



Universidad de Ixtlahuaca CUI, A.C.  
Licenciatura en Psicología  
Incorporada a la UAEMéx Clave 091-D



Elaboración de Instrumentos

Antología

Por

Luis Casas Vilchis

Ixtlahuaca, Méx.

Enero, 2017

## Índice

Antecedentes: .....	2
Aportaciones a la Teoría de la Medida .....	4
Conceptos Básicos .....	7
Calificar.....	7
Dato.....	7
Medición .....	7
Evaluación.....	8
Principios de la Evaluación .....	8
Funciones específicas: .....	8
Tipos de Evaluación: .....	9
Sustento de la Ciencia .....	9
Concepto, Constructo, Factor, Variables, Indicadores, Índices .....	10
Concepto .....	10
Constructo .....	10
Definición conceptual .....	11
Definición operacional .....	11
Factores o variables.....	11
Índice .....	12
Formas de medición de las variables .....	12
Pasos para la operacionalización de las variables .....	12
Niveles de Medición .....	12
Nivel Nominal.....	13
Nivel Ordinal.....	13
Nivel Intervalar .....	13
Nivel de Razón o Proporción .....	13
Características del Examinador o Evaluador .....	13
Características de la Población.....	14
Contextos de evaluación.....	15
Alcance de la evaluación.....	15
1.1 Valoración de necesidades.....	15

1.2 Seguimiento y control de programas .....	15
A nivel del programa:.....	16
A nivel de la población.....	16
1.3 Valoración del impacto .....	16
2. Esquema del alcance .....	16
Ética en la evaluación.....	17
Derechos Humanos: .....	17
Consideraciones en el proceso de valoración .....	17
Aspectos individuales.....	17
Aspectos sociales.....	18
Código Ético del Psicólogo .....	18
Consentimiento Informado.....	19
Consentimiento informado en el caso de niños .....	20
Pasos para elaborar un instrumento .....	20
Planteamiento del problema: .....	20
Determinar las características del fenómeno de estudio .....	21
Estructuración del instrumento y su aplicación .....	21
Procesamiento de la información y análisis de resultados .....	21
Conclusiones y sugerencias .....	22
Adaptación de los test .....	22
Métodos de Adaptación de los Test.....	22
Técnicas de Traducción.....	22
Métodos para establecer la equivalencia entre Test .....	23
Fuentes de Sesgo .....	24
a) Sesgo de constructo:.....	25
b) Sesgo metodológico: .....	25
Naturaleza y uso de las pruebas psicológicas (Instrumentos y Escalas de Medición).....	26
Encuesta .....	26
Cuestionario .....	26
Test.....	27
Examen .....	27
Inventario .....	28
Escalas de Actitudes .....	28

Clasificación de los Instrumentos.....	30
Desde el punto de vista de su objetivo.....	30
De ejecución máxima o de potencia. ....	30
Pruebas de ejecución típica: .....	30
Por la forma o tipo de respuesta que exigen .....	30
a) Pruebas objetivas.....	30
b) Pruebas subjetivas. ....	30
Por la forma de administración de la prueba .....	30
a) De aplicación colectiva.....	30
b) De aplicación individual.....	31
c) <b>Autoadministradas</b> .....	31
De acuerdo a la libertad de ejecución .....	31
a) Pruebas de poder .....	31
b) Pruebas de velocidad: .....	31
c) Pruebas de libre ejecución: .....	31
Por la forma de dar instrucciones .....	31
a) Oral .....	31
c) Escrita.....	31
d) A través de ejemplos .....	32
De acuerdo al material empleado .....	32
a) Lápiz y papel .....	32
b) Verbal.....	32
c) Manipulable.....	32
d) Aparatos especiales .....	32
Por su forma de calificación .....	32
a) Contando puntos.....	32
b) Interpretación .....	32
De acuerdo a los límites de edad .....	32
Según el modo de expresión de la respuesta.....	33
De acuerdo con el sector que exploren .....	33
Reactivo .....	33
Requisitos para el Desarrollo de Reactivos .....	34
a) Pregunta directa o afirmación .....	34

b)	El tema central del reactivo, debe referirse a una sola idea.....	34
c)	Complejidad .....	34
d)	El fenómeno a apreciar .....	34
e)	Claridad en la expresión .....	34
f)	Brevedad en la expresión.....	35
g)	Construcción gramatical engorrosa .....	35
<b>h)</b>	<b>Afirmaciones tímidas o ambiguas .....</b>	<b>35</b>
<b>i)</b>	<b>Evitar adiciones inconexas y falsas.....</b>	<b>35</b>
<b>j)</b>	<b>Evitar fuentes ajenas de dificultad de los reactivos.....</b>	<b>35</b>
<b>k)</b>	<b>Evitar las formas estereotipadas .....</b>	<b>35</b>
<b>l)</b>	<b>Incluir un máximo de términos comunes en el reactivo .....</b>	<b>36</b>
<b>m)</b>	<b>Los reactivos deberán presentarse en forma afirmativa.....</b>	<b>36</b>
<b>n)</b>	<b>Resaltar cualquier forma negativa de presentación.....</b>	<b>36</b>
<b>o)</b>	<b>Procurar evitar el uso de la palabra NO.....</b>	<b>36</b>
<b>p)</b>	<b>Sesgo cultural y de género.....</b>	<b>36</b>
Tipos de Reactivos .....		36
a)	Reactivos de Verdadero y Falso .....	36
b)	Reactivos de Selección Múltiple (Opción múltiple) .....	36
c)	Reactivos de Aparejamiento (Relación de columnas).....	37
d)	Reactivos de Jerarquización .....	37
e)	Reactivos por Completar (Complementación).....	37
f)	Reactivos de Ensayo.....	37
g)	Problemas.....	37
h)	Reactivos de Ejecución.....	37
Número de reactivo: .....		37
Dificultad y Discriminación del los Reactivos .....		38
a)	Índice de Dificultad o Aprobación del Reactivo .....	38
b)	Índice de Discriminación del Reactivo .....	38
<b>Validez y Confiabilidad .....</b>		<b>39</b>
<b>Validez.....</b>		<b>39</b>
Tipos de validez .....		39
Validez de contenido.....		39
Validez de Contenido a través de un Cuadro Comparativo .....		41

Validez de criterio .....	43
Validez de constructo .....	44
Nomología.....	44
Otros Tipos de Validez.....	45
<b>Confiabilidad</b> .....	45
Tipos de Confiabilidad .....	45
Confiabilidad Test-Retest:.....	45
Confiabilidad de Pruebas Paralelas: .....	46
Confiabilidad por Mitades:.....	46
Consistencia Interna: .....	46
Estandarización o normalización.....	47
Fractiles .....	47
Estaninas .....	47
Puntuación $z$ .....	48
Puntuación $T$ .....	48
Estructuración del Instrumento.....	49
1. Elementos de la cabeza:.....	49
1.1. Datos de Identificación .....	49
1.2. Objetivo .....	49
1.3. Confidencialidad .....	49
1.4. Instrucciones.....	49
1.5. Agradecimiento por la participación.....	50
2. Elementos del cuerpo .....	50
2.1. Protocolo de respuestas .....	50
2.2. Cuaderno de reactivos .....	51
Ficha Técnica .....	52
Referencias.....	53

La palabra *antología* emana del griego: *anthos*, que puede traducirse como *flor*, y el verbo *legein*, que significa *escoger*. En la actualidad hace referencia a la recopilación de textos o párrafos seleccionados de los mismos, que por su valor o precisión sirven como referencia a considerar para futuras consultas o lecturas, y no está exenta de la aportación de quien se encarga de recopilar la información a través del parafraseo o la presentación de comentarios y ejemplos que apuntalen o contextualicen la información, en el sentido de tratar de hacer más entendibles las ideas de los textos presentados, evidentemente siempre respetando y haciendo presente la autoría original.

La elaboración de pruebas psicométricas o *test*, es una labor importante y delicada al tratar de hacer evidente las características psicológicas de quienes las responden, y para este quehacer es importante el contar con elementos que permitan señalar con claridad la validez y confiabilidad de la información de quien participa en el proceso de medición psicométrica y de los postulados que habrán de sustentar la información obtenida.

El formular una antología relacionada con la unidad de aprendizaje *Elaboración de Instrumentos*, no es una labor sencilla en el hecho de que hay una amplia colección de textos relacionados con el tema y en los cuales se hacen aportaciones por demás valiosas y útiles para la elaboración de un instrumento de medición en psicología y por tal, el seleccionar aquellos textos que sean en sí los *más valiosos*, en el sentido de que su aportación sea de manera esencial de lo más significativa para cubrir los aspectos básicos de la elaboración de una prueba psicométrica no está exenta de hacer exclusiones involuntarias.

Por tal, la presentación de una *Antología de Elaboración de Instrumentos*, busca hacer el primer acercamiento en la estructuración de un documento que permita apoyar al conocimiento en el proceso inicial de esta labor, por lo que cualquier aportación que permita enriquecer, precisar o corregir los preceptos vertidos será bienvenida.

Luis Casas Vilchis

**Antecedentes:**

A lo largo de la historia el ser humano a tratado de establecer criterios que le permitan diferencia tanto los objetos entre sí como a las personas mismas, de ese modo se han establecido procesos de medición que permiten observar las diferencias existentes aún entre objetos o personas que aparentemente tienen características muy similares.



Por ejemplo, para medir la longitud existente entre dos puntos se han creado medidas de tipo arbitrario pero que son aceptadas por la sociedad y se establecen medidas como el centímetro o la pulgada, para el caso del peso de los metales preciosos como el oro se establece el quilate, para el peso el gramaje, la luz los lúmenes, etc.



De igual forma para las ciencias sociales, considerando una perspectiva positivista, se ha buscado establecer criterios que permitan medir las conductas típicas y atípicas de los grupos sociales y los individuos que la conforman



En la historia contemporánea, en Francia, a principios del siglo XX (1905), Binet creó una prueba que ayudaría a colocar a los niños parisienses en edad escolar en la clase apropiada, una década después se establece esta misma prueba, con los ajustes necesarios para el mismo proceso en Estados Unidos.



Para 1917 con el inicio de la primera guerra mundial, en Estados Unidos se empezaron a emplear las primeras pruebas para encontrar en los reclutas probables problemas intelectuales y emocionales, siendo aun más fuertes su dependencia en la segunda guerra mundial e introduciendo pruebas y condiciones de aplicación más estrictas y sistematizadas, que permitían llevar a cabo un mejor proceso de selección



Esta fuerte dependencia a las pruebas trajo posteriormente el uso de manera más amplia dentro de las organizaciones, las instituciones de salud mental y las instituciones educativas en la década de los 50's y 60's, manteniendo desde entonces un papel relevante en el proceso de valoración psicológica, a pesar de cuestionamientos principalmente de índole ético que han sufrido.



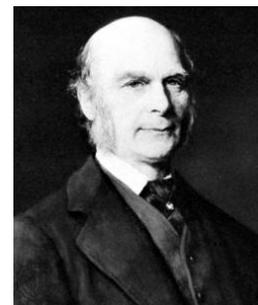
### Aportaciones a la Teoría de la Medida



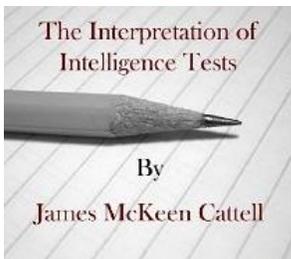
Wilhelm Wundt: Fundador del primer laboratorio de psicología experimental, en Leipzig (1879), a través del estructuralismo o anatomía de los procesos consciente utilizando el método introspectivo (mirar hacia adentro), en los cuales los sujetos examinan y describen sus experiencias conscientes en condiciones controladas de laboratorio. Estos experimentos se basaban principalmente al estudio de la sensibilidad a estímulos visuales, auditivos y de otros sentidos, y el tiempo de reacción simple. Esto fue el indicio de la formación de los primeros tests psicológicos. El estructuralismo basa los procesos conscientes en tres elementos: sensaciones, imágenes y sentimientos.



Francis Galton: Publicó su obra *El Genio Heredado*, y después, *Investigación del Desarrollo de las Facultades Humanas*, aquí se encuentran los principios de los estudios científicos de las diferencias individuales y sus primeros intentos por medirlas. Creó el primer laboratorio para la medición antropométrica, para la medición de las características físicas y sensorio motoras de los seres humanos. Estableció los principios del empleo del método estadístico como auxiliar importante para obtener medidas más objetivas y precisas en sus investigaciones. Utilizó el concepto de correlación, posteriormente desarrollado por Karl



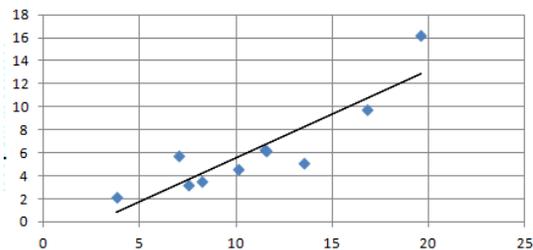
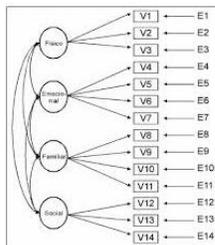
Pearson. Con él se inició el estudio científico de las diferencias individuales y la evaluación de las pruebas mentales, apoyado en la instrumentación matemática.



