



Mi Universidad

LIBRO

Psicometría

Licenciatura en Psicología

Tercer Cuatrimestre

Mayo – Agosto

Marco Estratégico de Referencia

Antecedentes históricos

Nuestra Universidad tiene sus antecedentes de formación en el año de 1979 con el inicio de actividades de la normal de educadoras “Edgar Robledo Santiago”, que en su momento marcó un nuevo rumbo para la educación de Comitán y del estado de Chiapas. Nuestra escuela fue fundada por el Profesor Manuel Albores Salazar con la idea de traer educación a Comitán, ya que esto representaba una forma de apoyar a muchas familias de la región para que siguieran estudiando.

En el año 1984 inicia actividades el CBTiS Moctezuma Ilhuicamina, que fue el primer bachillerato tecnológico particular del estado de Chiapas, manteniendo con esto la visión en grande de traer educación a nuestro municipio, esta institución fue creada para que la gente que trabajaba por la mañana tuviera la opción de estudiar por las tardes.

La Maestra Martha Ruth Alcázar Mellanes es la madre de los tres integrantes de la familia Albores Alcázar que se fueron integrando poco a poco a la escuela formada por su padre, el Profesor Manuel Albores Salazar; Víctor Manuel Albores Alcázar en julio de 1996 como chofer de transporte escolar, Karla Fabiola Albores Alcázar se integró en la docencia en 1998, Martha Patricia Albores Alcázar en el departamento de cobranza en 1999.

En el año 2002, Víctor Manuel Albores Alcázar formó el Grupo Educativo Albores Alcázar S.C. para darle un nuevo rumbo y sentido empresarial al negocio familiar y en el año 2004 funda la Universidad Del Sureste.

La formación de nuestra Universidad se da principalmente porque en Comitán y en toda la región no existía una verdadera oferta Educativa, por lo que se veía urgente la creación de una institución de Educación superior, pero que estuviera a la altura de las exigencias de los jóvenes

que tenían intención de seguir estudiando o de los profesionistas para seguir preparándose a través de estudios de posgrado.

Nuestra Universidad inició sus actividades el 18 de agosto del 2004 en las instalaciones de la 4ª avenida oriente sur no. 24, con la licenciatura en Puericultura, contando con dos grupos de cuarenta alumnos cada uno. En el año 2005 nos trasladamos a nuestras propias instalaciones en la carretera Comitán – Tzimol km. 57 donde actualmente se encuentra el campus Comitán y el corporativo UDS, este último, es el encargado de estandarizar y controlar todos los procesos operativos y educativos de los diferentes campus, así como de crear los diferentes planes estratégicos de expansión de la marca.

Misión

Satisfacer la necesidad de Educación que promueva el espíritu emprendedor, aplicando altos estándares de calidad académica, que propicien el desarrollo de nuestros alumnos, Profesores, colaboradores y la sociedad, a través de la incorporación de tecnologías en el proceso de enseñanza-aprendizaje.

Visión

Ser la mejor oferta académica en cada región de influencia, y a través de nuestra plataforma virtual tener una cobertura global, con un crecimiento sostenible y las ofertas académicas innovadoras con pertinencia para la sociedad.

Valores

- Disciplina
- Honestidad
- Equidad
- Libertad

Escudo



El escudo del Grupo Educativo Albores Alcázar S.C. está constituido por tres líneas curvas que nacen de izquierda a derecha formando los escalones al éxito. En la parte superior está situado un cuadro motivo de la abstracción de la forma de un libro abierto.

Eslogan

“Mi Universidad”

ALBORES



Es nuestra mascota, un Jaguar. Su piel es negra y se distingue por ser líder, trabaja en equipo y obtiene lo que desea. El ímpetu, extremo valor y fortaleza son los rasgos que distinguen.

Psicometría.

Objetivo de la materia:

La medición es una de las principales herramientas en el quehacer del psicólogo. Por tal motivo es esencial que dentro de su formación cuenten con el conocimiento sobre los aspectos esenciales para la aplicación e interpretación de instrumentos de medición como de los test ya validados y estandarizados que miran diferentes aspectos y de mayor uso en el campo de la psicología.

Criterios de evaluación:

No	Concepto	Porcentaje
1	Trabajos Escritos	10%
2	Actividades Áulicas	20%
3	Trabajos en plataforma Educativa	20%
4	Examen	50%
Total de Criterios de evaluación		100%

INDICE

Unidad 1

Introducción a la medición en psicología

- 1.1 Definición de psicometría
- 1.2 Consideraciones éticas del uso de pruebas psicométricas
- 1.3 Definición de test/ pruebas psicológicas
- 1.4 Clasificación de los test psicológicos
- 1.5 Diferencia entre pruebas proyectivas y psicométricas

Unidad 2

Características determinantes para la selección de instrumentos psicométricos

- 2.1 Confiabilidad
 - 2.1.1 Definición de confiabilidad
 - 2.1.2 Métodos de la evaluación de la confiabilidad: temporal, equivalentes, consistencia interna
- 2.2 Estandarización
 - 2.2.1 Definición de estandarización
- 2.3 Validez
 - 2.3.1 Definición de validez

2.3.2 Validez de constructo

2.3.3 Validez de criterio

2.3.4 Validez de contenido

Unidad 3

Normas de calificación

3.1 Normas de calificación y tipificación

3.2 Desviación estándar

3.3 Comparación interindividual (confronta de resultados de otros sujetos)

3.4 Comparación intraindividual (dos medidas del mismo sujeto)

Unidad 4

Procesos de aplicación de pruebas

4.1 Procedimientos de aplicación de pruebas: requisitos de aplicación, criterio de inclusión y exclusión

4.2 Calificación de la prueba

4.3 Estructura y calificación del informe

4.4 El psicodiagnóstico y diagnóstico

4.5 Manejo y aplicación de la prueba proyectiva: dibujo de la figura humana de Karen Machover

4.5.1 Calificación de la prueba proyectiva: dibujo de la figura humana de Karen Machover

4.5.2 Interpretación y redacción de informe de la prueba proyectiva: dibujo de la figura humana de Karen Machover

4.6 Manejo y aplicación de la prueba BETA III

4.6.1 Calificación de la prueba de inteligencia BETA III

4.6.2 Interpretación y redacción de informe de la prueba de inteligencia BETA III

4.7 Manejo y aplicación de la prueba neuropsicológica de Bender

4.7.1 Calificación de la prueba neuropsicológica de Bender

4.7.2 Interpretación y redacción de informe de la prueba neuropsicológica de Bender

Unidad I

Introducción a la medición en psicología.

I.1. Definición de psicometría.

La Psicometría es una disciplina científica encuadrada dentro del marco de la Metodología de las Ciencias del Comportamiento y directamente relacionada con el campo de la medición psicológica.

Hace uso intensivo sobre todo de cálculos y análisis estadísticos para extraer información útil a partir de la administración repetida de un mismo test a un grupo amplio de personas.

La psicometría es la rama de la psicología que se encarga de estudiar las teorías, métodos y técnicas que permiten o mejor dicho sustentan las medidas indirectas de los fenómenos psicológicos. Desde sus inicios como ciencia hasta la actualidad, los avances en este campo, realizados por psicólogos, educadores, estadísticos, entre otros, han sido impresionantes, convirtiendo lo no observable, al menos de forma directa, en fenómenos observables, es decir, lo subjetivo en algo objetivo, tarea nada fácil, pero que ahora es posible, obteniendo mediciones con bastante precisión. Dentro de la obra de Livia y Ortiz (2014) se permite conocer el gran avance conseguido en esta área.

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 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”.

Antecedentes históricos de la psicometría.

El hecho de que las personas difieran en su comportamiento y que esas diferencias puedan medirse se ha reconocido desde los albores de la civilización. Platón y Aristóteles escribieron sobre las diferencias individuales hace más de 2000 años, y los chinos, desde la dinastía Chang (1115 a.C.), ya tenían un programa de pruebas para el ingreso de los funcionarios públicos que evaluaba destrezas importantes para la época, tales como arquería, equitación, música, escritura y matemática (Cohen y Swerdlik, 2000).

No obstante, en su acepción actual, el empleo de los tests psicológicos se inició en Europa a fines del siglo XIX. Durante la Edad Media la preocupación por la individualidad era prácticamente inexistente, permitiéndose poca libertad para la expresión y el desarrollo de la personalidad (Aiken, 2003). Es en el Renacimiento y la Ilustración cuando resurge el interés por el aprendizaje y la creatividad.

Sin embargo, recién a finales del siglo XIX se inicia el estudio científico de las diferencias individuales en lo que respecta a habilidades y rasgos de personalidad. Los tests se desarrollaron dentro del contexto de la formulación de la teoría de la evolución de las especies y las fases tardías de la Revolución Industrial, en el marco de una creciente preocupación por el aumento de la población, la mano de obra desocupada y la paulatina democratización de las escuelas.

