Catell como la de Reidl y Juárez, acerca de la influencia de factores culturales en la conducta medida con este instrumento; además, enfatizó la edad por la que atravesaban los grupos estudiados.

En 1971, la psicóloga Rosa María Cravioto realizó un estudio para establecer el coeficiente de estabilidad temporal (test-retest). Primeramente utilizó una muestra de estudiantes de la Escuela de Educación Física de Puebla, en quienes aplicó las formas A y B. Este estudio tuvo que desecharlo Cravioto, pues las variables controladas en el test no pudieron ser manipuladas del mismo modo en el retest, además de que los resultados presentaban varias dudas. Posteriormente, esta investigadora volvió a insistir en la obtención del coeficiente de estabilidad temporal, empleando

para ello una muestra de estudiantes de psicología de la UNAM y aplicándoles sólo la forma B. Los sujetos de la muestra fueron estudiantes del cuarto semestre de la carrera de psicología, cursos matutino y vespertino, de ambos sexos y de diversos niveles socioeconómicos, a quienes se les aplicó el retest un mes después del test. El rango de edades osciló entre 19 y 27 años con un promedio de 21 años, y la muestra total fue de 43 sujetos. Los resultados fueron los siguientes: en la mayor parte de los factores se observaron correlaciones significativamente importantes (0.01 en trece de ellos, 0.05 en otros dos y en el factor N no hubo). Además, se trabajó sobre la consistencia interna de la prueba para obtener el coeficiente, mediante el método de las mitades, utilizando dos procedimientos: uno, teniendo en consideración las mitades formadas por diferentes números de ítems, en vez de tener en cuenta el número de reactivos por factor; y el otro, tomando el número de reactivos de tal manera que quedaran equilibrados en mitades iguales para cada factor. Se hizo todo esto para cada una de las aplicaciones. En el segundo método se obtuvieron correlaciones significativas al 0.01 en todos los factores. En el primer método, solamente en algunos factores hubo diferencias significativas al 0.05 en tres, y al 0.01 en dos; en los demás no hubo correlaciones importantes.

En este año, la psicóloga Marisol Orcajo realizó un estudio para observar el comportamiento de los factores de la prueba de R. Catell a través de la edad y de la escolaridad. Así, comparó la influencia del nivel socioeconómico, utilizando los datos de Reidl y los de sus propias muestras. Los resultados fueron los siguientes:

- a) Se observaron diferencias significativas estadísticamente en 14 de los factores de la prueba de Catell, a través de las edades de 12, 13, 14, 15 y 16 años.
- b) En algunos de los factores se presentaron diferencias significativas a partir de la escolaridad, en tanto que en otros no las hubo.
- c) Al comparar estos datos con los de Reidl, vuelve a ratificarse el hecho de la no existencia de diferencias estadísticamente significativas entre ambas muestras en cuanto al sexo (femenino), ni en edad, ni en escolaridad.

Al revisar cuidadosamente cada una de las investigaciones mencionadas, puede afirmarse que la prueba de los 16 P.F. de Catell es un instrumento capaz de obtener características de la personalidad en aquellos sujetos en los que sea necesario utilizarlo.

Inventario de la personalidad de D. N. Jackson (Personality Research Form, P.R.F.)

D. N. Jackson, a partir de la hipótesis acerca de los modernos conceptos de la personalidad y de la medición psicológica, construyó su prueba, teniendo como objetivos fundamentales el desarrollar escalas de la personalidad y tener en reserva un gran número de ítems que permitan, en un momento dado, planear y llevar a cabo investigaciones sobre la personalidad, a fin de medir los rasgos más relevantes que se observan en escuelas, universidades, clínicas, centros de orientación y en el campo del trabajo. Este inventario constituye una aplicación práctica y sumaria de todas las teorías e investigaciones en el área de la personalidad, en su medición, y un ejemplo de cómo debe construirse una prueba de personalidad.

El inventario de la personalidad, de Jackson, estudiado en México por vez primera en el INCCAPAC, está basado en las teorías de Henry A. Murray, para quien la personalidad es el órgano

gobernante, comprometido desde el nacimiento hasta la muerte, en la ejecución de operaciones funcionales de transformación. Es ella la que establece orden e integración entre el medio externo y la estructura interna del individuo. “Sus funciones, dice Murray, son: ejercitar los procesos de desarrollo de la personalidad misma, expresarse, aprender a generar y reducir tensiones, disminuyéndolas para resolver conflictos, utilizando mediadores que satisfagan las necesidades insistentes, disminuyendo la tensión. Una necesidad es una construcción o concepto hipotético que se mantiene por una carga de energía. Puede ser provocada por procesos internos o por exigencias externas al organismo, y va siempre acompañada de un sentimiento o emoción, adoptando sus peculiares estilos para anticipar su dirección y su grado de exigencia.”