James McKeen Cattell: Analizó exhaustivamente el desarrollo y depuración de los instrumentos de medición psicológica, empleando por primera vez la expresión “test mental”, los cuales en un principio fueron creados solo para la medición de capacidades sensoriales y otras funciones simples, desde el punto de vista del empirismo inglés, el cual afirmaba que el conocimiento se adquiría mediante la experiencia; de tal, que las medidas relevantes fueron de tipo sensorio motriz: tiempo de reacción, capacidades auditivas y visuales, sensibilidad al dolor, al gusto, etc.



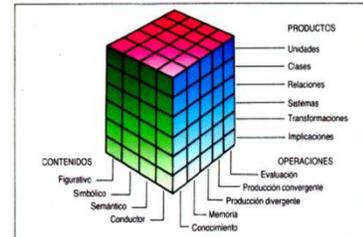
Spearman: Establece uno de los primeros modelos de inteligencia general, el cual propuso una teoría de la estructura intelectual, basada en el análisis estadístico de datos de pruebas psicológicas. Propuso una teoría de dos factores: La capacidad general, o un factor de energía mental, que constituía un factor de razonamiento (denominado g. Además se suponía que cada prueba específica medía habilidades específicas que eran para esa prueba en particular. El segundo componente incluía pruebas específicas. De esta manera, aunque se le asignaba importancia primordial al factor general, la teoría reconocía que era importante tomar en consideración otros factores específicos.



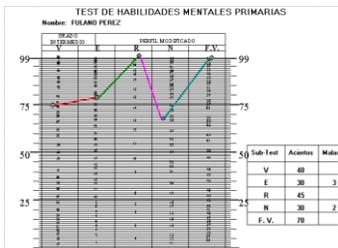
J. P. Guilford: Concibe que el funcionamiento intelectual tiene tres dimensiones: operaciones, contenidos y productos. Las operaciones son los procesos incluidos en la conducta intelectual (cognición, memoria, pensamiento divergente, pensamiento convergente y evaluación). Los contenidos pueden ser: figurativos, simbólicos, semánticos o conductuales, y los productos pueden ser: unidades, clases, relaciones, sistemas, transformaciones o implicaciones.



Teoría multifactorial: Guilford



Thurstone: Desarrolla la prueba de capacidades mentales primarias, basada en el supuesto de que cierto número de capacidades diferentes constituyen lo que caracterizamos como inteligencia. Empleando el método estadístico de análisis de factores, Thurstone, aísla cierto número de factores relativamente independientes, estos son: *Significado verbal*; Representado por una prueba de vocabulario. *Percepción espacial*; comprende dibujos geométricos ideados para valorar la capacidad individual “para visualizar los objetos en dos o tres dimensiones en el espacio”. *Números*; Se presentan dos columnas de problemas y se debe indicar si las respuestas que se presentan son correctas o erróneas. *Razonamiento*; Valora la capacidad para resolver problemas lógicos. *Fluidez verbal*; Se escriben tantas palabras como sea posible en determinado tiempo, que comiencen con una letra específica.



## Conceptos Básicos

### Calificar

Es el registro inmediato y carente de ambigüedad de la respuesta, la cual es respaldada con un argumento que nos permite formular una clave.

### Dato

Información concreta sobre hechos, elementos, etc., que permite estudiarlos, analizarlos o conocerlos, es “una medida de cualquier variable” (López, 2007, pág. 49)

### Medición

El proceso de medición en psicología va más allá de la concepción física de las características de la persona, es evidente que existen elementos intrínsecos que se requieren retomar para conocer realmente al individuo, elementos que conforman sus características psíquicas y se desglosan en ámbitos como inteligencia, actitudes y valores, por retomar las de mayor relevancia.

Estos elementos pueden ser medidos de manera más o menos objetiva con base a los conocimientos que se tengan respecto a estos mismos, la cantidad y sobre todo la calidad de la información con que se cuenta son imprescindibles para lograr la mayor precisión en el proceso de medición, de esta manera el hombre a tratado de hallar una forma de medir aspectos tales como la inteligencia, las actitudes, las habilidades, etc.

Así, en ese sentido se trata de conformar los aspectos que den respaldo a las posibles mediciones de aquellas características que interesa cuantificar, con la finalidad de establecer parámetros que permitan darle una utilidad a esa concepción numérica que se da al fenómeno de estudio en cuestión.

Se debe por principio de cuentas, definir lo que es entendible por medir; en términos generales, se puede entender el medir como “la descripción de datos en términos de números” (Guilfor, 1954 en Brown 1980, pág. 8), aunque con mayor precisión se define como la “asignación de números a objetos o eventos, de acuerdo a reglas explícitas” (Stevens, 1951 en Brown, 1980, pág. 8).

De igual forma coinciden Siegel y Castellan al definir la medición como “el proceso de mapear o asignar números a objetos u observaciones” (Siegel & Castellan, 2005, pág. 53)

Carmines y Zeller (1991, citados en Hernández, Fernández, y Baptista, 2010, pág. 276) definen a la medición como el *proceso de vincular conceptos abstractos con indicadores empíricos* a través de un plan explícito, organizado y que permite clasificar y cuantificar datos disponibles de acuerdo a indicadores

Es decir, la medición “responde a la pregunta ¿cuánto? [y] proporciona una descripción de la ejecución de la persona; [pero] no nos dice nada sobre el valor de dicha ejecución” (Brown, 1980, pág. 13).

Se puede observar que se procura otorgar un valor numérico a ciertas características que interesan en un proceso de investigación, características que no necesariamente han de ser palpables, sino que se derivan, a decir de Hernández, et al (2006), de abstracciones formuladas en el desarrollo del conocimiento, es decir, a través de la formulación de *constructos*.

## **Evaluación**

El término evaluar se deriva del latín *valere* que significa *ser fuerte*. La evaluación señala una valoración *externa* a lo que es el objeto a evaluar. El proceso de medición conlleva entonces, el poder ser interpretado y tener un sentido de bajo o alto, bueno o malo, etc. con base a las reglas previamente establecidas y parámetros de comparación, es decir, se procede a evaluar y permite la adecuada toma de decisiones.

La evaluación psicológica es la “recopilación e integración de datos relacionados con la psicología con el propósito de hacer una valoración psicológica, lograda con el uso de herramientas como pruebas, entrevistas, estudios de caso, observación conductual y aparatos y procedimientos de medición diseñados en forma especial” (Cohen & Swerdlik, 2001) (López, 2014, pág. 49)

Luego entonces, se debe entender la evaluación como el proceso de dar un juicio sobre el fenómeno que se mide. “Tratamos de responder la pregunta: ¿Hasta qué punto es bueno? Esto es evaluación” (Brown, 1980).

### **Principios de la Evaluación**

1. Finalidad: Identificar los objetivos de la evaluación y no perderlos de vista y seleccionar los instrumentos sobre la base de esos objetivos
2. Continuidad: Usar variadas y distintas técnicas de manera ordenada y calificadas sin perder de vista las limitaciones de dichas técnicas
3. Diferencialidad: La evaluación es el punto de partida para mejorar las prácticas de enseñanza, orientación y administración, por lo que se requiere la participación de diversas personas en diferentes partes del proceso para lograr una evaluación libre de sesgos.
4. Integralidad: La evaluación no debe basarse exclusivamente en evidencias objetivas, considerando que se trabaja con personas y que las diferencias individuales pueden tener diversas causas que las impulsan.
5. Cientificidad: Se evalúa para averiguar resultados, por lo que requiere de métodos claramente establecidos y validados.
6. Individualidad: Toda evaluación exige comparación de los hechos y sus evidencias, con patrones normales de algún tipo, se requiere ser consciente de las diferencias que existen entre los individuos

### **Funciones específicas:**

- a) Motivación
- b) Diagnóstico
- c) Pronóstico
- d) Orientación y asesoría
- e) Promoción
- f) Selección

**Tipos de Evaluación:**

1. Funcionales y formales: Son funcionales porque procuran averiguar por qué se actúa de cierta manera y utilizan varias técnicas para entender qué causa las conductas de estudio. Son formales porque se planifican para ello, ya que demandan atención, recursos y esfuerzos especiales en su desarrollo, debido a las implicaciones de sus resultados sobre las personas o las instituciones involucradas.
  - 1.1. Diagnóstico: Pretende evidenciar en cuál nivel, dentro de un proceso de medición, se encuentra alguna o algunas características que son objeto de estudio, para señalar lo que ocurrirá en un determinado proceso y adecuar los métodos o estrategias de trabajo que permitan definir y alcanzar un objetivo.
  - 1.2. Sumarias: Es la suma total de los resultados de todas las pruebas y medios de evaluación a fin de tomar una decisión.
  - 1.3. Evaluación formal: Llevadas de acuerdo a un plan y permite respaldar la toma de decisiones.
2. Tipos de análisis e interpretación
  - 2.1. Por normas: Son los resultados promedio obtenidos a partir de las mediciones de un número suficiente de individuos, lo cual permite comparar resultados individuales con los resultados de los otros individuos de la población.
  - 2.2. Por criterios: De acuerdo a rasgos o característica adoptada como parámetro para juzgar una característica psicológica.

**Sustento de la Ciencia**

Toda ciencia, requiere para su constitución general de tres elementos:

- Objeto
- Método
- Fin

De acuerdo con Carl Gustav Hempel la ciencia es el modo de conocimiento que aspira a formular, mediante el uso de un *lenguaje riguroso y apropiado*, teorías o leyes por medio de las cuales se rigen los fenómenos, se clasifica de dos tipos:

- Formales: Cuyos objetos son ideales y atemporales (lógica y matemática), las cuales se basan en la demostración.
- Fáticas o empíricas: Fenómenos reales: Naturales, estudian los fenómenos no producidos por el hombre (Física, Química, Astronomía, Biología) y Sociales, basada en la actividad humana y sus productos (sociología, historia, psicología, antropología, economía).

Cualquier postura científica se basa en tesis que habrá de demostrarse para eliminar su noción de falsedad, por lo que requiere de recurrir a la interdisciplinariedad que conlleva realidades objetivas, pero también la perspectiva ideológica.

En ese aspecto la ciencia se conforma, con base a las diversas disciplinas, sean en su utilidad a las ciencias exactas o las ciencias humanas, en dos tipos de relaciones:

- a) De aplicación: Implica una dualidad de objetos, en el que es aplicado difiere sobre el que se aplica, por lo que conlleva de igual manera exterioridad, es decir, la instrumentalidad del primero con respecto al segundo.
- b) De constitución: La ideología científica forma un solo bloque con la práctica científica.

Respecto a la relación entre disciplinas científicas y literarias, hay que recordar lo señalado por Kant, en el sentido de que *pueden existir ciencias cuyos objetos no existen, y pueden existir ciencias sin objeto. Las ciencias humanas son ciencias sin objeto.*

### **Concepto, Constructo, Factor, Variables, Indicadores, Índices**

La competencia argumentativa es una mezcla conceptual, no un referente.

#### **Concepto**

El concepto se puede entender como “una abstracción creada por una generalización a partir de instancias particulares” (Kerlinger, 2002, pág. 36). De acuerdo con Galimberti (2002, pág. 226), el concepto es el resultado de un proceso de abstracción, en la que se da una categorización de objetos o acontecimientos basándose en cualidades comunes y suele ser indicado por una palabra. Siendo el significado de un término, representa la unidad mínima de un pensamiento o una proposición, y su validez universal se relaciona con su comunicabilidad.

#### **Constructo**

El constructo, es la posición lógico-hipotética adoptada para prever una serie de fenómenos cuya realización no es observable, sino deducibles del constructo adoptado, es decir, se conoce la existencia pero su definición no es clara, precisa o aceptable del todo, pudiendo incluso crear controversias en su acepción. En la construcción de test, se adopta el uso de los constructos que tienen su formación en los métodos hipotéticos-deductivos.

El constructo, a decir de Kerlinger (2002, pág. 36), es un concepto con el “significado agregado de haber sido inventado o adaptado para un propósito científico especial, de forma deliberada y consciente”.

Por lo que, a decir del mismo Kerlinger (2002, pág. 36), se usan de dos formas: 1) incorporando los esquemas teóricos y relacionándolos con otros constructos, 2) definiendo el elemento de estudio de tal forma que puede ser observado y medido.