En este contexto, surge un llamativo interés por las diferencias individuales, especialmente las de carácter hereditario, así como también por la adaptabilidad diferencial de los seres humanos a las exigencias de un entorno cambiante. Esta filosofía, denominada “darwinismo social”, centraba su interés en las diferencias de naturaleza hereditaria y la adaptabilidad de los seres humanos a las exigencias de la sociedad industrial (Sternberg, 1987).

Las diferencias observadas por el astrónomo Friedrich Bessel a comienzos del siglo XIX en los registros del paso de las estrellas a través de una línea del campo visual del telescopio, realizados por distintas personas, se convirtió en la primera evidencia de que algunas capacidades humanas podían cuantificarse. El matemático belga Adolphe Quetelet fue el primero en plantear que la teoría estadística de la probabilidad podía aplicarse a la medición del comportamiento humano (Herrera Rojas, 1998).

A estas contribuciones se sumaron las de los primeros estudiosos de la psicofísica, tales como Gustave Fechner y Ernst Weber, y los fundadores de la psicología experimental, destacándose la figura de Wilhelm Wundt con su fuerte interés por medir la magnitud de propiedades psicológicas elementales a fin de formular leyes científicas. Estos autores también pusieron de manifiesto la necesidad de controlar las condiciones de prueba y tipificar los procedimientos. No obstante, el interés de los investigadores pioneros de la medición psicológica se orientó principalmente a la formulación de leyes generales que permitiesen predecir el comportamiento, y no tanto hacia la explicación de las diferencias individuales.

Es de particular relevancia la figura de Sir Francis Galton (1822-1911), primo del célebre Charles Darwin, quien a partir de sus estudios sobre la heredabilidad de la inteligencia, fue el principal responsable del inicio del movimiento psicométrico y del interés por la medición de las diferencias individuales. Este investigador inglés, interesado por el estudio de la herencia, creó un laboratorio antropométrico en Kensington, Inglaterra, donde cualquier persona podía evaluar su estatura, peso corporal, fuerza muscular, agudeza visual y otra serie de características sensoriales y motoras. Galton construyó varios tests de discriminación sensorial con la convicción de que éstos le permitían medir la inteligencia, y fue el primer investigador en adaptar algunas técnicas estadísticas para el análisis de los resultados de los tests, constituyéndose en el precursor del uso de procedimientos de análisis cuantitativos en investigación con humanos (Herrera Rojas, 1998). Con sus estudios sobre gemelos fue también uno de los fundadores de la genética del comportamiento, uno de los campos más influyentes en la psicología contemporánea (Loelhin, 1992).

En sintonía con las ideas de Galton, James Catell construyó diferentes tests de tiempos de reacción y otras funciones mentales simples. A este autor se le debe, además, la rápida difusión de los tests en los Estados Unidos y los primeros intentos por validarlos en relación con criterios externos, es decir, comprobar si efectivamente predecían comportamientos reales diferentes de la situación de evaluación, tales como el éxito académico de los estudiantes universitarios. Sin embargo, su aporte más significativo es el de haber introducido en la literatura psicológica el término *test mental* (Muñiz, 2001).

En el año 1895, el psicólogo francés Alfred Binet publicó un artículo en el cual criticaba los tests existentes en ese momento, considerando que medían funciones muy elementales y que poseían escasa capacidad predictiva en relación con criterios externos relevantes, tales como el rendimiento académico. Binet propuso crear tests de medición de funciones mentales más complejas, tales como juicio, memoria y razonamiento. Por su parte, Wissler (1901) demostró a comienzos del siglo XX que los tests sensoriales o de reacciones mentales simples no predecían en forma adecuada el rendimiento académico de los estudiantes. Todo esto propició la creación de medidas psicológicas más semejantes a las actividades de la vida cotidiana.

En este contexto se creó la primera escala de inteligencia, que integró las experiencias anteriores e introdujo ítems relacionados con juicio, comprensión y razonamiento. Binet y

Simon, a pedido del gobierno francés, utilizaron por primera vez en 1905 una escala para identificar, entre los niños que ingresaban a primer grado, aquellos que padecían debilidad mental. Esta escala consistía en 30 problemas de dificultad creciente (comprensión verbal y capacidad de razonar con materiales no verbales) y representa el desempeño típico de los niños a una edad determinada.

En 1908 estos autores desarrollaron la noción de edad mental y también una escala más refinada que se constituyó en el prototipo de los tests individuales de inteligencia. En la revisión de la escala Binet-Simon, realizada por Terman y conocida como Stanford-Binet, aparece la noción de Cociente

Intelectual (CI). La propuesta de Terman del CI como unidad de medida de la inteligencia, con todas sus limitaciones (entre las que se destaca el hecho de que los cocientes intelectuales no serían comparables entre edades, debido a diferencias en la variabilidad de la ejecución del test), tiene una gran importancia en la psicometría, al punto tal que el CI se convirtió casi en un mito.

En esta época también fueron muy importantes los descubrimientos de un grupo de investigadores que perfeccionaron diferentes índices y modelos de análisis estadísticos, particularmente en la medición de la inteligencia. Se destacaron los trabajos de Karl Pearson (1857-1936), discípulo de Galton, quien desarrolló el coeficiente de correlación que lleva su nombre (“producto momento de Pearson”), sentando las bases para el análisis estadístico que se realiza actualmente en psicología.

Por otra parte, Charles Spearman (1927) inició una serie de investigaciones sobre las funciones cognitivas que lo llevaron al desarrollo del análisis factorial. Apoyándose en la observación de correlaciones entre tests, Spearman plantea su famosa teoría de dos factores. Según esta teoría, las puntuaciones de los tests pueden explicarse a través de dos factores: uno general, conocido como el factor g , que es común a todas las variables medidas, y uno específico, s , que sería exclusivo de cada una de esas variables.

Pocos acontecimientos en la historia de los tests mentales han tenido una importancia tan grande como la formulación de la teoría de los dos factores de la inteligencia. Sobre ese fundamento se han construido numerosos tests, no sólo de inteligencia sino también de personalidad, intereses y otros constructos psicológicos. Spearman concibió también la teoría

de la confiabilidad de los tests y, junto a Thorndike, el modelo estadístico de puntuaciones conocido luego como Teoría Clásica de los Tests (Martínez Arias, 1995).

La Primera Guerra Mundial generó grandes problemas para la selección y adiestramiento de millones de combatientes. El programa de selección masiva en el que se involucraron los psicólogos más capaces de la época significó una prueba de la madurez de la teoría y la técnica psicométrica. Se elaboraron losDe este modo, en el nivel (edad mental) de 3 años se agruparon todos los ítems que resolvía el 80% de los niños normales de esa edad y así sucesivamente hasta los 13 años (Binet y Simon, 1916; Anastasi y Urbina, 1998).

primeros tests colectivos de inteligencia para la clasificación de grandes masas de reclutas, los célebres tests Alfa y Beta del ejército norteamericano. El test Army Alfa, elaborado por psicólogos militares dirigidos por Yerkes (1921), estaba constituido por ocho subtests que medían aspectos tales como razonamiento práctico, analogías y razonamiento matemático. El Army Beta era una versión no verbal del anterior, utilizada para la evaluación de combatientes con capacidades lingüísticas limitadas o que no eran angloparlantes.

Woodworth (en Anastasi y Urbina, 1998) desarrolló su Personal Data Sheet, un autoinforme con preguntas sobre sintomatología mental, tales como ¿usted toma whisky todos los días? (*en mi caso debo reconocer que sí*). La finalidad de este instrumento era detectar soldados con trastornos psicológicos y que no fueran aptos para el servicio militar durante la Primera Guerra Mundial. Este inventario se convirtió en modelo para los inventarios de personalidad posteriores, más sofisticados, que revisaremos más adelante. La amplia difusión de los tests colectivos durante la primera conflagración mundial fue observada con interés por los educadores, dada la practicidad de estos instrumentos.

Como consecuencia de todas estas innovaciones se produjo una actividad creciente de construcción de pruebas y se desarrollaron las nociones iniciales de estandarización y validación de los tests mentales.

Es importante destacar en esta época la obra de Rorschach (1921), el psiquiatra suizo que publicó una técnica de psicodiagnóstico basada en una serie de láminas con manchas de tinta, recomendando su uso como herramienta de investigación. El test de Rorschach configuró una nueva tendencia en la evaluación psicológica vinculada con modelos teóricos psicodinámicos.

El año 1935 ha sido calificado como “bisagra” entre el período

“histórico” y “moderno” dentro de la psicometría (Sternberg, 1987). Ese año se fundó la Sociedad Psicométrica por un grupo de investigadores agrupados en torno a la figura de L. Thurstone, investigador de la Universidad de Chicago. También en este año surge la primera publicación especializada en la medición psicológica, *Psychometrika*, que continúa vigente en la actualidad.

La mayoría de los tests publicados hasta ese momento se basaban en la concepción de la inteligencia como rasgo unitario.