Para Murray, las necesidades son: a) primarias; b) manifiestas y latentes; c) difusas y focales, y d) proactivas y reactivas. Además, este investigador establece que hay diferencia entre la actividad de proceso y las necesidades modales: las primeras son aquellas que sólo intentan llegar a la satisfacción, y las segundas son propositivas, es decir, plantean ejecuciones con dirección e, incluso, perfección. Sin embargo, entre todas las necesidades hay interacción, o sea, no actúan independientemente. Esta interacción es de suma importancia, ya que a la larga da significado y coherencia a una amplia gama de conductas del individuo.

Sería necesario escribir demasiadas páginas para ser más explícitos en la relación de los postulados e hipótesis de las teorías de Murray; no obstante, aquí nuestro interés es solamente fundamentar el instrumento a que nos hemos referido en un principio: la prueba de Jackson.

Jackson ha realizado múltiples estudios, tendientes a investigar la forma más completa y segura de medir la personalidad sobre los lineamientos de Murray, y son los siguientes: a) la forma A, constituida por 300 ítemes, divididos en 15 escalas de 20 ítemes cada una; b) la forma B, similar a la forma A, construida exactamente igual y aplicables una u otra intercambiadamente o combinadas (ambas formas tienen un tiempo de duración de 30 a 45 minutos para ser contestadas por el sujeto), y c) las formas AA y BB, que contienen 440 reactivos, distribuidos en las 15 escalas de las dos primeras formas, más siete escalas, con 20 reactivos cada una. Su tiempo de ejecución es de 40 a 70 minutos. Las hojas de respuesta han sido especialmente diseñadas para la eficiencia de la calificación manual.

Jackson ha llevado a cabo muchos estudios con estas pruebas. Su fin principal ha sido obtener la mayor confiabilidad y validez en sus trabajos. Sus estudios metodológicos han sido muy rigurosos. Ha establecido la validez convergente y discriminante; ha realizado investigaciones mediante el método de análisis factorial múltiple, a fin de establecer relaciones entre la personalidad y los intereses vocacionales. Asimismo, ha efectuado estudios de personalidad, sus investigaciones se han dirigido a eliminar, hasta donde sea posible, los prejuicios que operan en la calificación e interpretación de la misma y que conducen a distorsiones fatales. Asimismo, Jackson se ha interesado en detectar dos tipos de respuesta en este tipo de inventarios: las respuestas verdaderas y la aprobación de ítemes, llegando a encontrar que cuando una alta saturación de contenido influye en la elección de ítemes, aparecen escalas homogéneas, libres de respuestas estereotipadas.

Para establecer la veracidad de las respuestas, este investigador ha empleado métodos indirectos, entre ellos utilizar el juicio como único punto de vista sobre algunos atributos de un estímulo. Además, ha realizado investigaciones con sus pruebas en diversos tipos de sujetos con

alteraciones conductuales, tales como los alcohólicos y aquellos que presentan desviaciones de comportamiento.

Sus estudios, desarrollados sobre la construcción de escalas para poder llegar a formular escalas de personalidad, serán de gran utilidad en el futuro. El grado de validez y confiabilidad que Jackson ha obtenido en sus investigaciones con la prueba del P.R.F., ha tenido el carácter de descubrimiento inédito para pruebas de personalidad de escalas múltiples.

En 1971, en el INCCAPAC de México, con el fin de obtener un perfil sobre la personalidad del adolescente mexicano, las psicólogas Alicia Velázquez Medina y Silvia Furszyfer Mostkow llevaron a cabo una investigación con el P.R.F., prueba que forma parte de la batería utilizada en la investigación sobre el desarrollo de la personalidad del niño escolar mexicano (IDPEM), iniciada en 1964, con el objeto de llegar a saber si existen diferencias significativas en cada una de las escalas, a través de tres edades y dos sexos.