Los conceptos empleados dentro de la psicología son constructos hipotéticos, ya que se supone su existencia a través de factores que los conforman como fenómenos deducibles de la expresión conductual del sujeto. Las investigaciones sustentan o refutan la existencia de los constructos si existen explicaciones que les permitan seguir siendo factibles o no. Los estados psicológicos se infieren a través de las manifestaciones conductuales ya que no son percibidas por sí mismas como tal de modo que se crea un constructo hipotético, es decir “una alusión a un estado de interno que se supone que desempeña su función entre todos los efectos en el comportamiento humano” (Coolican, 2005, pág. 21).

Los constructos encierran la idea principal de trabajo de los investigadores derivada del sustento teórico, por tal, se inventan o adoptan de acuerdo a un propósito científico en especial, de forma deliberada y consciente, incorporando los esquemas teóricos y relacionándolo con otros constructos y definiéndola y especificándola de forma tal que pueda ser observada y medida.

Para llevar a cabo el proceso de medición y observación es necesario la conformación de dos elementos importantes que conformarán la estructura de la construcción de la investigación científica, en la definición de las características del fenómeno de estudio: la definición conceptual y la definición operacional.

### **Definición conceptual**

La definición conceptual o constitutiva es dada por los libros especializados que encierran la esencia o características de la variable, objeto o fenómeno siendo definiciones reales, esta definición es respaldada por los argumentos que presenta el autor de las mismas, así como los elementos teóricos que lo sustentan y permiten la comprensión del fenómeno.

### **Definición operacional**

La definición operacional de las variables especifica las actividades requeridas para la recolección de datos y poder llevar a cabo el proceso de medición y evaluación. Ya que “proporciona un método más o menos válido para medir cierta parte de un constructo hipotético; pocas veces cubre la totalidad de lo que se tiende normalmente por ese constructo” (Coolican, 2005, pág. 22) Establece las normas y procedimientos que seguirá el investigador para medir las variables, por lo que, en cualquier proceso de investigación, es importante plantear de la manera más precisa posible aquellos *factores* que se consideran como *indicadores* del constructo en el que se está interesado.

### **Factores o variables**

*Factor* se refiere a una *variable categórica* con un pequeño número de niveles, que se tiene como posible fuente de variación, siendo una variable categórica explicativa (Everitt y Wykes, 2001). Es decir, el factor es una variable o característica que puede ser medida en diferentes individuos, y es susceptible de adoptar diferentes valores.

Las variables deberán ser susceptibles de medirse, observarse, evaluarse e incluso inferirse a través de la obtención de datos de la realidad. (Hernández, Fernández y Baptista, 2006, p. 148)

El transformar conceptos [constructos] a índices requiere, de cuatro etapas (Padua, 1996):

- a) Imagen inicial: Es pasar de la imagen inicial a una más clara a través el respaldo bibliográfico, lo que conlleva a la segunda etapa.
- b) Dimensiones, a través de la descomposición del concepto original en las “partes” que lo conforman (Factores).
- c) Posteriormente se señala cuáles son los indicadores para cada una de estas dimensiones, a través de la elaboración de reactivos que permitan obtener la respuesta del sujeto.
- d) Con los elementos conjuntados se procede a la elaboración de índices, lo que permite la plena operación de las variables.

Las variables, por su cambio de *valor*, retomando de nuevo a Padua, incluyen el atributo (medición nominal), cuasi variable (medición ordinal) o variables como tal (medición intervalar y de razón)

**Índice**

El índice es la reconstitución de un concepto original que se ha dimensionalizado y en que en cada dimensión tiene diversos indicadores.

El índice es “un número que está compuesto de dos o más números” (Kerlinger & Lee, 2002, pág. 189), ya que este es resultado final de diversas medidas, permitiendo la simplificación y comprensión de los hechos, tomando por lo general forma de cocientes que varían de entre -1.00 y 1.00, pasando por 0.

**Formas de medición de las variables**

Directa: Unidimensional

Indirecta: Unidimensional

Variables → Indicadores

Otra variable
Productos
Razones
Proporciones
Puntuaciones

Indirecta multidimensional

Variable	Dimensión 1	Indicador 1.1
		Indicador 1.2
		Indicador 1.3
	Dimensión 2	Medición directa
	Dimensión 3	Indicador 3.1
		Indicador 3.2

**Pasos para la operacionalización de las variables**

1. Identificar la naturaleza de la variable
2. Identificar la forma de medir la variable (directa o indirecta)
3. Seleccionar indicadores (de ser necesario)
4. Elegir la escala de medición (de las variables o indicadores)
5. Elegir los instrumentos y procedimientos de medición
6. Definir la forma en que se expresaran los varoles de la variable para el análisis

**Niveles de Medición**

La medición solo es permisible solo cuando hay cierto grado de isomorfismo entre los sistemas y sus estructuras, entre el sistema numérico y el sistema y la estructura del objeto que se está midiendo. Dos sistemas son isomorfos si sus estructuras son idénticas en las relaciones y operaciones que permiten (Herrans, 2000).

Existen cuatro niveles de medición:

### **Nivel Nominal**

Es el nivel más elemental de medición, asigna a las personas a categorías cualitativamente distintas, y se pretende determinar si existen dos personas miembros de la misma categoría o clase, es decir, las categorías deben ser mutuamente excluyente. Esta nivel se caracteriza por la relación de equivalencia (los sujetos que se agrupan en una categoría específica tienen que ser iguales en relación con el rasgo que se usó) ( $=$ ,  $\neq$ ).

### **Nivel Ordinal**

Clasifican a las personas en alguna dimensión, se tiene la clasificación y la magnitud, sin tener conocimiento del tamaño de las unidades del nivel de medición. Usualmente se usa el símbolo  $>$  (Ej. Clasificación de los más altos promedios). ( $=$ ,  $\neq$ ,  $>$ ,  $<$ ).

### **Nivel Intervalar**

La diferencia de magnitud significa lo mismo en todos los puntos de la Nivel. Implica clasificación, magnitud y unidades de tamaños iguales o estándar y uso un punto cero arbitrario. Su uso en la psicología es importante, ya que los puntajes se pueden transformar en cualquier clase de calificación lineal; sumando o restando una constante o multiplicándolas o dividiéndolas por la misma constante. En este sentido, las calificaciones de una Nivel se pueden convertir a las de otra que utilice unidades diferentes. Las estadísticas más utilizadas presuponen una Nivel de intervalos de medición. ( $=$ ,  $\neq$ ,  $>$ ,  $<$ ,  $+$ ,  $-$ ,  $X$ ,  $\div$ ).

### **Nivel de Razón o Proporción**

Posee todas las propiedades de la escala de intervalos y un punto cero real o verdadero en su origen, llega al nivel de medición más alto y preciso. La razón entre dos puntos cualesquiera es independiente de la unidad de medición. Se logra crear una escala de razón cuando cuatro relaciones son factibles operacionalmente: la relación de equivalencia, la relación de más que, la razón entre dos intervalos cualesquiera y la razón entre dos valores cualesquiera de la escala. ( $=$ ,  $\neq$ ,  $>$ ,  $<$ ,  $+$ ,  $-$ ,  $X$ ,  $\div$ ,  $0$ ).

### **Características del Examinador o Evaluador**

El proceso de aplicación de cualquier tipo de prueba implica el adecuado conocimiento y entrenamiento de quien lleve a cabo este proceso, en este sentido, los elementos de mayor importancia a tomar en consideración son los siguientes:

- a. Conocimiento técnico actualizado de la prueba y su forma de administración.
- b. Capacidades y aptitudes del evaluado.
- c. Conocimiento del contexto cultural del evaluado.

- d. Discriminación (cultura, nivel socioeconómico, educación, edad, género y origen étnico)
- e. Neutralidad en el proceso (Reforzamiento de respuesta y Expectativa).
- f. Rapport y Empatía: generar confianza a los evaluados y crear credibilidad, ser cooperativo al resolver dudas, expresarse de un modo no amenazador, la paciencia y flexibilidad, Ser asertivos.
- g. Ética: objetividad e imparcialidad.

### **Características de la Población**

El objetivo de cualquier estudio científico radica en hacer generalizaciones a partir de determinadas muestras, por lo que la población la constituye todos los miembros existentes de un grupo que tienen ciertas características relevantes a la investigación, en ese sentido son conocidos como casos. Por lo que la elección de un determinado número de casos dentro de una población se denomina muestra. El determinar muestras representativas permite generalizar los resultados, con cierto margen de error, hacia la población objeto de estudio.

La *población* también es conocida como *universo de estudio*, *población meta* o *marco de muestreo*, este último permite identificar mejor a aquellos individuos que cumplan las características requeridas para delimitar la población, enumerarlos y seleccionarlos en caso de requerir una muestra.

La *muestra* es un subgrupo representativo de la población meta, es decir, aquella sobre la cual se recolectarán los datos y que a su vez debe definirse anticipadamente con precisión.

Existen varias formas de selección para obtener una muestra representativa, de las que destacan:

- a. Selección aleatoria o probabilística, en la que se da la selección completamente al azar, considerando una lista de individuos como guía y los cuales serán seleccionados de acuerdo a números aleatorios.
- b. Selección casual o no probabilística, en el que se elige a quien esté disponible o el primero de una lista.

De las características de este marco de muestreo, se derivan elementos o variables extrañas que pueden afectar la calidad de la medición, entre los elementos intrínsecos que afectarán a los evaluados se encuentra:

- a) Ansiedad y el grado de ésta.
- b) Capacidad y disposición para cooperar.
- c) Dolor, molestias e incomodidad física propias del sujeto.
- d) Nivel de alerta o somnolencia.
- f) Predisposición al acuerdo o desacuerdo hacia los reactivos.

- g) Preparación previa a la prueba.
- h) Importancia que atribuya a la autodescripción positiva o negativa.
- i) La suerte del evaluado al responder sin conocer de lo que responde.

### **Contextos de evaluación**

El contextualizar “articula una realidad a otra más compleja o tratar de ajustarla a los hechos que median entre ellas para que cobre sentido y se inserte en el proceso social” (Reimers, 2018). El contexto media, altera y determina las circunstancias de la realidad a la que se circunscribe, las asimila y cambia en un permanente juego interactivo. Todo está entonces sujeto a ciertos valores y creencias inscritos en el contexto y en ese sentido la sociedad se desenvuelve y los docentes obramos inconsciente o concientemente de acuerdo a lo que él prescribe.

El contexto social de la evaluación refiere al conjunto de instituciones, prácticas e interacciones entre grupos interesados en la evaluación que resultan en la regulación de qué se estudia, de qué manera, por quién, con conocimiento de quién y con cuáles resultados (Cabral, 2008).

El contexto social de la evaluación define, pues, la cultura de la evaluación que resulta de normas, instituciones y transacciones políticas entre actores sociales e institucionales.

La educación constituye una práctica política, sus procesos de evaluación también lo son, por lo que, el contexto no deja de prescribir o trazar una perspectiva que resulta útil a la hora de desarrollar propuestas que contribuyan a evaluar.

### **Alcance de la evaluación**

#### **1. Establecimiento de los objetivos**

*El establecimiento de los objetivos de evaluación, da pauta a formular diversos tipos de estrategias de evaluación, a su vez, estas requieren identificar las preguntas claves necesarias para responder de acuerdo al tipo de evaluación correspondiente (fig. 1).*

##### **1.1 Valoración de necesidades**

Se debe conocer cuáles son las problemáticas o necesidades a indagar, así como sustentarlas de acuerdo a datos que las demuestren y delimiten para alcanzar el objetivo que de ellas surja.

##### **1.2 Seguimiento y control de programas**

Consiste en técnicas de evaluación, para medir algún aspecto del desempeño del programa, teniendo dos propósitos principales:

- *Mejorar los programas* identificando los aspectos que funcionan de acuerdo al plan y aquellos que requieren corrección a mitad de un proceso, curso, actividad, etc.
- *Rastrear los cambios en los resultados obtenidos* en comparación de los resultados deseados.

El seguimiento incluye medir la situación vigente y el cambio a través del tiempo en cualquier componente del programa y en ciertos aspectos de la población. Permite evaluar aspectos conexos y derivados del trabajo de evaluación desarrollado:

**A nivel del programa:**

- Insumos: suministros, personal administrativo y fondos para la implementación
- Productos: capacitación, acceso, calidad y utilización.

**A nivel de la población:**

- Productos: (por ejemplo: tasas de deserción escolar, estrés en madres, rendimiento laboral, etc.)

Idealmente, el seguimiento y control de programas incluye técnicas de investigación cuantitativas y cualitativas. La recolección de datos puede incluir cualquiera de las técnicas estándar usadas en la investigación de las ciencias sociales (encuestas, grupos focales, entrevistas en profundidad, observación, informantes clave, etc.) además del análisis de datos del programa (por ejemplo, estadísticas de servicios). Las evaluaciones pueden ser completas, cubriendo una gran variedad de temas o concentrarse en un solo tema, como, por ejemplo, la calidad de los servicios institucionales.