Los investigadores nucleados en la Sociedad Psicométrica desarrollaron una innovación fundamental: el análisis factorial moderno, un método que demostraba con claridad que la inteligencia es algo más que una capacidad unitaria. El psicólogo estadounidense Thurstone realizó una serie de aportes a la lógica y los fundamentos matemáticos del análisis factorial, logros que facilitaron la medición de aptitudes más específicas, que contribuyen al desempeño cognitivo más allá de la influencia de la inteligencia general o *g*. El test de Aptitudes Mentales Primarias de Thurstone (1935) fue un modelo para las baterías de tests multifactoriales posteriores, inaugurando una nueva manera de concebir y medir la inteligencia.

Como hemos dicho, con el empleo del análisis factorial se construyeron numerosos tests, no sólo de inteligencia, sino también de personalidad, intereses y otros atributos psicológicos.

Teorías contemporáneas tales como la de la inteligencia fluida (*Gf*) y cristalizada (*Gc*) de Cattell (1967), la teoría de los cinco factores de la personalidad (Norman, 1963; Costa y Mc Crae, 1999; Goldberg, 1999) y otros modelos semejantes (Caroll, 1993) constituyen un refinamiento de los postulados precursores de Spearman y Thurstone.

La Segunda Guerra Mundial, con sus necesidades de incorporación de millones de reclutas, también estimuló la construcción de tests de aptitudes específicas, que fueron muy útiles para seleccionar pilotos, bombarderos, operadores de radio y otras funciones militares especializadas. Por esa época, Guilford (1967) construyó para la fuerza aérea una batería de tests que medían diferentes factores de la estructura de la inteligencia humana. El Test de Aptitudes Diferenciales (Bennet, Seashore y Wesman, 2000), entre otros similares, son herederos de esos descubrimientos.

La década de 1950 es considerada como una fase “madura” de la teoría de los tests, puesto que aparecieron textos que con el tiempo serían clásicos y dejarían establecidos los

fundamentos teóricos de la psicometría. Surge en ese momento histórico una corriente de revisión y análisis de la fundamentación científica de las pruebas. Los trabajos realizados en este período versan en su gran mayoría sobre teoría de la medición, los principios y fundamentos de la medición en psicología, los problemas de validez y confiabilidad y, en síntesis, la construcción de una teoría psicométrica. Así, pueden mencionarse *Theory of Mental Tests* (Gulliksen, 1950) y las normas técnicas iniciales de la APA, entre otras obras valiosas (Hogan, 2004).

En la década de 1960 se comenzó a criticar esta concepción clásica de la teoría de los tests, al tiempo que aparecían teorías alternativas. Hay dos modelos originados en esa época que prevalecen en la literatura psicométrica actual: el de maestría de dominio y el de rasgo latente. Dentro del primero se ubican los denominados tests con referencia a criterio, término introducido por Glaser (1963), que miden un dominio de conocimiento claramente delimitado. Estos tests están íntimamente ligados al campo educativo. Por otra parte, la teoría de rasgo latente (Rasch, 1963) derivó en la Teoría de Respuesta al Ítem (Lord, 1980), uno de los paradigmas relevantes de la psicometría contemporánea. Ambos enfoques serán revisados más adelante.

En los últimos años del siglo XX se produjo un acercamiento entre la psicometría y la psicología cognitiva, y se elaboraron modelos psicométricos denominados “modelos componenciales” que incorporan los diferentes componentes de los procesos cognitivos en la resolución de un problema (Van der Linden y Hambleton, 1997; Prieto y Delgado, 1999). Estos modelos también se conocen como “evaluación inteligente”; en ellos se presentan tareas que son comunes en la vida real. Un modelo componencial requiere: a) un análisis de las operaciones mentales (componentes cognitivos) que intervienen en la resolución de los ítems y b) un modelo matemático que estime la probabilidad de responder correctamente el ítem teniendo en cuenta sus propiedades psicométricas y el nivel de conocimiento del sujeto.

El uso de las computadoras en psicometría tuvo un notable incremento desde la década de 1980, aplicándose en casi todas las instancias de la evaluación psicológica. Debido a su consistencia, la computadora lleva al extremo la estandarización y objetividad de un test. Algunas de las aplicaciones más interesantes de la informática en los tests psicológicos son:

a) *Bancos de ítems*: La forma de presentación más habitual de un test es un cuadernillo impreso con los ítems a resolver. Sin embargo, un test bien ajustado para determinados propósitos puede ser rápidamente creado a partir de un banco de ítems. Una escuela puede solicitar a un editor algunos tests adecuados a los contenidos de su currícula. En la actualidad es posible elaborar un banco de ítems y, a partir del mismo, construir tests con una computadora. Otra posibilidad es generar un número ilimitado de formas equivalentes de un test, seleccionándolas con el mismo criterio de un banco de ítems. Un procedimiento denominado GAI (generación automática de ítem) permite generar ítems mediante determinados algoritmos, que requieren programas específicos como el Rasch Item Calibration Program (RASCAL, 1989) u otros semejantes. La Teoría de Respuesta al Ítem, revisada al final de este texto, constituye el marco conceptual y metodológico para el desarrollo de bancos de ítems de tests.

b) *Administración y puntuación asistida por computadora*: Las computadoras son adecuadas para administrar y puntuar los tests. Al aplicar una prueba en formato computarizado se obtienen mediciones precisas e instantáneas, no hay errores en la corrección y se consiguen informes legibles con posibilidad de transmisión y multicopias impresas.

Además la interacción con las computadoras fascina a las nuevas generaciones y se espera que esto vaya en aumento creciente. Un inconveniente es que la informatización puede ocasionar una pérdida de la riqueza de las observaciones no formales que realizan los administradores expertos durante la aplicación de un test individual, algo que puede atenuarse si el administrador acompaña el proceso de respuesta del individuo al test computarizado (Cronbach, 1998).

c) *Software de simulación*: Las nuevas tecnologías incrementan notablemente la variedad de los estímulos incluidos en los tests. Los simuladores de vuelo, por ejemplo, representan de modo realista el instrumental que deben manipular los pilotos y proporcionan continua retroalimentación de los resultados de sus operaciones. Aunque fueron diseñados para entrenamiento, estos dispositivos pueden también ser empleados para la evaluación del progreso en el aprendizaje en cualquier dominio.

La mayoría de los tests empleados en las diversas áreas de la psicología disponen de versiones computarizadas. El directorio de *software* psicológico de la APA (1999) describe regularmente centenares de programas para administrar y/o interpretar tests por computadoras. Estas

tecnologías permiten economizar el proceso de puntuación y elaboración de perfiles y, al mismo tiempo, mejoran la precisión y objetividad de los tests convencionales (de lápiz y papel) al eliminar los errores que se cometen durante la puntuación manual de los mismos.

El alcance de la informática en psicometría no se limita al empleo de tests asistidos por computadora. En las últimas décadas se ha diseñado una amplia variedad de programas estadísticos que incluyen rutinas y menús adecuados para resolver problemas de investigación en este dominio (estudios correlacionales, análisis factorial, entre otros). En el apéndice de este libro presentamos ejemplos de algunos análisis psicométricos realizados con *software* estadístico moderno.

La búsqueda de información relacionada con los tests también se ve sumamente facilitada por los recursos disponibles en Internet, donde se encuentran bases documentales de gran utilidad como la que ofrece el sitio web de la American Psychological Association (www.apa.org), así como revistas y portales científicos *on line* (www.sciencedirect.com, por ejemplo), y editoriales abocadas exclusivamente a la publicación de tests, tales como TEA en España (www.teaediciones.com).

En la Facultad de Psicología de la Universidad Nacional de Córdoba existe una revista electrónica especializada, *Evaluar*, que periódicamente publica trabajos teóricos y empíricos relacionados con la medición psicológica y educativa (www.revistaevaluar.com.ar).

En síntesis, la psicometría moderna evidencia tres características fundamentales: a) la importancia de la teoría de respuesta al ítem en la construcción de tests, coexistiendo con la teoría clásica de los tests (y en algunos casos reemplazándola); b) la presencia creciente de los tests basados en computadora en lugar de los tests de lápiz y papel, y c) el diseño de ítems más atractivos y realistas, que incorporan recursos audiovisuales y de simulación computarizada, hecho que promete una nueva generación de tests de mayor validez (Kaplan y Saccuzzo, 2006; Moreno, Martínez y Muñiz, 2004).

Las escalas más utilizadas para la medición de la inteligencia en nuestro medio son las elaboradas por David Wechsler en 1939, con varias actualizaciones posteriores; las últimas referentes al WISC-IV (Wechsler, 2005), para niños y adolescentes, y el WAIS-III (Wechsler, 1999), para adultos. Todas las escalas de Wechsler comprenden subtests verbales y de ejecución. Los ítems de los subtests verbales plantean problemas del tipo *¿Qué significa*

arrogante?, o Menciona un planeta de nuestro sistema solar que no sea la Tierra; los subtests no verbales consisten, por ejemplo, en ensamblar objetos a la manera de un rompecabezas.