La muestra utilizada fue una ampliación de la muestra original del IDPEM descrita en páginas anteriores, cuyas características fueron las siguientes:

1. Es un estudio transversal que cubre tres años de edad del desarrollo de escolares mexicanos del Distrito Federal.
2. Los escolares pertenecían a tres sistemas educativos, que para diferenciarlos, se enunciaron de la siguiente manera: sistema I, que corresponde al sistema educativo privado; sistema II, correspondiente al sistema oficial de unidades habitacionales, y sistema III, perteneciente al sistema oficial propiamente dicho.
3. Los escolares utilizados fueron del nivel secundario: desde primer grado hasta tercero.
4. Al igual que en los demás estudios, los sujetos fueron probados cuando tenían las siguientes edades: los de primer grado, 12 años 8 meses; los de segundo, 13 años 8 meses, y los de tercero, 14 años 8 meses, con una desviación de tolerancia en la aplicación, para todos ellos, de 15 días antes y 15 días después de la fecha indicada.
5. Se tomaron 60 sujetos por cada grado y por cada sistema, divididos en ambos sexos (véase tabla 6).

Tabla 6. Muestra para el P.R.F.

	SISTEMA I		SISTEMA II		SISTEMA III	
	H	M	H	M	H	M
Primer grado: 12 años 8 meses	30	30	30	30	30	30
Segundo grado: 13 años 8 meses	30	30	30	30	30	30
Tercer grado: 14 años 8 meses	30	30	30	30	30	30
Total: 540 casos						

Las características de la muestra fueron las mismas que para toda la investigación de la IDPEM. Se utilizó la forma A del P.R.F., que consta de 15 escalas con 20 reactivos cada una, los cuales se presentan como afirmaciones o negaciones y se contestan en una hoja de respuestas, marcando falso o verdadero –según el caso – para cada ítem.

Las escalas llevan los siguientes títulos:

- Escala 1 : necesidad de logro.
- Escala 2 : necesidad de afiliación.
- Escala 3 : necesidad de agresión.
- Escala 4 : necesidad de autonomía.
- Escala 5 : necesidad de dominancia.
- Escala 6 : necesidad de resistencia.
- Escala 7 : necesidad de exhibicionismo.
- Escala 8 : necesidad de evasión.
- Escala 9 : necesidad de impulsividad.
- Escala 10: necesidad de interés por los demás.
- Escala 11: necesidad de orden.
- Escala 12: necesidad de humorismo.
- Escala 13: necesidad de reconocimiento social.
- Escala 14: necesidad de conocimiento intelectual.
- Escala 15: necesidad de ser veraz (evitar errores).

Para mayor información acerca de este estudio, será necesario consultar la tesis profesional de Velásquez y Furszyfer, hecha en 1972, denominada “Estudio preliminar del inventario de personalidad, de Douglas N. Jackson, en estudiantes de secundaria del Distrito Federal”. En estos párrafos sólo se indicarán las conclusiones acerca de este estudio.

La escala de veracidad merece especial atención, puesto que se elaboró para controlar las otras escalas. Sirve para detectar el error no intencional por parte del examinado. Los ítemes de esta escala expresan situaciones tan comunes a la población en general, que son obvias las respuestas, y si éstas no se producen, se detectan fácilmente. El análisis de los reactivos de esta escala puso de manifiesto este hecho. De tal suerte, mediante esta forma es posible evaluar la calidad de la respuesta, incrementándose la utilidad del instrumento.

Al respecto, se formularon varias hipótesis, que son las siguientes:

- a) Existen diferencias estadísticamente significativas entre cada uno de los factores de la personalidad medidos mediante la prueba de Jackson, para los diferentes grados escolares.
- b) Existen diferencias estadísticamente significativas entre los rasgos de personalidad para los dos sexos, es decir, entre los jóvenes y las jóvenes de México.

- c) Entre el grupo de hombres hay mayor número de diferencias estadísticamente significativas, que en el grupo de mujeres.
- d) Hay mayor número de diferencias estadísticamente significativas entre los hombres y las mujeres del tercer grado, que en los otros dos.
- e) La escala de interés por los demás presentó diferencias estadísticamente significativas en primero, segundo y tercer grados.
- f) Las escalas de necesidad de evasión y de afiliación presentan diferencias estadísticamente significativas en primero, segundo y tercer grados.
- g) La escala de necesidad de dominancia las presenta en los primeros y terceros grados, entre hombres y mujeres.
- h) En general, en las demás escalas hubo diferencias marcadamente significativas, principalmente en los sujetos de tercer grado, y en éstos entre hombres y mujeres.