**1.3 Valoración del impacto**

El propósito es medir el grado de cambio atribuible al programa o intervención. A diferencia del seguimiento y control de programas, que rastrea el cambio, la valoración del impacto aborda la causalidad.

La evaluación del seguimiento y control de programas, pueden incluir las mismas variables para el seguimiento de programas. La diferencia entre ambos es el diseño de la evaluación y las técnicas de análisis utilizadas.

**2. Esquema del alcance**

Se presenta en forma escrita lo que se va a evaluar. La descripción del alcance de la evaluación incluye:

- Metas y objetivos del programa
- Marco conceptual que muestre los enlaces entre insumos, procesos, productos y resultados
- Objetivos de la evaluación

Tipo	Pregunta(s) Asociada
<b>Valoración de necesidades</b>	¿Qué debe incluir el programa y cuál es la mejor forma de organización del servicio para satisfacer las necesidades del grupo objetivo?
<b>Seguimiento y control de programas</b>	<p><b>Insumos:</b> ¿Estuvieron disponibles los insumos (por ejemplo, equipo, suministros, medicinas, materiales, personal) para el programa, en las cantidades y tiempos especificados por el programa?</p> <p><b>Procesos:</b> ¿Se llevaron a cabo las actividades programadas conforme al plan? ¿En qué medida fueron llevadas a cabo apropiadamente?</p> <p><b>Productos:</b> ¿Ocurrieron los cambios esperados al nivel de programa, en términos de acceso a los servicios, calidad de la atención, utilización del</p>

	servicio? <b>Resultados:</b> ¿Ocurrió el cambio esperado al nivel de la población (no atribuible al programa)? <b>Costos:</b> ¿Cuál fue el costo extra de expandir una actividad, produciendo una unidad mayor de producto y logrando el cambio ocurrido?
<b>Valoración del impacto</b>	¿Cuál y cuánto cambio ocurrió (al nivel de programa o de población) que sea atribuible al programa?

Fig. 1

### Ética en la evaluación

Cualquier tipo de valoración en psicología requiere, de quienes en ello trabajan, del conocimiento íntegro de las implicaciones éticas de la actividad de evaluación o valoración. Es necesario entender el alcance de la valoración en psicología y no anteponerlas sobre el respeto que merece el individuo, desde este plano es importante observar los elementos que permiten proteger al individuo en el proceso de valoración psicológica.

#### Derechos Humanos:

Son el principal elemento que sirve para normar el quehacer del trabajo en psicología, de entre lo más relevante se encuentra:

- Derecho a no ser evaluado: A excepción de ser ordenadas por la ley; al existir consentimiento implícito en actividades educativas, institucionales u organizacionales; o al evaluar capacidades de decisión.
- Derecho a conocer las calificaciones de la prueba y su interpretación, y las bases de cualquier decisión que afecte a sus vidas.
- Derecho a conocer quien tendrá acceso a la información de la prueba.
- Derecho a la confidencialidad de los resultados de la prueba

#### Consideraciones en el proceso de valoración:

##### Aspectos individuales:

La valoración en sí, conlleva varias características a observar y de las cuales es importante considerar a efecto de no perturbar el bienestar del individuo.

- Etiquetar o clasificar a las personas: En el proceso de identificación diagnóstica de una enfermedad no es malo al tratar de entender y atender algún problema psicológico, lo delicado es caer en *etiquetamientos*, que limiten o alteren la vida de la persona, es decir se puede incurrir en la *estigmatización*.
- Invasión a la privacidad: Se da al no emplearse la información de manera adecuada, no se debe solicitar ni revelar ninguna información adicional más allá de lo necesario, para el propósito de la prueba. Las personas deben conocer los límites de la confidencialidad y que datos de una prueba pueden ser requeridos por mandato judicial.

- Lealtades divididas: El conflicto se da cuando el bienestar del individuo está en contra del bienestar de la institución que requiere de la evaluación. Es obligación del psicólogo señalar su posicionamiento en cuanto a lealtad. Señalando a los sujetos como se utilizarán las pruebas y describir los límites de confidencialidad Y a las instituciones, se les proporcionará el mínimo de información necesaria.
- Responsabilidad del usuario y de los desarrolladores de pruebas: Se deben utilizar instrumentos cuya validez y confiabilidad haya sido establecido para el uso en miembros de la población que está siendo evaluada, al mismo tiempo, se debe considerar que las diferencias situacionales, personales, lingüísticas y culturales afectan el juicio o reducen la precisión en las interpretaciones

#### **Aspectos sociales:**

- Deshumanización: El manejo de equipos computarizados a permitido el rápido análisis de la información que se genera, sin embargo, carece de elementos de juicio que permita valorar desde una perspectiva más flexible el por qué de los resultados de una prueba o varias pruebas, en un grupo de sujetos. Esto es importante al hacer una valoración clara y precisa de los resultados obtenidos tras la aplicación de una prueba.
- Utilidad de las pruebas: La aplicación de pruebas, debe tener una justificación clara y precisa, no debe permitirse la aplicación de pruebas sin una razón de ser de la misma. Cualquier instrumento de medición debe salvaguardarse de una utilización sin sentido, ya que este acto por si mismo puede provocar contaminación en la población y afectar los resultados en futuras aplicaciones que lleven un verdadero interés científico.
- Acceso a servicio de pruebas psicológicas. El hacer cualquier tipo de valoración implica un costo, los servicios de bienestar social, salud pública y educación en nuestro país prácticamente, no consideran la aplicación de pruebas para determinar la magnitud de cualquier problemática desde la perspectiva psicológica, lo cual nos deja en gran desventaja ante otras naciones que analizan y entienden que ocurre con la población y definen y afinan estrategias de trabajo de acuerdo a los datos encontrados.

#### **Código Ético del Psicólogo**

- a) La aplicación e interpretación de las pruebas e instrumentos psicológicos son exclusivos de quienes posean la preparación profesional adecuada y hayan aceptado las obligaciones y consecuencias de esta práctica.
- b) Seguir los procedimientos científicos al construir, desarrollar, adaptar, administrar o usar cualquier tipo de técnica o instrumento, verificar su base en datos sólidos de validez, confiabilidad y normatividad, así como su uso apropiado
- c) La toma de decisiones, reconoce los límites de la certidumbre con la que es posible diagnosticar, emitir juicios, o hacer predicciones acerca de individuos o grupos.
- d) Identifica situaciones en las que ciertas técnicas de valoración o normas no son aplicables o requieren de ajustes para su administración o interpretación (cuestiones de género, la edad, la raza, el origen étnico, nacionalidad, orientación sexual, discapacidad, idioma, o nivel socioeconómico)
- e) Evitar la indiscriminada comercialización y distribución de pruebas para uso profesional, incluyendo manuales o información que expresen sus motivos o fines,

su desarrollo, su validez, y el nivel de entrenamiento necesario para aplicarlas e interpretarlas.

- f) Emplear instrumentos como se indica en los manuales respectivos, respetar el rigor metodológico en la aplicación e interpretación de los mismos.
- g) Usar las pruebas e instrumentos en proceso de validación sólo con fines de investigación o docencia, previa aclaración al respecto y con las debidas reservas.
- h) Considerar a las pruebas psicológicas como instrumentos auxiliares que de ninguna manera son suficientes para elaborar un diagnóstico.
- i) Explicar al consultante sobre la naturaleza, propósitos y resultados de la prueba en lenguaje comprensible y constructivo, salvaguardándolo de cualquier situación que ponga en peligro su estabilidad emocional.
- j) Dar a conocer a los consultantes los resultados e interpretaciones de los instrumentos psicológicos empleados, evitando aportar información que pueda comprometer el funcionamiento de la prueba, pero explicando las bases de las decisiones que puedan afectar al consultante o a quien dependa de él.
- k) Evitar aplicar cualquier instrumento de evaluación psicológica a familiares o amigos.
- l) Asegurarse de que la aplicación y resultados de instrumentos de evaluación psicológica sean estrictamente confidenciales. [\(volver a PP\)](#)

### **Consentimiento Informado**

Es probable que éste sea el requisito más importante, y el formato de consentimiento informado podría ser la herramienta que por sí sola asegure un comportamiento ético. Sin duda, todo proyecto de investigación que utilice participantes humanos debe incluir un formato de consentimiento informado el cual leerá y firmará cada participante o la persona que autorice la participación (en el caso de un niño, el padre firmaría el formato).

El consentimiento informado es una descripción de lo que sucederá durante el curso de la investigación. Una carta de este tipo debe contener al menos la siguiente información:

- El propósito de la investigación
- La identidad y la filiación del investigador (o los investigadores)
- Lo que está haciendo el investigador (o los investigadores)
- Cuánto tiempo durará la participación
- El derecho del sujeto a retirarse del estudio en cualquier momento y por cualquier razón.
- Los posibles beneficios para el individuo y para la sociedad
- Los posibles daños, riesgos o molestias para el individuo
- Un compromiso de mantener la confidencialidad estricta de los resultados
- Cómo obtener una copia de los resultados
- Cómo ponerse en contacto con el investigador (o los investigadores) en caso de tener dudas

Un lugar para la firma del sujeto que indique el estar de acuerdo en participar y que entienda el propósito de la investigación.

La carta se presenta en papel membretado e ilustra todos los puntos anteriores. No está escrita en jerga científica, y es tan sencilla y clara como es posible. El objetivo aquí es informar, no coaccionar ni persuadir a las personas para que participen.

### **Consentimiento informado en el caso de niños**

Existe un problema obvio cuando se trata de obtener el consentimiento informado en el caso de niños. Un ejemplo es una investigación del desarrollo visual de niños pequeños, en la que el niño es demasiado joven para dar cualquier tipo de consentimiento. Queda a juicio de los padres decidir si permitirán o no que su hijo participe.

Sin embargo, la ética de la investigación con niños implica una gran cantidad de cuestiones que van mucho más allá del difícil proceso de asegurar que los niños no se expondrán a ningún peligro, sea físico o psicológico. Por ejemplo, ¿los niños de 12 años tienen edad suficiente para tomar la decisión de retirarse del estudio, opción que debe quedar bien clara en el formato de consentimiento? ¿Pueden ellos entender las implicaciones a largo plazo de la investigación en la que están participando? ¿Y los posibles riesgos? Es aquí dónde debe entrar en juego el buen juicio del investigador. Si un niño está convencido de que no desea participar, tal vez usted pierda ese sujeto y esos datos, pero es preciso respetar sus deseos tanto como los de un adulto. Además, una participación forzada podría causar el descontento o enojo del niño y dar lugar a datos poco confiables.

A medida que aumenta la edad de los niños, el problema se vuelve más complejo. Por ejemplo ¿qué hay con el niño de 12 años que tiene edad suficiente para entender el propósito del experimento? ¿Debe él o ella firmar el formato de consentimiento además del padre, la madre o ambos? Ningún investigador cuerdo se abstendría de obtener primero la autorización de los padres o de alguno de ellos. Además, cuando se utilizan niños de edad escolar en investigaciones, las autoridades de educación podrían exigir que un comité de investigación de la escuela revise la propuesta. Por añadidura, cada vez más investigadores en Estados Unidos, por ejemplo, sienten la necesidad de contar con un seguro de responsabilidad para protegerse en caso de que algún padre enojado promueva una demanda o que ocurra un perjuicio no intencional.

### **Pasos para elaborar un instrumento**

#### **Planteamiento del problema:**

1. Relevancia social  
Cuáles son las circunstancias sociales, clínicas, educativas, organizacionales o de cualquier índole que implique la intervención del psicólogo para identificarlas y que permita establecer criterios para su medición
2. Estudios relacionados  
Conocer de qué forma ha sido abordado el tema o temas similares por otros investigadores para conocer los enfoques metodológicos en los que se han basado

para llevar a cabo su investigación, y considerar aquellos que conlleven procesos cuantitativos y comprender sus metodologías y adaptarlas o contemplarlas al establecer criterios propios de medición.

### 3. Enfoques teóricos

Es evidente la importancia de entender los diversos enfoques teóricos que se establecen y que permiten el manejo de posturas cuantificables para el desarrollo de la investigación científica en psicología, ya que de esta forma se permite delimitar un trabajo de medición y de igual forma, se establecen los criterios que se han de considerar relevantes para conformar o configurar el perfil de aquel fenómeno psicológico que nos interese investigar.