Análisis psicométricos contemporáneos de las escalas Wechsler identificaron cuatro factores de inteligencia subyacentes (organización perceptual, memoria de trabajo, comprensión verbal y velocidad de procesamiento). En la última versión del WISCIV las puntuaciones se interpretan en función de esos cuatro factores y no en la forma tradicional de inteligencia verbal y de ejecución. En las versiones actuales de las escalas Wechsler los ítems están ordenados según los parámetros de dificultad y discriminación de la teoría de respuesta al ítem (Hogan, 2004).

Aplicación de la psicometría en diversos campos.

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 educativa, 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 las 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 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. 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 abogados 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.

1.2. Consideraciones éticas del uso de pruebas psicométricas.

Protecciones. Las normas éticas de los psicólogos (APA, 1992), requieren que estos profesionales empleen sólo técnicas o procedimientos que estén dentro de su competencia.

Además en general los editores de materiales de pruebas restringen su venta a individuos o instituciones que pueden demostrar su capacidad para aplicar, calificar e interpretar pruebas. Entonces, la venta de pruebas no está abierta, sino que depende de la capacitación del usuario. La mayoría de las personas suponen que tienen que revelar tanto tan poco como deseen de sus actitudes, sentimientos, temores o aspiraciones. Por supuesto, con los procedimientos sutiles o indirectos de evaluación, una persona evaluada no siempre puede juzgar con completa certidumbre si una respuesta dada es deseable. Pero cualquiera que sea la naturaleza de la prueba el individuo tiene derecho a una explicación plena de sus propósitos y del empleo que se dará a los resultados.

Sólo deben aplicarse a la persona las pruebas pertinentes para los propósitos de evaluación.

Confidencialidad.

Cuando se evalúa a un individuo, se deben hacer todos los esfuerzos por explicar el los propósitos y utilización de las pruebas y proporciona información acerca de las personas o instituciones que tendrán acceso a los resultados. Si un individuo da su consentimiento informado, el procedimiento de evaluación puede continuar.

Predisposición de la prueba.

El clínico debe puede superar la predisposición de la prueba mediante la utilización de ecuaciones de predicción diferentes (y más apropiadas) para diferentes grupos. Es decir, la predisposición entra en juego cuando el psicólogo clínico hace predicciones son base a asociaciones empíricas que so características de otro grupo de interés (por ejemplo, varones), pero no los miembros del grupo de interés (mujeres). La meta consiste en investigar la posibilidad de validez confidencial y, de encontrarse, se debe emplear la ecuación apropiada de predicción para ese grupo

1.3. Definición de test/pruebas psicológicas.

Los tests psicológicos se construyen, en general, para medir constructos que no pueden observarse directamente. Nunnally y Bernstein (1995) afirmaron que nunca se miden las personas sino algunos de sus atributos, es decir, características particulares de los individuos.

En psicología nadie se propone “medir” un niño, sino su inteligencia, estabilidad emocional o autoestima, por ejemplo. Por otro lado, las operaciones de medición en psicología son casi siempre indirectas, vale decir, suponen la determinación de los indicadores del fenómeno a medir.

Como argumentó Martínez Arias (1995), el estatus actual de la psicología genera una serie de dificultades para el desarrollo de instrumentos científicos de medición, a saber:

- a) Un mismo constructo psicológico puede ser definido de manera diferente, por lo cual distintos procedimientos de medida pueden conducir a inferencias disímiles en relación a aquél.
- b) Es difícil determinar las características de una muestra de elementos (ítems) de un test para que sea representativa, en cuanto a extensión y variedad de contenidos, del dominio o constructo que se quiere medir.
- c) Como consecuencia de lo expresado en los puntos anteriores siempre existen errores en las medidas.
- d) Las escalas de medición usadas en psicología carecen, casi siempre, de cero absoluto y de unidades de medidas constantes.

Aun con estas deficiencias, el nivel de precisión alcanzado por la medición en psicología permite exhibir algunas ventajas respecto a la observación natural o no formal del comportamiento, entre ellas:

- Una de las principales es la objetividad, que implica que una afirmación fáctica es posible de verificar por otros científicos en forma independiente.
- La posibilidad de medición de las variables facilita el desarrollo de investigaciones. Según Nunnally (1991), los avances en las ciencias en general, y en la psicología en particular, se relacionan con los adelantos en los métodos de medición, aunque lo opuesto también es una realidad constatable en la historia de la ciencia.
- Los índices numéricos utilizados por los tests permiten comunicar los resultados de una evaluación con mayor precisión.

De este modo, los tests proporcionan discriminaciones más sutiles que la clasificación intuitiva que un maestro podría hacer de sus estudiantes, incluyéndolos en categorías poco discriminativas como “brillante”, “promedio” o “debajo del promedio”, por ejemplo.

- El desarrollo de tests es un proceso complejo, pero el resultado final es un procedimiento estandarizado más sencillo y breve que la observación. Pensemos, en relación con esta última aseveración, en el tiempo requerido para administrar y puntuar un test en comparación con el tiempo que demandaría la observación del desempeño o comportamiento de una persona en su ambiente natural (escuela o trabajo, por ejemplo).

La delimitación del concepto de *tests psicológicos* no es sencilla, y a lo largo de la historia de la psicología ha suscitado innumerables polémicas. El término inglés *test* (prueba, examen) proviene del vocablo latino *testa-testis*, que denominaba una balanza utilizada en la antigüedad para pesar vasijas de oro (Cortada de Kohan, 1999).

De acuerdo con Anastasi y Urbina (1998), un test es un instrumento de medición del comportamiento de un individuo, a partir del cual pueden inferirse otros comportamientos relevantes. En 1999 la American Psychological Association (en adelante APA) definió a los tests como *un procedimiento por medio del cual una muestra de comportamiento de un dominio especificado es obtenida y posteriormente puntuada, empleando un proceso estandarizado*. Esta definición comprende no sólo a los tests de ejecución máxima, donde las respuestas son evaluadas por su corrección y calidad sino también a los de comportamiento típico (inventarios de personalidad, por ejemplo) siempre que respeten el postulado anterior.

El concepto de “evaluación” es más comprensivo que el de test y se refiere al proceso que permite integrar la información obtenida por medio de tests con la proveniente de otras fuentes, tales como la información relacionada con la historia clínica, familiar, ocupacional o educacional de una persona.

1.4. Clasificación de los test psicológicos.

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 aquella 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. 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.

Diferencia entre las pruebas proyectivas y psicométricas. .

Pruebas psicométricas:

Las pruebas o test psicométricos son una medida objetiva y tipificada de una muestra de conducta. Fundamentalmente nos permiten hacer descripciones y comparaciones de unas personas con otra y también de una misma persona en diferentes momentos de su vida;

Este tipo de exámenes psicométricos deben saber utilizarse y sus mediciones son bastante correctas si se utilizan bien. Dependiendo de la naturaleza de la cuestión a medir se elegirá un tipo de test u otro y datos obtenidos con este tipo de test pueden ser más fiables en unas personas que en otras, por tanto, hay que usarlos adecuadamente.

Las pruebas psicométricas son las encargadas de medir cualidades psíquicas del individuo, las hay de varios tipos: de medición de inteligencia (que tanto cotejan edad mental y edad cronológica), rasgos de personalidad (como tiendes a reaccionar), factores de personalidad (cotidianamente como te comportas).

Son una herramienta tanto para conocer su vida, estado emocional, intelectual e inclusive si existe alguna anomalía en su funcionamiento cerebral o simplemente para ubicarnos en su forma de pensar.

Pruebas proyectivas:

Las técnicas proyectivas son “unos instrumentos considerados como especialmente sensibles para revelar aspectos inconscientes de la conducta, ya que permiten provocar una amplia

variedad de respuestas subjetivas, son altamente multidimensionales y evocan respuestas y datos del sujeto, inusualmente ricos con un mínimo conocimiento del objetivo del test, por parte de éste” (Lindzey 1961) Las técnicas proyectivas constituyen junto con las técnicas psicométricas el grueso de la evaluación tradicional. Se suele separar las técnicas proyectivas de las técnicas psicométricas, considerando que las primeras intentan esencialmente comprender, en tanto que las segundas se ocupan de medir y evaluar.

Unidad 2

Características determinantes para la selección de instrumentos psicométricos.

Confiabilidad.

2.1.1 Definición de confiabilidad

En la actualidad coexisten dos teorías generales de los tests, la teoría clásica de los tests y la de respuesta al ítem.

La hipótesis fundamental de la teoría clásica de los tests (TCT) es que la puntuación observada de una persona en un test es una función de dos componentes: su puntaje verdadero (que es inobservable) y el error de medición implícito en toda medición.