Es importante hacer notar que aún quedan por realizar muchos estudios con este inventario, sobre todo aquellos destinados a convertirlo en un instrumento más válido y confiable para los sujetos mexicanos, así como desarrollar análisis estadísticos que pongan de manifiesto los alcances de este nuevo método de investigación de la personalidad.

CONSIDERACIONES FINALES ACERCA DE LA MEDICIÓN Y DE LA PSICOLOGÍA CLÍNICA

Lo expuesto anteriormente ha sido sólo una rápida revisión de algunas formas de medición y de ciertos instrumentos utilizados en la práctica de la psicología clínica. Además, se ha hecho hincapié primero en aquellos tests más conocidos, y después en los más válidos y confiables (por tanto, de pronóstico más útil y seguro), de los cuales se conocen ampliamente investigaciones y metodologías, verificadas para lograr los criterios y requisitos fundamentales que permitan detectar, ética y honestamente, en el campo profesional del psicólogo, las situaciones normales, anormales o patológicas, sin perjuicio de la salud mental y física de quienes necesiten la ayuda del psicólogo, y con la seguridad de que éste no cometerá equivocaciones tales que no sólo pongan en entredicho, como ha ocurrido, el prestigio de su profesión sino que también compliquen mucho más la problemática de quienes confiadamente acuden al psicólogo en busca de una solución para sus problemas.

Asimismo, cabe enfatizar, por último, que las técnicas utilizadas para la medición en el campo de la psicología clínica no son los únicos instrumentos de que puede disponer el psicólogo, así como tampoco éste es sólo un “aplicador de pruebas”; sus objetivos son desempeñar funciones más importantes que esa. Sin embargo, tales técnicas sirven para elaborar, una vez detectados los problemas de cada individuo, la instrumentación necesaria para operar los cambios conductuales, a fin de ayudar a solucionar los problemas humanos.

En esta era de las máquinas fantásticas que “todo lo hacen”, se ha llegado a afirmar absurdamente que el hombre puede ser sustituido por cualquiera de ellas, para desarrollar aquellas actividades en las cuales aún hace 20 años se requería absolutamente su presencia. Esto puede ser cierto sólo relativamente, puesto que hay actividades para las cuales las máquinas *nunca* podrán ocupar el lugar del hombre; sin embargo, éste puede usarlas como auxiliares para su mejor rendimiento.

De tal suerte, se puede decir que el hombre *crea*, y las máquinas no. El hombre siempre estará capacitado para diseñar o inventar nuevas formas de tecnología, a fin de facilitar la vida humana, lo cual le está vedado a una computadora, por ejemplo, porque carece de esa habilidad inherente al ser humano, que se llama *creatividad*.

Se hace hincapié en esto porque ha llegado a afirmarse que en un momento dado, el psicólogo será sustituido por las computadoras. Esto es absolutamente imposible, sobre todo en el campo de la psicología, donde se requiere, la mayor parte del tiempo, la confrontación del psicólogo con la problemática humana. En esta relación, el psicólogo al observar, medir y evaluar el comportamiento de dos individuos en la situación específica que se trate, tiene que crear o adaptar técnicas psicológicas –ya conocidas o inventadas por él sobre su preparación profesional – a las circunstancias particulares, porque tampoco puede generalizar ni aplicar el mismo tratamiento a todos aquellos sujetos que lo requieran, aún cuando el o los problemas sean aparentemente iguales. De igual forma, no debe olvidar las diferencias individuales; de tal suerte, si un médico debe conocer diversos medicamentos para tratar un padecimiento específico, teniendo en cuenta el organismo humano de que se trate, asimismo el psicólogo tiene igual obligación en cuanto al empleo de las técnicas de tratamiento psicológico.

Cada sujeto necesita diferentes formas de solución a sus alteraciones conductuales o a sus problemas específicos, aún cuando éstas se deriven de las técnicas básicas; y esto sólo puede hacerlo el hombre-psicólogo. Incluso dentro de la propia psicología, todas las técnicas que tiendan cada vez más a considerar al individuo como una computadora ambulante, tenderán también a desaparecer y a caer por sí solas si no intentan un enfoque más humano, pues, de no hacerlo así, no resolverán problemas a largo plazo.

“Psicometría Aplicada”, Morales, María Luisa, Ed. Trillas, 2ª. Edición, México, D.F. 1990.