## **Determinar las características del fenómeno de estudio**

- 1) Definir el objetivo de medición
- 2) Definición conceptual y definición operacional  
Indispensable considerar para la elaboración de cualquier instrumento de medición en psicología (leer *Concepto, Constructo, Factor, Variables, Indicadores, Índices*, en esta antología).
- 3) Determinar los factores o variables a medir  
Cualquier fenómeno psicológico se conforma de características que son sustentadas a través de las diversas teorías que lo conforman, es necesaria una lectura amplia de las investigaciones que las sustentan.
- 4) Reactivos
  - 4.1) Estímulos (afirmaciones o preguntas)
  - 4.2) Tipos de respuesta
- 5) Primera validación del instrumento
  - 5.1) Validez de contenido
  - 5.2) Por *sujetos tipo*
  - 5.3) Determinar el peso de los reactivos (puntajes de los reactivos)
  - 5.4) Asignación de puntajes totales
  - 5.5) Análisis de los reactivos
  - 5.6) Versión final de la escala

## **Estructuración del instrumento y su aplicación**

- 1) Instrucciones para el evaluador
  - 1.1) Consideraciones previas a la aplicación
  - 1.2) Consideraciones durante la aplicación
  - 1.3) Consideraciones posteriores a la aplicación
- 2) Instrucciones para el evaluado
- 3) Datos de identificación
- 4) Cuaderno de reactivos
- 5) Protocolo de respuesta

## **Procesamiento de la información y análisis de resultados**

- 1) Bases de datos
- 2) Alfa de Cronbach
- 3) Estadística descriptiva

- 4) Validez
- 5) Confiabilidad
- 6) Estandarización y normalización

### **Conclusiones y sugerencias**

Consideraciones en relación a la primera aplicación en la que se habrá de observar los cambios necesarios para el ajuste y corrección del instrumento creado.

### **Adaptación de los test**

(Mikulic I. M., s.f.)

### **Métodos de Adaptación de los Test**

La adaptación implica considerar no sólo las palabras utilizadas al traducir la prueba sino también las variables culturales involucradas. Así una correcta traducción no asegura un significado unívoco. Van de Vijver y Leung (1997) establecieron tres niveles de adaptación de las pruebas psicológicas.

#### 1. Aplicación

La simple y llana traducción de un test de un idioma a otro, asume la equivalencia de constructo. Es el método más común y utilizado en el mundo. La traducción de una prueba no indica ningún nivel de equivalencia entre ambas versiones de la misma.

#### 2. Adaptación

A la traducción se agrega la transformación, adición o substracción de algunos ítems de la escala original para evitar errores por su significado a través de las culturas y, por lo tanto, necesitan modificaciones o ser eliminados. Así mismo, los ítems que no existen en la versión original del test pueden representar mejor al constructo en la población en la cual se administrará la nueva versión.

#### 3. Ensamble

El instrumento original ha sido modificado tan profundamente que prácticamente se ha transformado en un nuevo instrumento original con los nuevos elementos. Ocurre cuando muchos ítems del test original son inadecuados para representar el constructo a medir. El ensamble también se da cuando el constructo no está representado de forma adecuada por la versión original en la cultura a la que se quiere adaptar la prueba.

### **Técnicas de Traducción**

El proceso de traducción es complejo e implica más que la traducción lineal de las palabras escritas a un nuevo lenguaje. Existen dos métodos comunes: la traducción directa o *forward translation* y la traducción inversa o *backward translation*. En el método de traducción directa un traductor, o preferentemente, un grupo de traductores, traducen el test desde el idioma original al nuevo idioma. Luego, otro grupo de

traductores, juzga la equivalencia entre las dos versiones. De este modo pueden realizarse las correcciones pertinentes en las dificultades o errores identificados por los traductores. En el caso de la traducción inversa, el más utilizado de los métodos, un grupo de traductores realiza una traducción desde el idioma original al nuevo idioma; luego un segundo grupo de traductores toma el test traducido (en el nuevo idioma) y vuelve a traducirlo al idioma original. Seguidamente se realizan las comparaciones entre la versión original y la versión retraducida al idioma original para determinar su equivalencia. Ambos métodos poseen diversas ventajas y desventajas.

### **Métodos para establecer la equivalencia entre Test**

Una vez que se ha traducido convenientemente una prueba es necesario realizar un estudio para establecer si la traducción es equivalente a la original. Para ello habrá de implementarse un diseño experimental y un análisis de datos obtenidos a través de ese diseño. Hambleton (2000) señala que existen básicamente 3 métodos:

- 1) Administración de la prueba en el idioma original y de la prueba traducida a sujetos bilingües. Se administrará ambas versiones de la prueba (la original y su traducción al nuevo idioma) a sujetos que hablan ambos idiomas. Este método, posee ventajas (controlarse las diferencias de las características de los participantes en el test ya que ambas pruebas son administradas a la misma persona), y desventajas (el diseño está basado en la premisa de que los sujetos son igualmente competentes en ambos idiomas, lo cual es difícil de sostener y el hecho de que no puede asegurarse que los bilingües posean el mismo nivel de competencia que la población general).
- 2) Administración de la versión original y su traducción inversa a monolingües en el idioma original. La equivalencia de los ítems se determina comparando el desempeño de cada sujeto en cada ítem de ambas versiones. La ventaja está en el control de las diferencias en las características de los participantes. La primera desventaja está en que este diseño no permite obtener datos con la versión en el idioma meta (*target*) del test. De esta manera, no es posible obtener puntajes de sujetos que hablen el idioma al que se intenta traducir el test. La segunda desventaja reside en el hecho de la falta de independencia entre los puntajes obtenidos, al existir un efecto de aprendizaje luego de la administración de la primera versión de la prueba, especialmente si la primera es la original. La administración al azar de una de las versiones en primer lugar puede reducir la importancia del efecto de aprendizaje.
- 3) Administración de la versión original a monolingües que hablan el idioma original y de la versión traducida a monolingües que hablan el idioma al que ha sido traducida la prueba. Una dificultad reside en asumir que los sujetos de ambas muestras poseen una habilidad comparable. Sin embargo, tal obstáculo puede

superarse si los análisis son desarrollados con la Teoría de Respuesta al Ítem, en la que se asume que utilizando distintos conjuntos de ítems pueden obtenerse las mismas estimaciones de aptitud. Igualmente, administrando esos ítem a distintas muestras de examinados las estimaciones de parámetros obtenidas serán iguales. Una vez obtenidos los datos por medio de los diseños revisados existen varias posibilidades estadísticas para su análisis. Básicamente el análisis estará destinado a identificar la existencia de Funcionamiento Diferencial de Ítems (FDI) es decir, ítem que se comportan en forma diferente a través de las diversas muestras transculturales.

### **Fuentes de Sesgo**

Si deseamos usar las pruebas para predecir resultados en alguna situación futura, necesitamos instrumentos con alta validez predictiva del criterio particular. Este requisito suele descuidarse en el desarrollo de las llamadas *pruebas libres de influencia cultural*. En un esfuerzo por incluir en esas pruebas solo las funciones comunes a diferentes culturas o subculturas, puede elegirse un contenido que tenga poca relevancia para el criterio que se pretenda predecir. Una mejor solución es elegir un contenido relevante para el criterio e investigar luego las posibles diferencias poblacionales de la efectividad de la prueba para el propósito pretendido.

Desde mediados de la década de los setenta se ha observado una rápida acumulación de investigaciones sobre problemas de sesgo de la prueba. En este contexto, el término “sesgo” se emplea en su bien establecido sentido estadístico, para designar un error constante o sistemático en contraste con uno que se debe al azar. Las principales preguntas que se han planteado con respecto al sesgo de la prueba, tiene que ver con el coeficiente de validez (sesgo de la pendiente) y la relación entre las medias del grupo en la prueba y en el criterio (sesgo de intersección). Si una prueba produce un coeficiente de validez significativamente diferente en dos grupos, la diferencia se describe como sesgo de la pendiente y esta clase de diferencia entre grupos se conoce como *validez diferencial*. Una prueba exhibe sesgo de intersección si sistemáticamente subpredice o sobrepredice una ejecución del criterio para un grupo particular.

El problema del sesgo de la intersección se relaciona mas con lo que ha sido llamado *equidad de la prueba*. Aunque los términos “equidad” y “sesgo” de la prueba a veces se usan indistintamente para cubrir todos los aspectos del uso del instrumento con minorías culturales. Modelos de decisión para el uso justo de las pruebas: Gradualmente empezó a cambiar el interés de la investigación en la evaluación del sesgo de las pruebas al diseño de estrategias de selección para su uso justo con minorías culturales. Entre las metas por reconciliar están las de proporcionar iguales oportunidades a todos los individuos, elevar al máximo la tasa del éxito y la productividad, incrementar la mezcla demográfica y la representatividad y extender el

tratamiento preferencial a grupos desfavorecidos por inequidades anteriores. Van de Vijver y Tanzer (1997) identificaron diferentes fuentes de sesgo, que a continuación se explicitan:

**a) Sesgo de constructo:**

Se da cuando el constructo medido no es idéntico a través de los grupos culturales. La importancia que cada cultura otorga a ciertas conductas se encuentra en esta categoría. Conductas de ética y civismo que en algunas sociedades pueden ser normales en otras pueden constituir un verdadero rasgo de rigidez y asemejarse a una conducta obsesiva-compulsiva.

**b) Sesgo metodológico:**

Este sesgo reconoce tres formas.

a) El sesgo de muestra

Se da cuando las muestras son incomparables entre sí. La cantidad de años de escolaridad que poseen los sujetos de una muestra es una variable determinante en el desempeño del mismo en un test determinado, especialmente si se trata de un test de habilidad. El nivel sociocultural, la motivación, la composición por género y edad de los sujetos son otras variables que pueden hacer incomparables a dos muestras que pueden mostrar resultados muy diferentes en un test determinado. El sesgo en el instrumento: que puede provenir de las características del instrumento. La familiaridad que los sujetos tienen con los estímulos presentados tiene una gran importancia. Algunos estímulos tales como objetos, dibujos, figuras u otros elementos utilizados en algunas culturas no existen en otras o son irrelevantes. El idioma es otra fuente de sesgo de instrumento. La traducción de un idioma a otro frecuentemente subestimada, es un problema importante que requiere una metodología específica a seguir. Los problemas son mayores cuantas más diferencias hay entre idiomas como entre el inglés y el chino o el árabe, idiomas cuya lectura se realiza de izquierda a derecha a diferencia del de derecha a izquierda. También la disposición del texto tiene importancia en el completamiento de frases o de interpretación de textos, mayor es el problema cuando implican conectar letras y números siguiendo un orden alfabético o numérico y los caracteres de idiomas como el español, ruso, griego, árabe, hebreo o chino son tan diferentes. También entre los idiomas occidentales existen diferencias como que en inglés no existe la “ñ” que el alfabeto sueco contiene más vocales, y en portugués existen distintos tipos de “a”. Los métodos de respuesta constituyen otra fuente de sesgo del instrumento.

b) Sesgo de administración

Incluye problemas tales como dificultades en la comunicación, es decir, dificultades para que el entrevistado entienda las instrucciones del entrevistador

ya sea por la dificultad de las palabras utilizadas, el modo de explicación de las instrucciones o un inadecuado manejo del idioma de alguna de las partes. También se incluyen las alteraciones en la manera de administrar las pruebas. Normalmente los manuales incluyen instrucciones de administración que en muchos casos no son adecuadas para la población a aplicar. Los administradores del Test entonces adaptan esas instrucciones según su criterio personal. Otro punto importante es el uso de cronómetros que produce serias alteraciones en los resultados.

c) Sesgo de ítem

Se produce cuando el mismo tiene diferentes significados en distintas culturas. Ciertos grupos culturales pueden obtener puntajes significativamente distintos en un ítem determinado a pesar de obtener un puntaje total similar. La deseabilidad social o la relevancia cultural, entre otros factores, pueden producir el sesgo de ítem. Tanzer (1995) demostró que aunque la estructura factorial de un Test de autoconcepto académico era similar cuando se lo aplicó a estudiantes australianos y singaporeanos, existían diferencias sustanciales entre estas muestras cuando se compararon algunos ítems específicamente. Este tipo de sesgos también actúa en test neuropsicológicos.

## **Naturaleza y uso de las pruebas psicológicas (Instrumentos y Escalas de Medición)**

### **Encuesta**

Existe una confusión entre lo que es un *cuestionario* y una *encuesta* y se llegan a tomar en ocasiones como términos sinónimos, cuando no es así. Mientras que el término encuesta viene del francés *enquête* (investigar) o que a su vez se deriva de *inquiere* (interrogar o investigar). García (2004, pág. 29), hace una clara diferenciación entre ambos y señala que “la encuesta es una actividad consciente y planeada para indagar y obtener datos sobre hechos, conocimientos, opiniones, juicios y motivaciones”, lo que a decir del autor, conlleva el diseñar las *estrategias*<sup>1</sup> para obtener la información en la población objeto de estudio. En este sentido Coolican (2005, pág. 110), lo reafirma al señalar que “consiste en pedir información a un número relativamente grande de personas”. Señala además, que es factible también aplicarlo a pocas personas, dando como argumento que sustenta el uso de un *cuestionario* estructurado en que las respuestas a los reactivos “forman una unidad equivalente dentro de una amplia muestra” (pág. 110).