El postulado esencial de la TCT se expresa como:

$$O_i = V_i + E_i,$$

Esto es, la puntuación observada de un individuo es igual a la puntuación verdadera más el error. En el plano teórico, la puntuación verdadera puede entenderse como la media de las puntuaciones obtenidas por una persona en infinitas aplicaciones de un test (en diferentes momentos y condiciones), asumiendo que la forma de distribución de esas infinitas puntuaciones se aproxima a la normal.

La puntuación de error, por otra parte, es la suma de todos aquellos factores aleatorios que influyen y afectan el registro de los datos, introduciendo inconsistencia en el proceso y alejando la puntuación observada de la puntuación verdadera. Aun en ciencias naturales, donde se cuenta con instrumentos más precisos, existe esa posibilidad de error. En la medida en que controlemos las fuentes de error de una medición, más se acercará la puntuación observada (empírica) a la puntuación verdadera (teórica).

La confiabilidad puede entenderse como la exactitud o precisión de una medición, o el grado en el cual las puntuaciones de un test están libres de esos errores de medición. Esta exactitud o precisión de las puntuaciones permite que éstas se mantengan constantes en diferentes circunstancias.

Obviamente, debemos distinguir entre la falta de consistencia debido a un cambio real en el rasgo medido y las fluctuaciones azarosas de las puntuaciones ocasionadas por cambios en circunstancias personales u otras que afecten la situación de evaluación. Los cambios reales en

el atributo medido no son fuente de falta de confiabilidad, mientras que todo cambio debido a factores externos que se presenta en forma no sistemática afecta la confiabilidad de las puntuaciones.

Es importante diferenciar entre errores sistemáticos y aleatorios. Un error sistemático es aquel que genera constantemente una puntuación elevada o baja en una persona al margen de los cambios que puedan darse en el rasgo medido por el test. Este tipo de error se denomina sesgo de medición, que se define como toda aquella fuente de variación que sistemáticamente afecta a las puntuaciones de un determinado grupo que está siendo evaluado por un test, ya sea elevando o disminuyendo las mismas.

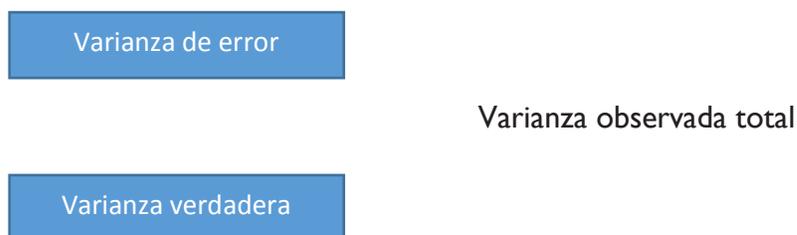
De esta manera, estos sesgos afectan directamente las inferencias o interpretaciones que podemos realizar a partir de esos puntajes. Un instrumento que mide aptitudes para la matemática puede estar construido de tal manera que sus ítems sean poco comprensibles para un grupo social o etario determinado, y esto podría ocasionar que ese grupo obtenga puntuaciones más bajas en el test independientemente de sus aptitudes matemáticas.

Si bien los sesgos están habitualmente presentes en todo proceso de medición, este tipo de error es predecible y los diferentes procedimientos que se exponen en el capítulo 4 sobre Validez apuntan a controlarlos. Por otra parte, los errores aleatorios que afectan la precisión o consistencia de las medidas realizadas son impredecibles y forman parte de todo proceso de medición.

El control de estas fuentes de error se relaciona con los métodos utilizados para verificar la confiabilidad.

En general, la confiabilidad se verifica mediante un coeficiente de correlación entre las medidas repetidas de un fenómeno. Para entender la lógica de la utilización de este coeficiente cabe realizar algunas observaciones previas. La puntuación verdadera también puede expresarse en términos de varianza de las puntuaciones de los tests. Recuérdese que la varianza indica la dispersión promedio de los valores (en este caso los puntajes de tests) alrededor de la media de un grupo de observaciones (más precisamente, el promedio de los cuadrados de la diferencia entre cada valor y la media). En la teoría clásica de los tests, la confiabilidad puede expresarse como la proporción de varianza observada de las puntuaciones de tests que se debe a la varianza verdadera (la variabilidad de la puntuación verdadera), es decir, a la varianza

del rasgo que se está evaluando y no a otros factores. Puede representarse lo afirmado anteriormente de la siguiente manera:



Los diferentes métodos utilizados para evaluar la confiabilidad utilizan el coeficiente de correlación como estadístico fundamental.

De este modo, un coeficiente de 0,80 sugiere que un 64% de la varianza observada es asimilable a la variabilidad de la puntuación verdadera, con un margen de error del 36%. Los diferentes métodos que se expondrán más adelante están diseñados para estimar la cuantía de error presente en las puntuaciones de un test determinado.

Como puede inferirse, la confiabilidad (como la validez) no es una característica del test en sí mismo, según la teoría clásica de los tests, sino una propiedad de las puntuaciones del test cuando éste se administra a una muestra específica y bajo condiciones particulares (APA, 1999). Ésta es una de las principales limitaciones de la TCT.

2.1.2 Métodos de la evaluación de la confiabilidad: temporal, equivalentes, consistencia interna.

En los métodos utilizados para corroborar la confiabilidad de un instrumento de medición se pueden distinguir dos instancias. Por una parte es necesario administrar el instrumento a una muestra según un diseño de investigación específico y, por otro lado, los datos que resulten de tal aplicación deben ser analizados mediante procedimientos apropiados para obtener un estadístico que represente la confiabilidad de las puntuaciones del test (Cortada de Kohan, 1999).

a) *Test-retest:*

Este método consiste en administrar un test en dos oportunidades a la misma muestra de sujetos, con un determinado intervalo entre las dos administraciones, y calcular la correlación entre los puntajes obtenidos en la primera y segunda vez.

Como se explicó previamente, todos los métodos utilizados para valorar la confiabilidad de un test se interesan por el grado de consistencia entre dos conjuntos de puntuaciones obtenidas independientemente, y por lo tanto pueden ser expresados en función de un coeficiente de correlación que exprese el grado de correspondencia o relación entre dos conjuntos de puntuaciones (Anastasi y Urbina, 1998).

b) *Formas equivalentes:*

A través de este método, también denominado “formas paralelas”, se puede evaluar la consistencia interna pero también la estabilidad temporal de un conjunto de puntuaciones. El procedimiento básico consiste en administrar dos formas equivalentes de un test a un mismo grupo de individuos. En el caso de que este método se utilice para verificar la estabilidad, la administración de la segunda forma se realiza transcurrido un tiempo a partir de la administración de la primera forma, y posteriormente se correlacionan los resultados obtenidos.

Si bien este método es el más completo para evaluar la confiabilidad, puesto que permite controlar la mayor cantidad posible de fuentes de error aleatorio (distinta muestra de ítems, distintas condiciones físicas y mentales de los examinados, diferente situación ambiental, posiblemente distinto administrador), en la práctica presenta algunos inconvenientes.

Para ser consideradas equivalentes, dos pruebas deben reunir ciertos requisitos, tales como tener las mismas características formales (cantidad de ítems, escala de respuesta, etc.) y estadísticas (tener medias y desviaciones estándar semejantes, coeficientes de correlación elevados entre ambas formas, etc.) (APA, 1999).

El procedimiento ha sido simplificado con la informatización de los tests, que posibilita utilizar las formas computarizadas como versiones alternativas de las originales en lápiz y papel. No obstante, tal como lo expresan las normas especiales de la

APA para tests computarizados, este tipo de tests no puede ser tratado automáticamente como forma paralela de los tests de lápiz y papel, sino que debe demostrarse el cumplimiento de los requisitos anteriormente mencionados c) *Partición en mitades*:

A través de este método se verifica la consistencia interna de las puntuaciones de un test, es decir, el grado en que las diferentes partes del test miden la misma variable. Se administra el test en una ocasión a una muestra de individuos y posteriormente se divide la prueba en dos mitades comparables, obteniendo de esta manera dos puntuaciones para cada individuo de la muestra. Finalmente, se correlacionan las puntuaciones correspondientes a ambas mitades del test por medio de un coeficiente de correlación.

Este método fue popular antes de que se dispusiera de computadoras personales, debido a que los estadísticos requeridos son más fáciles de calcular manualmente que el coeficiente alfa que se presenta en el apartado siguiente. Actualmente, el método de partición en mitades es poco empleado para verificar la consistencia interna de una prueba (Nunnally y Bernstein, 1995).

La dificultad inicial de este procedimiento es lograr que las mitades obtenidas sean realmente comparables. Muchos de los tests son contruidos con un nivel de dificultad creciente, y si se divide el test en la primera y segunda mitad seguramente éstas no resultarían comparables. Aun cuando no sean pruebas con dificultad creciente de sus ítems, otros factores pueden obstaculizar el contar con dos mitades estrictamente comparables.

En efecto, los examinados probablemente se vean más afectados por la fatiga hacia el final de la prueba, lo cual podría incidir en mayor medida en los puntajes de la segunda mitad.

El criterio habitualmente adoptado para dividir la prueba es el de separar los ítems del test en dos mitades, una de ítems pares y la restante de ítems impares. Un método más riguroso, pero que requiere mayor esfuerzo del investigador, exige el apareamiento de todos los ítems con un cierto criterio estadístico, asignándolos luego al azar a cada una de las partes o mitades del test.

d) *Métodos de covarianza de los ítems*:

Estos métodos comparten con el anterior dos aspectos importantes: por un lado, permiten verificar la consistencia interna de los puntajes del test y, por otra parte, requieren una sola administración de la prueba (Thorndike, 1989).

A partir de una única aplicación del test a una muestra se obtiene una estimación del grado de covarianza de los ítems, utilizando como estadístico el coeficiente alfa de Cronbach o la fórmula alternativa de Kuder Richardson (KR-20), cuando se trabaja con ítems dicotómicos (verdadero-falso, por ejemplo).

e) *Acuerdo entre examinadores:*

La dimensión evaluada por este método es la confiabilidad entre examinadores. El método consiste en administrar un test a una muestra, entregar los resultados (protocolos de respuesta) del test a un conjunto de jueces que los puntuarán independientemente.

A continuación, se verifica el grado de acuerdo que alcanzan los jueces luego de leer, registrar y codificar los mismos datos (Murat, 1985). Naturalmente, este procedimiento no se aplica en tests que se puntúan de manera objetiva (de opción múltiple, por ejemplo) y sólo adquiere importancia cuando interviene el criterio del examinador en el proceso de calificación.

2.2 Estandarización.

2.2.1 Definición de estandarización.

Se llama así al proceso mediante el cual se establecen procedimientos unívocos para la aplicación, calificación e interpretación de un test psicométrico (Cronbach, 1972). Cuando las condiciones de administración y calificación del test psicométrico están bien definidas y su utilización es idéntica en todos los sujetos examinados, entonces el aspecto más importante que queda por resolver es la interpretación de las puntuaciones logradas por los sujetos evaluados. Esta interpretación se realiza comparando el puntaje obtenido por el sujeto con las puntuaciones contenidas en el baremo o tablas de normas.

2.2.2 Estandarización en la calificación.

Los números que arrojan la medición de un atributo psicológico realizada con un test psicométrico se denominan puntajes o calificaciones directas. Estos puntajes en sí mismos no tienen un significado preciso, adquieren un significado psicométrico cuando se les compara con una tabla de normas o baremo, que ha sido previamente construida con las puntuaciones que en el test han obtenido un grupo de sujetos llamado grupo normativo. Al realizar esta comparación se puede hacer la clasificación de los sujetos examinados con lo cual se llega a cumplir la finalidad del test, que es clasificar a los sujetos examinados.

Los puntajes directos se transforman en varios tipos de puntajes derivados o unidades de medición que se presentan en las tablas de normas o baremos.

Un baremo es una tabla que sistematiza las normas (afirmación estadística del desempeño del grupo normativo en el test psicométrico) que transforman los puntajes directos en puntajes derivados que son interpretables estadísticamente. Puntajes derivados son los siguientes: a) percentiles, puntaje derivado que transforma el puntaje directo en una escala del 1 al 100, llamándose cada uno de los puntos un centil. Un examinado que tiene un puntaje directo equivalente al percentil 80, se dice que supera al 80% del grupo normativo en el aspecto evaluado; b) puntajes estándar, que son aquellos que tienen como unidad a fracciones de la desviación estándar, ejemplos: el puntaje CI, el puntaje enecatipo (escala del 1 al 9), el puntaje decatipo (escala del 1 al 10), entre otros. También es un puntaje derivado la edad mental.

2.3 Validez

2.3.1 Definición de validez.

La validez es un aspecto esencial de la medición psicológica y se relaciona con la investigación del significado teórico de las puntuaciones obtenidas por medio de un test (Oliden, 2003).

Las puntuaciones evidencian propiedades de validez cuando se verifica que el test realmente mide el constructo que pretende medir, justificando adecuadamente las inferencias realizadas en función de sus resultados (Nunnally, 1991).

Recordemos que “constructo” es la representación abstracta de un conjunto de comportamientos relacionados. Así, por ejemplo, “depresión” o “aptitud matemática” son constructos. Se podría construir un inventario de 20 ítems para medir depresión, pero alguna dimensión del constructo (como el componente emocional de la depresión) podría estar escasamente representada por el contenido de ese test. Esa dimensión no está adecuadamente representada por los ítems de un test ocasiona la *sub-representación del constructo*. Por otra parte, es probable que las puntuaciones en un inventario reflejen una tendencia a dar respuestas socialmente deseables (aceptables) por parte del examinado. Este aspecto de las puntuaciones no relacionadas con el verdadero propósito de medición del test se denomina *varianza irrelevante del constructo* (Hogan, 2004).

La situación ideal en lo referente a la validez es que un test represente adecuadamente y mida la varianza relevante del constructo, o expresado de otro modo, que las interpretaciones de los resultados de una prueba estén libres de sesgo de medición.

Según Oliden (2003), la teoría de la validez se relaciona con el concepto de sesgo, definido como un error sistemático que produce distorsión en las puntuaciones adulterando su significado teórico. Tal como afirma Muñiz (1998), el hecho que las puntuaciones de un test sean confiables es una condición necesaria pero no suficiente para que sean válidas.

A pesar de su importancia, el concepto de validez es uno de los más complejos y controvertidos de la teoría de los tests (Angoff, 1988; APA, 1999). Una breve introducción histórica puede contribuir a esclarecer la significación actual de este concepto.

Para Oliden (2003), la evolución de las teorizaciones acerca de la validez, desde un enfoque operacional y pragmático hasta la concepción contemporánea, refleja los cambios que se fueron dando en la ciencia psicológica en general.

En la historia del concepto de validez pueden identificarse tres etapas principales. Una primera etapa, operacional, en la que predomina una perspectiva exclusivamente pragmática de las aplicaciones de los tests. Este enfoque coincide con el operacionalismo dominante en la epistemología de la primera mitad del siglo pasado y se manifiesta en la noción de validez como sinónimo de la correlación entre las puntuaciones de un test y algún criterio que el test intenta predecir (Martínez Arias, 1995).

Como señala Angoff (1988), la concepción de validez con un sentido meramente predictivo dominó el escenario de la psicometría hasta los años cincuenta.

Posteriormente se comprendió que este concepto de validez exclusivamente ligado a la predicción de criterios externos no era útil para muchos tests en los que ellos mismos constituyen su propio criterio (por ejemplo, en pruebas de rendimiento) y esto condujo a introducir el concepto de validez de contenido.

Otro cambio importante se produjo con la aparición del clásico artículo de Cronbach y Meehl (1955) donde se presentó por primera vez el concepto de validez de constructo y se caracterizó a esta última como el aspecto fundamental e inclusivo de las restantes dimensiones de la validez (Martínez Arias, 1995). Esta publicación inició un segundo estadio teórico, en el

cual asume un papel fundamental la teoría psicológica. En esta fase se diferencian tres tipos de validez: de constructo, de contenido y predictiva.

Por último, el período actual o contextual se caracteriza por una extensión de la concepción anterior, a la que se agrega la importancia otorgada al uso propuesto para el instrumento. Esto significa que, en realidad, nunca se valida un test en sí mismo sino que su validez se verifica para determinados propósitos.

En esta nueva perspectiva ya no se habla de distintos tipos de validez sino de un proceso de recolección de diferentes tipos de evidencia para un concepto unitario.

Esta concepción contemporánea de validez se refleja en la última versión de las Normas Técnicas para los Tests Psicológicos y Educativos (APA, 1999), donde se la define como la adecuación, significación y utilidad de las inferencias específicas hechas a partir de las puntuaciones de los tests. Como expresamos, la validez es un concepto unitario y siempre se refiere al grado en que la evidencia empírica apoya las inferencias realizadas en función de los resultados de un test. La APA (1999) propuso cinco tipos de evidencia de validez, basadas en: el contenido del test, la estructura interna del test, el proceso de respuestas al test, las relaciones con otras variables externas al test y las consecuencias de su aplicación.

De esta manera, para verificar la validez de las inferencias realizadas a partir de las puntuaciones de un test se utilizan procedimientos semejantes a los implementados para contrastar cualquier hipótesis científica, vale decir la recolección de evidencias que confirmen o refuten esas inferencias.

En el estudio de la validez de constructo estas evidencias están relacionadas a cinco aspectos: a) Contenido (relevancia y representatividad del test); b) Sustantivo (razones teóricas de la consistencia observada de las respuestas); c) Estructural (configuración interna del test y dimensionalidad); d) Generalización (grado en que las inferencias hechas a partir del test se pueden generalizar a otras poblaciones, situaciones o tareas); e) Externo (relaciones del test con otros tests y constructos); f) Consecuencia (consecuencias éticas y sociales del test) (Messick, 1995).