### **Cuestionario**

Proviene del latín *questionarius* (lista de preguntas). Se puede definir como un “sistema de preguntas racionales, ordenadas de forma coherente... expresadas en un lenguaje sencillo y comprensible, que generalmente responde por escrito la persona interrogada, sin que sea necesaria la intervención del encuestador” (García, 2004, pág. 29). Aunque Hernández, Fernández, y Baptista (2010, pág. 217), simplemente lo definen como “un conjunto de preguntas respecto a una o más variables a medir”. Los cuestionarios, en su

---

<sup>1</sup> La estrategia conlleva una serie de acciones planificadas hacia un objetivo.

forma más elemental, son los instrumentos escritos que se utilizan para la obtención de información, los cuales pueden ser con reactivos cerrados, abiertos o mixtos y, regularmente recopilan datos concretos tales como la edad, ocupación, número de hijos, programa favorito, tiempo que dedica a hacer deporte y más, a decir de Coolican (1997, pág. 164), este tipo de instrumentos sólo cubren lo que denomina *validez de aspecto*. Siendo dirigida directamente a la persona que posee la información deberá de poseer un lenguaje sencillo y comprensible al mismo, buscan la “introspección del sujeto y se logra uniformidad en la cantidad y calidad de la información” (García, 2004, pág. 29)

Es decir, mientras con el cuestionario, se habla del instrumento que contiene los reactivos del tema que interesa conocer, la encuesta señala el procedimiento a utilizar para obtener la información, desde la forma de acercarse a la población, hasta las instrucciones, el seguimiento y el cierre que se tendrá en dicho procedimiento.

Los cuestionarios, en sus formas más sencilla no llegan a profundizar más allá de las opiniones<sup>2</sup> de quienes los responden, mientras que en sus formas más sofisticadas, conlleva estructurar reactivos más especializados que permiten resaltar aspectos relacionados a constructos teóricos que tratan de poner en evidencia el fenómeno a medir, de entre estos destacan las siguientes denominaciones:

### **Test**

Es una palabra de origen inglés, derivada del latín *testimonium*, y hace referencia a la prueba. Se adopta en la psicología a nivel internacional para designar a las pruebas que proponen examinar las cualidades psíquicas de los individuos. Procura obtener una “medición objetiva y estandarizada que permita analizar las diferencias entre las reacciones psíquicas de varios individuos o [...] las del mismo individuo en diferentes momentos o condiciones” (Galimberti, 2002, pág. 1054). Un test es un instrumento o técnica de medición utilizada para cuantificar el comportamiento o ayudar a intentar a predecir la conducta, a través de una situación experimental estandarizada que estimula un comportamiento, con la que se mide una muestra representativa de la conducta del individuo evaluándolo a través de una comparación estadística en relación a otros individuos colocados en la misma situación, permitiendo clasificar al sujeto evaluado cuantitativamente, y buscando de esta manera ser objetiva y estandarizada. La conducta manifiesta es la actividad observable de un individuo. Algunos *tests* intentan medir el grado en que una persona podría mostrar o emitir alguna conducta manifiesta en particular y otras pruebas miden el grado en que se presentó en el pasado una conducta manifiesta.

### **Examen**

Se entiende como el proceso en sí de llevar a cabo la “indagación evaluativa para comprender a un individuo de acuerdo con los cuadros de referencia previstos para ello” (Galimberti, 2002, pág. 466). Diferenciándose en: Clínico, Psíquico y Escolar o Profesional. Evidentemente, implica la utilización de diversas técnicas (entrevista,

---

<sup>2</sup> La *opinión*, a decir de Galimberti (2002, pág. 766), es “el conocimiento o creencia que no incluye ninguna garantía de verdad”, es decir “son atendibles pero apodícticamente incontrolables”. Lo *apodíctico*, hace referencia a aquello que expresa o contiene una verdad concluyente sin lugar a duda o discusión.

observación) o instrumentos (tests) que permitan tener una evidencia tangible de la características a estudiar en el individuo.

### **Inventario**

Señala Dorsch (1981, pág. 518) que es el “conjunto de datos sobre el individuo examinado, considerando sus funciones psíquicas en general o un aspecto determinado, sean sus conductas o inteligencia”, y tiene una acepción general al comprender a los tests, cuestionarios, listas de tareas propuestas, de preguntas, e incluso a un conjunto de pruebas.

### **Escalas de Actitudes**

Una escala es la “base objetiva para la medición, confrontación y la clasificación que permite transformar un sistema empírico en un sistema numérico” (Galimberti, 2002, pág. 416) El establecimiento de una escala se da cuando:

A partir de una distribución de frecuencias multivariada de un universo de atributos, es posible derivar de una variable cuantitativa con la cual caracterizar los objetos de un modo tal que cada atributo sea una función simple de aquella variable cuantitativa (Stoufer, 1966, citado por Padua, 1996, pág. 160),

En el caso de la psicología, se habla de la medición de las actitudes, como parte del proceso de entender la perspectiva del individuo con relación al entorno (fenómeno) que interesa conocer al investigador. En este mismo sentido Pádua (2004, pág. 161) señala que las actitudes se deben vislumbrar como *síntomas*, que en su conjunto determinarán la perspectiva peculiar del individuo. Señala como conceptos relacionados con las actitudes: Creencia, Sesgo (bias), Doctrina, Fe, Ideología, Valor y Opinión

Son de alguna forma, tipos de cuestionarios que principalmente, “se centran en los pensamientos, percepciones, sentimientos y conductas del individuo, hacia un objetivo cognoscitivo” (Kerlinger, 1973, citado en Brown, 1980, pág. 517), y a decir de Coolican (2005, pág. 118), utilizan afirmaciones más que preguntas en las que quien responde señala estar de acuerdo o en desacuerdo. Regularmente las escalas de opiniones *encubren* su objetivo. Entre los más comunes se encuentran:

#### a) Escalamiento de Likert

Los reactivos son presentados en forma de afirmaciones, que permite medir la postura del sujeto en tres, cinco o siete categorías. Los reactivos cumplen con el criterio de direccionalidad que permite hacer planteamientos sobre un determinado tema en sentido positivo o negativo (Hernández, Fernández, & Baptista, 2010, pág. 245).

#### Características

Es una escala psicométrica utilizada para la comprensión de las opiniones y actitudes. Se utiliza para medir y conocer el grado de acuerdo o desacuerdo de una

persona o encuestado hacia una oración afirmativa o negativa, por lo que utiliza un formato de respuestas fijos.

Asume que la fuerza e intensidad de la experiencia es lineal, por lo que va desde un totalmente desacuerdo a un totalmente de acuerdo, asumiendo que las actitudes pueden ser medidas. Las respuestas se presentan en diferentes niveles de medición, permitiendo escalas de 5, 7 y 9 elementos configurados previamente. Siempre se debe tener un elemento neutral para aquellos usuarios que ni de acuerdo ni en desacuerdo.

De acuerdo con el sitio QuestionPro (2018), la escala Likert presenta las siguientes ventajas y desventajas:

#### Ventajas

- Es de fácil aplicación y diseño.
- Puede utilizar reactivos no directamente relacionados con la expresión.
- Gradúa la opinión de las personas encuestadas.
- Es sencilla de contestar.

#### Desventajas

- Puede existir sesgo en la escala, ya que regularmente las respuestas positivas superan las negativas.
- Estudios indican que los encuestados tienden a contestar *de acuerdo* por implicar menor esfuerzo mental a la hora de contestar la encuesta.
- Dificultad para establecer con precisión la cantidad de respuestas positivas y negativas.

#### b) Escala de intervalos de Thurstone

Para cada afirmación de la escala se proporciona una puntuación equivalente a la fortaleza de esta afirmación, la cual es valorada por expertos, dando a cada reactivo un determinado puntaje, el cual será considerado como base para su valoración dentro de la prueba (Padua, 1996, pág. 180).

#### c) Diferencial Semántico de Osgood

Busca medir el significado connotativo del objeto para el individuo, es decir, las asociaciones que tiene el término para el individuo, estableciendo series de pares de adjetivos extremos que sirven para calificar la actitud del individuo, ubicándola en una categoría para cada par (Hernández, Fernández, & Baptista, 2010, pág. 255).

#### d) Escala de Guttman

Se conoce además como método del escalograma o análisis de escalograma. Define claramente lo que está midiendo, para establecer que al responderse un reactivo de manera positiva, los anteriores de igual forma se habrán respondido. Trata de encontrar la coherencia de la respuesta del sujeto, estableciendo un coeficiente de

reproductividad (Padua, 1996, pág. 195) y (Hernández, Fernández, & Baptista, 2010, pág. 260).

## **Clasificación de los Instrumentos**

### **Desde el punto de vista de su objetivo**

#### **De ejecución máxima o de potencia.**

Se exige el máximo rendimiento en la ejecución de una tarea. Existen tres factores determinantes en la calificación de estas pruebas:

- a) Habilidad innata: Capacidad potencial para ejecutar cualquier tarea que se presente a los sujetos.
- b) Habilidad adquirida: El ejercicio y el entrenamiento a través del aprendizaje, que a través de su interacción con el ambiente va obteniendo el individuo.
- c) Motivación: Grado de disponibilidad en que se encuentra el sujeto para ejecutar la tarea.

#### **Pruebas de ejecución típica:**

Se pueden designar como cualitativas o de rasgos. Se incluyen todas las pruebas que miden rasgos de personalidad.

### **Por la forma o tipo de respuesta que exigen**

#### **a) Pruebas objetivas.**

La respuesta deberá ser precisa de acuerdo al objetivo que persigan los reactivos, es decir, no deberá dejar lugar a duda respecto a lo que responde el examinado. Siendo entonces que se tiene una relación directa con las pruebas de ejecución máxima.

#### **b) Pruebas subjetivas.**

La respuesta depende del criterio de elección del sujeto y por tal motivo carece de un valor de bueno o malo por sí misma, requiriendo de conjuntar con las respuestas de otros sujetos para lograr la adecuada apreciación de la misma y señalar que tanto se acerca o aleja de la *normalidad*.

### **Por la forma de administración de la prueba**

#### **a) De aplicación colectiva**

Este tipo de pruebas son las más comunes, ya que tienen la ventaja de una aplicación a gran escala y son relativamente fáciles de calificar e interpretar. Comúnmente son confundidas con las pruebas de aplicación individual, por el hecho de que el ser posible aplicarla tanto a grupos grandes como a una sola persona le da una apariencia de aplicación individual. De igual forma el tener, por lo común una instrucción sencilla, pudiera permitirle al individuo por sí solo aplicarlo por lo que pudiera confundirse con lo que son las pruebas autoadministradas.

**b) De aplicación individual**

Requiere del registro de respuestas, el tiempo de las mismas, así también ciertas conductas que sea necesario que el investigador constante por sí mismo, por lo que no es posible su aplicación conjunta sea de dos o más sujetos.

**c) Autoadministradas**

Su principal característica es que tanto instrucciones, como aplicación, calificación e incluso la evaluación dependen exclusivamente del sujeto, evidentemente implica honestidad. Su utilidad se observa en los procesos previos de una evaluación en forma que se da por ejemplo, en instituciones y empresas dedicadas a los procesos de certificación.

**De acuerdo a la libertad de ejecución****a) Pruebas de poder**

Se le pide al sujeto que emplee toda su capacidad en el examen, es decir, que realice lo que más pueda sin escatimar esfuerzo en cuanto a sus habilidades y conocimientos, en cuanto a una actividad bien definida.

**b) Pruebas de velocidad:**

Se exige al examinado la ejecución más rápida que pueda realizar para terminar completamente la tarea asignada, en un tiempo determinado.

**c) Pruebas de libre ejecución:**

No existen límites de tiempo bien delimitados para la ejecución de la tarea aunque siempre se cuidará el tipo de prueba a aplicar para evitar que este sea excesivo.

**Por la forma de dar instrucciones**

Puede hacerse uso de las tres formas de instrucción al momento de aplicar alguna prueba, aunque siempre se deberá de tener cuidado en verificar cual es la idónea para hacer especial énfasis en esa, ya puede resultar molesto, innecesario e incluso humillante emplearlas de manera desmedida. La instrucción, en gran medida afecta la calidad de la respuesta, ya que una instrucción mal dada afecta de manera directa a la ejecución de la actividad a realizar.

**a) Oral**

Se utilizan principalmente en actividades lúdicas o cuando la persona no es capaz de leer, sea por no saber o por no poder hacerlo.

**c) Escrita**

Cuando la persona puede leer, es la mejor opción a utilizar en pruebas escritas, siempre cuidando el nivel de la persona para evitar presentar instrucciones que pudieran ser complejas en su entendimiento

**d) A través de ejemplos**

Se manejan principalmente en infantes, conlleva necesariamente el saber ejecutar adecuadamente la actividad a realizar por parte del instructor, ya que de lo contrario se puede tener un efecto adverso en la respuesta del sujeto.