2.3.2 Validez de constructo.

Validez de Constructo.- El constructo viene a ser un concepto hipotético que forma parte de las teorías que intentan explicar la conducta humana: inteligencia, creatividad, dependencia de campo, etc. La validez de constructo es la obtención de evidencias que apoyan que las conductas observadas en un test son (algunos) indicadores del constructo. Este tipo de validez responde a la pregunta "¿cómo se puede explicar psicológicamente la puntuación del test?". La respuesta a esta pregunta puede verse como la elaboración de una "miniteoría" acerca de una prueba psicológica. La lógica de la validez de constructo en muchos aspectos así como en sus métodos, es esencialmente la del método científico.

El proceso de validación de constructo implica a partir del establecimiento de deducciones de la teoría:

- a) Formular hipótesis y relaciones entre elementos del constructo, de éste con otros constructos de la teoría y con otros constructos externos.
- b) Seleccionar ítems o tests (indicadores) que representen manifestaciones concretas del constructo.
- c) c) Recogida de datos.
- d) d) Establecer consistencia entre datos e hipótesis, y examinar el grado en que los datos podrían explicarse mediante hipótesis alternativas.
- e) Hay diversos procedimientos para establecer la validez de constructo. Si elaboramos una "miniteoría" esta tendrá tres pasos: (1) en base a la teoría sostenida en ese momento respecto del test, el psicólogo deduce ciertas hipótesis sobre la conducta esperada de las personas que obtienen puntajes diferentes en el test, (2) se reúne datos que confirman o no esas hipótesis, (3) en base a los datos acumulados, se toma la decisión relativa a si la teoría explica adecuadamente los datos. Si no es así se tiene que revisar la teoría y repetir el proceso hasta lograr una explicación más adecuada. El proceso de validación, en ese sentido, es de continua reformulación y refinamiento.

Al determinar la validez de construcción, el propósito es identificar todos los factores que influyen en la ejecución del test y determinar el grado que influyen cada uno de ellos.

Ejemplo: Un psicólogo construye un test de ansiedad y elabora una "microteoría" cuya contrastación le dirá si el test tiene validez de constructo. Las hipótesis a verificar son las siguientes:

1. Los que obtienen puntuaciones elevadas acabarán probablemente en clínicas psiquiátricas en comparación con aquellos de puntuaciones más bajas.
2. Será más fácil que les receten drogas psicotrópicas a los que tienen altas puntuaciones que a los de bajos puntajes.
3. Los hijos de los de puntuaciones altas tendrán mayores probabilidades de tener una puntuación alta en el test que los hijos de quienes tuvieron puntuaciones bajas.
4. El test de ansiedad se correlacionará alta y significativamente (más allá de 0.60) con otros test de ansiedad.
5. El test de ansiedad no se correlacionará con variables que no resulten conexas con la misma.
6. Los grupos psiquiátricos caracterizados como ansiosos alcanzarán en el test unas puntuaciones más altas que los del grupo control.

2.3.3 Validez de criterio.

Validez Predictiva (evidencia externa) - ¿Predicen las puntuaciones del test un rendimiento o conducta futura? (Junto con la validez concurrente se le denomina también validez empírica del test). Un uso común de los tests es predecir la conducta futura; utilizamos el test para ayudarnos a tomar alguna decisión práctica (selección, clasificación, etc.). En cada una de estas situaciones, cuanto mayor es la exactitud de predicción del resultado (es decir del criterio externo), tanto más útil será la prueba. Por ejemplo, el test será un componente aceptable de un proceso de selección de personal, si sus calificaciones o puntuaciones predicen la ejecución de algún componente importante del trabajo (criterio externo); en otras palabras, para que el test se pueda utilizar como parte de un proceso de selección es preciso demostrar la validez de la prueba relacionándola con los criterios pertinentes. En este sentido, el contenido de la prueba pasa a tener un lugar secundario, siendo el interés fundamental del psicólogo el averiguar si el test predice un criterio determinado. se relacionará las puntuaciones del test sean criterios validos y confiables.

Un criterio es cualquier desempeño que los sujetos tienen en la vida real, por ejemplo, las medidas de rendimiento académico, medidas de rendimiento laboral, clasificaciones psiquiátricas, etcétera. En muchos casos resulta imposible hallar un criterio no ambiguo de un rasgo mental. Por ejemplo, dos psicólogos, Carla y Abel, que investigan el rasgo de aptitud

numérica pueden emplear diferentes criterios externos para correlacionar los puntajes del test que han creado. Así, Carla puede considerar que el criterio externo más adecuado son las calificaciones que reciben los sujetos en un curso de mecánica en taller; mientras que Abel puede considerar como criterio el periodo de tiempo que gastan los estudiantes en aprender una tarea mecánica y sencilla durante el entrenamiento en un fabrica. ¿Qué sucede si las pruebas que emplean ambos psicólogos correlacionan 0.006 con uno de los criterios, y 0.70 con el otro?, ¿cómo podemos afirmar que la prueba es valida cuando arrojan resultados de cierta clase?, ¿se trata en verdad de una prueba de aptitud mecánica? En razón a situaciones como esta se llevo a la conclusión de que la validación de un test es un proceso largo y no un hecho aislado. Solamente a través de estudios de correlación con una amplia variedad de criterios podremos comprender que mide la prueba. Así, una serie de investigaciones sobre la “prueba de actitud mecánica” nos puede demostrar que en realidad esta midiendo la habilidad para realizar movimientos fino y cuidadosamente controlados, siendo completamente independiente para comprender las reacciones complejas de la piezas mecánicas. De esta manera el test puede tener una alta correlación con las calificaciones obtenidas en el taller y ninguna con los trabajos e maquinarias. En el proceso de validación, la validez predictiva de un test (y también la concurrente) se expresa generalmente por medio de un coeficiente de correlación entre los puntajes y los denominados criterios. Este coeficiente se denomina coeficiente de validación. La interpretación de este coeficiente requiere un dominio excelente del análisis estadístico utilizado para obtenerlo. Después del criterio, los procedimientos estadísticos adquieren vital importancia para obtener esta categoría de validez. Incluso para un mejor análisis es conveniente contar con el dispersigrama o scattergrama o “nube de puntos” (gráfica del coeficiente de correlación entre las puntuaciones del test “X” y las del criterio “Y”).

2.3.4 Validez de contenido.

Validez de Contenido (evidencia del contenido).- ¿Los ítems que constituyen el test son realmente una muestra representativa del dominio de contenido o dominio conductual que nos interesa?

Es conveniente precisar que un dominio o campo conductual es una agrupación hipotética de todos los reactivos posibles que cubren un área psicológica particular. Al hablar de este conjunto de reactivos posibles, se emplean los términos de dominio, universo o población conductual como sinónimos. Por ejemplo: Un test de vocabulario debe ser una muestra adecuada del dominio o universo de ítems posibles en esta área.

La validez de contenido consiste en determinar lo adecuado del muestreo de reactivos del universo de reactivos posibles; en este sentido, es una “medida” de lo adecuado del muestreo. Ponemos “medida” entre comillas debido a que este tipo de validez consiste en una serie de estimaciones u opiniones, que no proporcionan un índice cuantitativo de validez (para su obtención no se utilizan procedimientos estadísticos). Este tipo de validez se asocia fundamentalmente a los tests de aprovechamiento o rendimiento (test de matemática, historia, etcétera); aunque no existen razones para que no pueda aplicarse a los otros tipos de pruebas psicológicas (pruebas de aptitudes, habilidades, etcétera).

Para su determinación se compara sistemáticamente los reactivos del test con el dominio conductual del contenido postulado. Por ejemplo: si tenemos una lista de 500 palabras que esperamos que los estudiantes de un curso sean capaces de escribirlas correctamente al final de este, su performance o rendimiento respecto a estas palabras será importante solamente en tanto que proporciona una prueba de su habilidad para escribir correctamente las 500 palabras. El test que construyamos tendrá una muestra de las 500 palabras, pero sólo tendrá validez de contenido en la medida en que proporcione una muestra adecuada de las 500 palabras que represente. Si seleccionamos solamente palabras fáciles o difíciles, o palabras que representen únicamente ciertos tipos de faltas comunes de ortografía, estaríamos propensos a obtener una validez de contenido muy baja. En consecuencia, el aspecto clave en la validez de contenido es el muestreo de los reactivos. En otras palabras, la validez de contenido es cuestión de determinar si la muestra de sus reactivos es representativa del universo o dominio conductual de ítems al que supuestamente representa. Para hacer esta determinación se recurre a “jueces” (o expertos, generalmente en número impar). El proceso es básicamente lógico y racional, los distintos jueces pueden no estar de acuerdo en la validez de contenido de un test; por ejemplo, la falta de claridad en la especificidad del dominio conductual, hará que resulten difíciles los juicios de validez de contenido

Unidad 3

Normas de calificación.