**De acuerdo al material empleado****a) Lápiz y papel**

En las pruebas digitalizadas el principio es el mismo, no confundir con las pruebas que requieren aparatos especiales

**b) Verbal**

Refiriéndose al uso de grabadoras para registrar las respuestas del sujeto, como en el caso de las entrevistas

**c) Manipulable**

Implica ejecución (material manual y visomotor)

**d) Aparatos especiales**

Para realizar determinada tarea, tales como el polígrafo, el taquistoscopio o el pletismógrafo.

**Por su forma de calificación****a) Contando puntos**

De forma manual, contando puntos, aplicando matrices de calificación, o usando computadoras.

**b) Interpretación**

Sin ningún tipo de registro cuantitativo como en el caso de las pruebas proyectivas.

**De acuerdo a los límites de edad**

a) Para infantes o bebés.

b) Para niños preescolares.

c) Para niños escolares.

d) Para adolescentes.

e) Para adultos.

f) Para senectud.

El rango de edad depende del marco teórico que lo sustente y las finalidades propias de la investigación y la población a la que se dirige el instrumento

**Según el modo de expresión de la respuesta**

- a) Verbales
- b) Manipulativos
- c) Impresos
- d) Gráficos

**De acuerdo con el sector que exploren**

## a) Test de eficiencia:

Estudian los aspectos intelectivos y cognitivos de la personalidad y de acuerdo con los aspectos que abordan se subdividen en:

- 1) Test de inteligencia.
- 2) Test de aptitudes.
- 3) Test de conocimientos y aprovechamiento.

Los test de inteligencia y aptitudes miden aspectos del funcionamiento del intelecto en su mayor parte independientes de sus adquisiciones culturales. Los test de conocimientos y aprovechamiento miden las adquisiciones culturales del sujeto en determinadas áreas o materias.

## b) Test de personalidad:

1) Miden los aspectos de la personalidad tales como la estabilidad emocional, sociabilidad, intereses, actitudes, y se subdividen en:

- 1) Métodos de observación normalizados (entrevista normalizada, escala de estimación)
- 2) Encuestas de personalidad
- 3) Test de conducta objetiva
- 4) Métodos expresivos
- 5) Test proyectivo

**Reactivo**

Estímulo externo que provoca una respuesta controlada y dirigida. Que produce reacción. En psicología, se aplica el fenómeno que es consecuencia de otro fenómeno, a la conducta que aparece inmediatamente después de estímulos exteriores.

Permite medir conocimientos en diferentes niveles de pensamiento, memorizar, recordar, reconocer, comprender, relacionar, sintetizar, analizar y evaluar. En general son breves, explícitos y la calificación está claramente determinada, libre de incertidumbre o error

## Requisitos para el Desarrollo de Reactivos

Se requiere un conocimiento detallado y completo del área a investigar y de sus finalidades específicas; el discernimiento adecuado de los procesos mentales que se pretenden evidenciar en las personas que se han de tomar la prueba.

### a) Pregunta directa o afirmación

Al construir un instrumento, una de los elementos a considerar es, si se colocan preguntas o afirmaciones, aunque en ocasiones y en apariencia son lo mismo, es claro el señalamiento que las escalas de medición se realizan principalmente con afirmaciones (ver Coolican (2005) y Hernández, Fernández y Baptista (2010). Se debe tener cuidado al plantear, pues no es simplemente el eliminar signos o agregarlos (¿Te gusta el pastel? / Te gusta el pastel), sino la intención de lo escrito con respecto a poner en evidencia alguna característica psicológica que se quiera evidenciar.

### b) El tema central del reactivo, debe referirse a una sola idea

Aunque se observa principalmente en procesos de entrevista, en la construcción de instrumentos es factible encontrar reactivos mal estructurados en los que se colocan al menos dos ideas que, al responder el encuestado difícilmente se sabrá que fue lo que respondió (¿Cómo se encuentra tu novio, tu familia?, ¿Qué puedes decir de tus conocimientos, habilidades? ¿Vienes a terapia por voluntad propia o porque te manda tu profesor?) En el último ejemplo, se dan incluso, alternativas de respuesta que rompe con la espontaneidad de la misma.

### c) Complejidad

Los reactivos han de ser sencillos y de fácil comprensión. Se debe evitar el saturar de información al reactivo que no sea necesaria al mismo, salvo que sea la intención para evidenciar procesos mentales complejos

La labor del psicólogo en la discriminación hacia grupos minoritarios debe ir más allá de mantener un papel neutral en este tipo de procesos y hacer patente su rechazo a cualquier expresión que afecte al ser humano, independiente de sus ideas, creencias o preferencias, asumiendo un papel activo de transformación social, en el cual rompa con esquemas arcaicos de pasividad y formar parte del cambio social.

### d) El fenómeno a apreciar

El objetivo se puede confundir dentro de la amplitud de información del tema, por lo que los reactivos habrán de evidenciar los procesos que quieren conocerse, y deberán de dirigirse a ese objetivo teniendo cuidado en no plantear los reactivos concernientes al tema, pero sin referir al objetivo (¿Te gusta el pastel? / ¿Sabes hacer pasteles?)

### e) Claridad en la expresión

Evitar términos o expresiones que confundan al examinado y que en consecuencia traiga respuestas igual de ambiguas (¿Qué tanto significan para ti las cantidades?), ¿Cómo entiendes el concepto de medición?, ¿Qué observas en la siguiente imagen?

**f) Brevedad en la expresión**

Procurar sean lo más corto posible, evitar una construcción amplia del reactivo para decir poco, hay que considerar la razón del mismo y evitar dificultar el reactivo de manera innecesaria (Si tienes la oportunidad de elegir entre visitar dos lugares diferentes como una gran ciudad o un bosque aislado de la civilización ¿qué prefieres? / ¿Prefieres visitar una gran ciudad o un bosque aislado de la civilización?

**g) Construcción gramatical engorrosa**

Al elaborar un reactivo debe evitarse manejar un lenguaje demasiado rebuscado o complejo que iría mejor en una obra literaria, tal como la poesía.

¿Es del agrado tuyo el ejecutar y realizar operaciones aritméticas?

¿Le gustaría paladear la corteza de cerdo en salsa esmeralda?

**h) Afirmaciones tímidas o ambiguas**

El reactivo debe ser claro y preciso en su redacción, evitando expresiones que provoquen confusión y denoten inseguridad en quien los presenta.

¿Cómo vez la aritmética?

¿Consideras que pudieras ser mi novia?

**i) Evitar adiciones inconexas y falsas**

Al redactar un reactivo debe evitarse presentar información que no sea acorde a la información que se presenta, principalmente en exámenes de ejecución máxima, ya que estresa de manera innecesaria al examinado.

Sigmund Freud en su proceso de *análisis cognitivo* determinó la influencia de la libido en la psique humana

**j) Evitar fuentes ajenas de dificultad de los reactivos**

Debe evitar sobrecargarse el reactivo con información innecesaria, ya que esto es más bien un distractor que poco ayuda a dar una respuesta adecuada.

**k) Evitar las formas estereotipadas**

Son monótonos y se puede provocar que el examinado se confunda al momento de responder y coloque su respuesta en un lugar equivocado dando la impresión de no saber la respuesta

Explique en qué consiste la teoría psicoanalítica y justifique su respuesta

Explique en qué consiste la teoría gestaltista y justifique su respuesta

Explique en qué consiste la teoría conductista y justifique su respuesta

Explique en qué consiste la teoría cognitiva y justifique su respuesta

Explique en qué consiste la teoría estructuralista y justifique su respuesta

Explique en qué consiste la teoría constructivista y justifique su respuesta

**l) Incluir un máximo de términos comunes en el reactivo**

Aunque se busca en ocasiones el manejo de cierto tipo de lenguaje que deberá ser conocido y dominado por quien es examinado, habrá de manejarse con mesura al momento de elaborarse los reactivos

Define el proceso de medición en una prueba de ejecución máxima, considerando una aplicación individual y respuesta de velocidad con reactivos de opción múltiple

**m) Evitar los reactivos con doble negación.**

- 1) ¿No se puede estudiar cuando no se cuenta con las condiciones adecuadas?
- 2) ¿No es verdad que no irás a la fiesta?
- 3) ¿Puedes no tener hambre cuando no has comido en días?

**n) Resaltar cualquier forma negativa de presentación**

- 1) ¿Estás segura de que ya *nunca* lograrás superar tus problemas?
- 2) ¿Cuál de las opciones es la *incorrecta*?
- 3) ¿Bajo que circunstancias es *inaceptable* la aplicación de una prueba psicométrica en casos periciales?

**o) Procurar evitar el uso de la palabra NO**

- 1) ¿No te gusta?
- 2) ¿Qué no le dijiste?
- 3) ¿No hay preguntas?

**p) Sesgo cultural y de género**

Procurar plasmar un vocabulario de uso común al utilizado por la población objeto de estudio, evitar regionalismos que dificulten la pregunta, o que ofenda a los participantes por su condición social, cultural o de género.

**Tipos de Reactivos**

**a) Reactivos de Verdadero y Falso**

El tipo de reactivos menos recomendable, desventajas: 1) Conduciendo a limitar la prueba a datos concretos; 2) Dificultad de incorporar dentro de los estímulos en sí la norma de veracidad o falsedad, por tal el estímulo será juzgado; 3) A menudo conduce a reacciones emocionales.

**b) Reactivos de Selección Múltiple (Opción múltiple)**

Es el más recomendable, ventajas: 1) Contienen en sí mismo la norma para escoger la mejor respuesta; 2) La mejor contestación que se da no tiene que ser la única e indiscutible respuesta a la pregunta; pero debe ser defendida como la mejor alternativa

que se presenta; 3) Puede presentarse como pregunta directa o afirmación incompleta; 4) Es flexible, varios reactivos pueden dirigirse al mismo conocimiento, o varios estímulos diferentes pueden seleccionarse a la vez en un grupo de alternativas; 5) Tiene todas las ventajas de la objetividad de la calificación. 6) Manejo de opciones no relacionadas que permite más certera la evaluación. 7) Permite la utilización de claves precisas y por ende una adecuada calificación.

**c) Reactivos de Aparejamiento (Relación de columnas)**

Son variante de los de selección múltiple, para cada estímulo de un reactivo, se elige de entre un grupo de alternativas. Fallas comunes: 1) Tener el mismo número de elementos en cada lista, se soluciona incluyendo en las alternativas, más elementos que en los estímulos, o la misma alternativa pueda contestar a más de un estímulo; 2) Incluir demasiados elementos en ambas listas, pudiendo superarse al presentar grupos homogéneos de estímulos y alternativas en listas separadas.

**d) Reactivos de Jerarquización**

El problema del sujeto es indicar, para los elementos que se le presentan, el orden apropiado, quizá cronológico o lógico. El problema estriba en calificar los segmentos de orden correcto que se den en la respuesta.

**e) Reactivos por Completar (Complementación)**

Son de dos formas: 1) De respuesta corta; 2) afirmaciones con más opciones en blanco. Su calificación no rutinaria y no es objetiva en su grado máximo; puede caer en el error de creer que sólo existe una forma de contestarlo eliminando otras opciones. (Brown, 1980) Miden conocimientos factuales, comprensión de principios y habilidad para identificar y definir conceptos.

**f) Reactivos de Ensayo**

No se diferencia claramente de los reactivos de completar, siendo más allá de la longitud y la función que desempeña la ejecución del reactivo. Proporcionan bases para evaluar capacidad de organización, integración y evaluación de conocimientos. Además de actitudes, creatividad y fluidez verbal del evaluado.

**g) Problemas**

Describe una situación y da ciertos datos pertinentes, la tarea del evaluado es resolver el problema presentado. Incluye identificar lo que se pregunta y parámetros y datos importantes, la formulación de ecuaciones o procedimientos analíticos y la aplicación de este último a los datos para obtener una solución.

**h) Reactivos de Ejecución**

La tarea se estructura de modo que los evaluados sepan exactamente lo que tienen que hacer y en qué condiciones. El saber qué hacer, se especifica en el objetivo de instrucción (el resultado deseado define la tarea). Las especificaciones de las condiciones, se refiere a factores como la longitud del tiempo de que se dispone para realizar la tarea, las ayudas que se pueden utilizar y otras variables que afecten la ejecución.

**Número de reactivo:**

No existe un acuerdo del número de reactivos que deberá contener una prueba, ni tampoco criterio o fórmula alguna que permita señalar con claridad cuántos reactivos deberá tener desde el inicio la prueba, y solo se realizan observaciones como las “justas necesarias para realizar el trabajo” (Murillo, 2006, pág. 4), evidentemente esto, carece de toda objetividad, aunque si hay alternativas para discriminar que reactivos son los que podrían ser considerados, tales como “sólo preguntas que no pueden responderse de otro modo o, mejor dicho, aquéllas que puede conseguirse la información por otras fuentes no deberían entrar en el cuestionario” (Murillo, 2006, pág. 4). Es decir, al elaborar reactivos se puede crear una gran cantidad de los mismos, pero es necesario hacer un posterior análisis que permita observar cuáles son los que mejor se dirigen al objetivo, y que no es posible adquirir la información por otros medios. En este sentido, la cantidad de reactivos dependerá de la creatividad del autor, así como de su capacidad de análisis para discriminar aquellos que mejor se ajustan a su objetivo.