3.1 Normas de calificación y tipificación.

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.

3.2 Desviación estándar.

La desviación estándar es un índice numérico de la dispersión de un conjunto de datos (o población). Mientras mayor es la desviación estándar, mayor es la dispersión de la población. La desviación estándar es un promedio de las desviaciones individuales de cada observación con respecto a la media de una distribución.

La desviación estándar es una medida de la dispersión de un conjunto de puntajes alrededor de la media. Para obtener la desviación estándar se empieza por restar la media de cada uno de los puntajes, con lo cual se llega a una nueva serie de valores denominados puntajes de desviación. Luego se elevan al cuadrado estos puntajes de desviación, se suman los cuadrados y se divide la suma por el número de valores que integran la serie, con el fin de obtener la desviación cuadrática media o variando.

3.3 Comparación interindividual (confronta de resultados de otros sujetos).

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.

3.4 Comparación intraindividual (dos medidas del mismo sujeto)

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.*

Unidad 4

PROCESOS DE APLICACIÓN DE PRUEBAS

4.1 Procedimiento de aplicación de pruebas: requisitos de aplicación, criterio de inclusión y exclusión.

Para asumir el procedimiento de aplicación de los test el examinador debe:

- Tener la habilidad que lo capacita para realizar la aplicación y calificación de pruebas.
- Antes de aplicar un test, familiarizarse con el modo en que se aplica ese test en particular.
- Ser riguroso seguidor de las instrucciones para la aplicación que tiene cada test.
- Observar una actitud científica e imparcial y no dejarse llevar por su inclinación de ayudar al examinado, proporcionándole datos que faciliten su ejecución en la prueba.
- Evitar expresiones mímicas que den, de una manera inconsciente, información al examinado al aplicar un test debemos permanecer indiferentes ante los errores a los aciertos del sujeto.
- Mantener una actitud amable con el examinado.
- Respetar las directrices éticas acerca del uso de los test.

Durante el procedimiento de aplicación de los test, el examinador debe:

- Considerar las condiciones ambientales en que se efectúa la aplicación.
- Observar el estado físico y psicológico en que se encuentra el examinado.
- Ofrecer instrucciones claras y precisas.
- Asumir responsabilidad en cuanto a la fidelidad con que expone las instrucciones.
- Los sujetos pueden preguntar todo lo que deseen, pero el examinador debe adelantarse a la mayor parte de las preguntas mediante instrucciones completas.
- Evitar excesiva rigidez.
- Prestar atención ante los posibles accidentes, como puede ser que uno de los examinados se enferme.
- Estar al tanto de las preguntas que pueden hacer los examinados.
- Ante preguntas no contempladas en las instrucciones proceder con cautela y juicio común.

4.2 Calificación de la prueba.

Esta fase requiere cautela y habilidad; a partir de la puntuación alcanzada en los testse elaborarán las hipótesis interpretativas que sirven de orientación a la labor diagnóstica, se puede decir, que el destino final de la puntuación es la elaboración de dichas hiotesis.

Cada test tiene su modo particular para ser aplicado y tiene su propio sistema de calificación. Es importante insistir que la calificación exige del cumplimiento estricto de los procedimientos que están establecidos para la puntuación.

En ocasiones, se producen observaciones por el examinador a favor de criterios acerca de que el sujeto no entendió algunos ítems de la prueba, o que tal respuesta es ocasionada por situaciones que distrajo la atención del sujeto, etc. estos criterios deben ser reflejados en las observaciones pero no pueden estas consideraciones, en modo alguno, modificar el sistema de calificación de la prueba.

Cada una de las pruebas psicológicas aplicadas deben reflejar datos generales del sujeto que las realiza así como la fecha de realización de la misma

4.3 Estructura y calificación del informe.

El informe es un documento oficial, en el cual se plasma el resultado de la investigación psicométrica realizada a una persona mediante la aplicación y calificación de instrumentos de medida psicológica, este informe no es definitorio de la evaluación psicológica que se le realiza un apersona, pero tiene valor judicial; dicho informe, conjuntamente con las hojas de respuestas sus perfiles, se le entrega solamente al psicólogo.

La confección del informe psicométrio dee cumplir los siguientes requisitos:

I.- Tiene que consignar los datos generales de sujeto examinado.

- Nombre
- Sexo
- Edad
- Fecha de nacimiento
- Edad
- Escolaridad

- Estado civil
- Dirección
- Número de historia clínica
- Donde estudia o trabaja
- 2.- presentar los datos de la observación realizada al sujeto, durante todo el proceso de aplicación, y ante cada una de las pruebas de modo específico, utilizando para ello una correcta redacción y ortografía, con lenguaje claro y preciso manejando el uso de los términos técnicos.
- A dicho informe se le adjuntan las pruebas y calificadas u hojas de respuestas con sus perfiles elaborados.
- El resultado de la calificación de las pruebas debe aparecer en cada test u hoja de respuesta y sus perfiles.
- En cada test, hoja de respuesta y perfil tiene que aparecer el nombre y apellidos de la persona, edad, sexo, escolaridad, y demás datos que pida a prueba, así como la fecha del día en que se realiza la misma y la hoja de comienzo y término de dicha prueba.
- En las observaciones puede aparecer recomendaciones al psicólogo a acerca del sujeto evaluado, acerca de alguna de las pruebas o de la evaluación en general; información que resulta valiosa para todo el proceso de evaluación.

4.5 El psicodiagnóstico y diagnóstico.

Una de las funciones del profesional psicólogo, tal vez incluso una de las que es considerada más típica del quehacer del psicólogo, es la de efectuar diagnósticos psicológicos. El término diagnóstico es ampliamente utilizado, no sólo en el ámbito de la psicología o de las ciencias de la salud, sino que en el contexto de muchas de las disciplinas profesionales, y pareciera que en general se entiende su significado. Sin embargo, cuando se busca una definición más precisa, nos encontramos con diversas ideas y conceptos vinculados a este término, como por ejemplo los de medición y evaluación, entre otros, para lo cual utilizamos pruebas o test psicológicos

además de otros instrumentos. Se hace necesario, entonces, precisar el significado que estos conceptos tienen en este contexto.

Las definiciones de diagnóstico son variadas, y en su mayoría derivadas del quehacer médico o del modelo médico. En el campo de la psicología, podemos decir que el psicólogo intentará hacer un psicodiagnóstico cuando, a través de los métodos propios de la psicología, intente llegar a una conclusión acerca de las características de una persona, de un grupo de personas o de un hecho psicológico en sí.

Por otra parte, es importante destacar que el psicodiagnóstico es el resultado de un examen detallado o de un estudio minucioso de algunas muestras de conducta. No debemos perder de vista el hecho de que los fenómenos psíquicos (o procesos psicológicos) no los podemos observar o medir directamente, sino sólo su exteriorización o manifestación a través de alguna forma de comportamiento o conducta.

4.5 Manejo y aplicación de la prueba proyectiva: dibujo de la figura humana de Karen Machover. (Se proporcionará a alumno manual completo en formato PDF de la prueba de dibujo de la figura humana de Karen Machover).

4.5.1 Calificación de la prueba proyectiva: dibujo de la figura humana de Karen Machover. (Se proporcionará a alumno manual completo en formato PDF de la prueba de dibujo de la figura humana de Karen Machover).

5.5.2 Interpretación y redacción de informe de la prueba proyectiva: dibujo de la figura humana de Karen Machover. (Se proporcionará a alumno manual completo en formato PDF de la prueba de dibujo de la figura humana de Karen Machover).

4.6 Manejo y aplicación de la prueba BETA III (Se proporcionará a alumno manual completo en formato PDF de la prueba BETA III).

4.6.1 Calificación de la prueba de inteligencia BETA III. (Se proporcionará a alumno manual completo en formato PDF de la prueba BETA III).

4.6.2 Interpretación y redacción de informe de la prueba de inteligencia BETA III. (Se proporcionará a alumno manual completo en formato PDF de la prueba BETA III).

4.7 Manejo y aplicación de la prueba neuropsicológica de BENDER sistema de puntuación Koppitz. (Se proporcionará a alumno manual completo en formato PDF de la prueba BENDER).

4.7.1 Calificación de la prueba neuropsicológica de Bender. (Se proporcionará a alumno manual completo en formato PDF de la prueba BENDER).

4.7.2 Interpretación y redacción de informe de la prueba Bender. (Se proporcionará a alumno manual completo en formato PDF de la prueba BENDER).

Bibliografía

- González Llaeza Felicia Miriam, instrumentos de evaluación psicológica, Edit. Ciencias médicas.
- Morales María Luisa, Psicometría aplicada. Edit. Trillas.
- Tornimbeni Silvia, Introducción a la psicometría. Edit. Paidós.