### **Dificultad y Discriminación del los Reactivos**

#### **a) Índice de Dificultad o Aprobación del Reactivo**

Se entiende como la proporción de personas examinadas que responden correctamente al reactivo y se establece a través de la siguiente fórmula:

$$p = \frac{\text{número de aciertos en el reactivo}}{\text{número total de examinados}}$$

$p_1$  = índice de dificultad del reactivo 1. Teóricamente el valor puede variar de 0 si nadie lo tuvo correcto a 1 si todos lo tuvieron correcto.

*Nota:* El término *Aprobación del Reactivo* se aplica en contextos como las pruebas de personalidad, en las que en conjunto de personas a las que se le aplica la prueba y que señalaran coincidencia al responder en alguna de las opciones, señalando esto el índice de aprobación del reactivo.

El cálculo del índice promedio de dificultad de los reactivos se calcula promediando los índices de dificultad del reactivo para todos los reactivos de la prueba.

$$\text{Índice promedio de dificultad} = \frac{\text{suma de índices de dificultad}}{\text{número de reactivos (total)}}$$

Incluye el índice simple de la dificultad de los reactivos, como el porcentaje de los sujetos que contestan correctamente y la medida de la relación del reactivo específico a la prueba total de la que es parte.

La dificultad óptima de la prueba se recomienda en un promedio del 50%, es decir, 50% de los estudiantes los resuelve correctamente y otros 50% erróneamente. Puede variar la dificultad de la prueba en una escala del 15% al 85%, aunque este es un criterio arbitrario.

#### **b) Índice de Discriminación del Reactivo**

“Es un índice estadístico de la eficiencia con que un reactivo discrimina entre personas que obtienen altas y bajas puntuaciones en la prueba completa” (Gregory, 2001, pág. 159).

Su cálculo es a través de la formula:

$$d = \frac{(S - I)}{N}$$

En donde S es igual al número de personas evaluadas que están ubicadas en el rango superior y respondieron adecuadamente el reactivo, I corresponde al número de personas evaluadas que están en el rango inferior y que respondieron adecuadamente la prueba y N corresponde al total de personas evaluadas en los rangos superior e inferior.

Para cálculo del rango superior e inferior no hay una propuesta específica, sino que es más bien el criterio arbitrario de quien hace la evaluación, pero se considera como aceptable considerar el 25% para cada extremo, es decir, si se cuentan con 100 sujeto en total, se tomará en consideración los 25 individuos que hayan sacado las puntuaciones más altas, así como los 25 con calificaciones más bajas.

Al obtener los resultados por cada reactivo, se buscarán valores positivos como los adecuados y más aún si se acercan al +1.0 ya que señalaría una clara diferencia entre los altos y bajos puntajes.

## **Validez y Confiabilidad**

### **Validez**

Grado de exactitud con que un instrumento mide realmente la característica de la personalidad o el modo de conducta que se propone medir (4).

“La validez, en términos generales, se refiere al grado en que un instrumento mide la variable que pretende medir.....La validez es una cuestión más compleja que debe alcanzarse en todo instrumento de medición que se aplica” (5)

La Validez responde a varias preguntas de suma importancia ¿Qué rasgos está midiendo la prueba? ¿Mide la prueba el rasgo para lo que fue construida? ¿Qué porcentaje de la varianza en las calificaciones de la prueba se pueden atribuir a las variables que mide? ¿Qué se puede predecir a partir de las calificaciones de la prueba? ¿Proporciona información que se pueda utilizar en la toma de decisiones? (Brown, 1980)

### Tipos de validez

Los dos principales tipos son: validez externa y validez interna:

La validez externa se refiere a la habilidad del instrumento para ser generalizado a través de personas, escenarios y tiempo.

La validez interna se refiere a la habilidad de un instrumento de medir lo que él se propone medir. Por lo general la validez interna es la que comúnmente se le dice validez.

Por otro lado, una ampliamente aceptada clasificación de validez interna consiste en tres principales tipos de evidencia: 1) validez de contenido, 2) validez relacionada con el criterio y 3) validez de constructo.

### Validez de contenido

Se refiere a que los reactivos que contenga el instrumento, concuerde con los objetivos que han sido reconocidos o establecidos. Los principios que lo rigen son los siguientes:

- a) El instrumento debe medir una sola función (por ejemplo: aprovechamiento, no actitudes).
- b) El contenido del instrumento debe ser representativo del objetivo.
- c) El instrumento debe contener suficientes reactivos de manera que se obtenga una muestra representativa del contenido y los objetivos. Durante la construcción del instrumento se establecen tablas de especificaciones de la proporción de preguntas que debe contener el instrumento en relación al énfasis y la prioridad de los objetivos.
- d) Los reactivos del instrumento deben estar bien formulados, asegurándose que no existan indicadores que faciliten la adivinanza al azar de las respuestas correctas.

La validez de contenido se establece mediante la inspección de los objetivos y los materiales con los que se cuenta para el cumplimiento y logro de objetivos.

Validez de contenido (Brown, 1980).Pregunta: ¿Cómo podría comportarse el individuo en el universo de situaciones de las que los reactivos de la prueba son una muestra?

Evaluación: Mediante la estimación de si la muestra de reactivos es o no adecuado.

Orientación: Hacia la tarea o conducta que procesa el instrumento.

**Validez de Contenido a través de un Cuadro Comparativo**

Datos Generales

Objetivo:	Definir qué es lo que se pretende medir
Población:	Señalar las características de la población a la que se habrá de dirigir los reactivos (edad, escolaridad, etc.)
Importancia del Trabajo:	Señalar el por qué es importante medir lo que se pretende medir
Forma de Responder a la Prueba:	Señalar el tipo de escala (Likert, Thurstone, etc.) así como la gama de opciones de respuesta a considerar (ejem. Nunca, A veces, Siempre)

Instrucciones para el Juez: Califique los reactivos de acuerdo a los criterios señalados:

Relación Factor-Reactivo				Anexo						
Marque con una X en la casilla correspondiente si a su parecer, los reactivos se dirigen de manera adecuada (3), Parcialmente adecuada (2) o Inadecuadamente (1) a describir el factor correspondiente:				Considere las opciones de este cuadro solo en caso de observaciones al reactivo						
3. Adecuada				El reactivo es entendible para el tipo de población con el que se pretende trabajar						
2. Parcialmente adecuado				La redacción general del reactivo es correcta y entendible						
1. Inadecuada				El reactivo predispone a la elección libre de cualquier opción de respuesta						
Factor A: (Concepto y Definición Operacional)	1	2	3	No	Si	No	Si	No	Si	Otras Observaciones
Reactivo 1										
Reactivo 2										
Reactivo 3										
Reactivo 4										
Reactivo k										

Nombre y Firma del Juez

Relación Factor-Reactivo				Anexo					
Marque con una X en la casilla correspondiente si a su parecer, los reactivos se dirigen de manera adecuada (3), Parcialmente adecuada (2) o Inadecuadamente (1) a describir el factor correspondiente:				Considere las opciones de este cuadro solo en caso de observaciones al reactivo					
3. Adecuada				El reactivo es entendible para el tipo de población con el que se pretende trabajar					

2. Parcialmente adecuado				La redacción general del reactivo es correcta y entendible						
1. Inadecuada				El reactivo predispone a la elección libre de cualquier opción de respuesta						
Factor B: (Concepto y Definición Operacional)	1	2	3	No	Si	No	Si	No	Si	Otras Observaciones
Reactivo 1										
Reactivo 2										
Reactivo 3										
Reactivo 4										
Reactivo k										

Nombre y Firma del Juez

Relación Factor-Reactivo				Considere las opciones de este cuadro solo en caso de observaciones al reactivo						
<p>Marque con una X en la casilla correspondiente si a su parecer, los reactivos se dirigen de manera adecuada (3), Parcialmente adecuada (2) o Inadecuadamente (1) a describir el factor correspondiente:</p>				El reactivo es entendible para el tipo de población con el que se pretende trabajar						
3. Adecuada				La redacción general del reactivo es correcta y entendible						
2. Parcialmente adecuado				El reactivo predispone a la elección libre de cualquier opción de respuesta						
1. Inadecuada										
Factor C: (Concepto y Definición Operacional)	1	2	3	No	Si	No	Si	No	Si	Otras Observaciones
Reactivo 1										
Reactivo 2										
Reactivo 3										
Reactivo 4										
Reactivo k										

Nombre y Firma del Juez

### Validez de criterio

El criterio queda definido como la medida externa que existe con anterioridad al instrumento en cuestión y que es aceptado por expertos como un índice adecuado o satisfactorio de la característica que el nuevo instrumento pretende medir. Se trata de medir la ejecución futura de un individuo en alguna variable significativa (el criterio). Este tipo de validez se divide en concurrente y predictiva.

a) Validez concurrente: Grado hasta el cual el instrumento mide lo mismo que miden otros instrumentos ya existentes, considerados como medidas adecuadas del mismo rasgo. De igual manera se da cuando el instrumento se correlaciona con algún criterio existente que mida lo mismo que el instrumento (ejemplo: en una fábrica se utiliza la producción de las personas empleadas como criterio para evaluar empleados nuevos y como criterio para validación de una prueba para selección de personal). La validez concurrente de un instrumento no indica nada acerca de la capacidad predictiva de éste, solo señala que el instrumento nuevo que se ha desarrollado para medir un rasgo es tan eficiente o válido como el criterio previo.

b) Validez predictiva: Es el grado hasta el cual la ejecución de un instrumento sirve para predecir eficientemente la conducta futura de las personas que se han examinado. Se deben señalar coeficientes de validez predictiva para cada tipo de predicción que se pretenda hacer con un mismo instrumento (ejemplo: exámenes de inteligencia para medir éxito académico de los sujetos en distintos grados escolares). La predicción del comportamiento futuro, con un instrumento que se administre en el presente, se limita al lapso incluido en la determinación de la validez predictiva del instrumento. Solo es válida su capacidad predictiva de acuerdo al lapso en que se ha probado. La validez predictiva es esencial y se requiere en las pruebas de aptitud e inteligencia. El psicólogo que hace predicciones sobre comportamiento futuro sin tener sustento para tales, incurre a una mala práctica sujeta a consideraciones éticas y legales. Aiken (2003), señala que la precisión en la predicción que ofrezca una prueba, encierra una *validez creciente*, que permite decidir si la aplicación de un instrumento, se justifica por su costo a través de la satisfacción de los propósitos de la evaluación sin necesidad de usar una prueba adicional, es decir, observa la calidad de la predicción, cuando no se usa otro instrumento o procedimiento de evaluación.

La diferencia entre ambas formas de validez (concurrente y predictiva), se da en el tiempo que transcurre entre la aplicación del nuevo instrumento por validarse y la aplicación del criterio que se da para validarlo. Si el lapso entre la aplicación del nuevo instrumento por validarse y la aplicación del criterio es corto, se está trabajando con validez concurrente. Pero si el lapso es largo, se está trabando con la validez predictiva. Ambos tipos de validez se miden computando un coeficiente de correlación entre los resultados de ambas administraciones (coeficiente de validez).

Validez de criterio (Brown, 1980).

Pregunta: ¿Qué tan bien predicen las calificaciones del instrumento la situación o ejecución de alguna medida independiente?

Evaluación: Mediante la comparación de calificaciones de la prueba con las de la medida independiente (cualitativamente diferente).

Orientación: Hacia el criterio, la variable predicha.

Ejemplo: Uso de una prueba de aptitudes escolares para predecir el promedio de calificaciones en la universidad; el empleo de una prueba de aptitudes para la mecánica, para predecir el éxito como mecánico automotriz; la utilización de un inventario de personalidad para predecir que conductores de automóvil tendrán accidentes.

### Validez de constructo

Mide una característica de la personalidad cuya existencia se infiere de las respuestas del sujeto a los reactivos. A través de ella se pretende medir un postulado hipotético. El psicólogo ha de desarrollar hipótesis en sus teorías psicológicas que puedan ser sometidas al proceso de medición y que, mediante dicho proceso se refuten o sostengan. Se requiere evidencia de este tipo de validez para todo tipo de instrumento. Conlleva la recopilación de datos y de estudios que sustenten y corroboren las premisas, corolarios e hipótesis que conformen la red nomológica de una teoría psicológica en particular.

### Nomología

Parte de una ciencia que establece leyes y reglas generales (hipótesis nomológica) aplicables a los hechos de los que se ocupa la ciencia en cuestión. El establecimiento de hipótesis nomológicas es tarea central de toda ciencia experimental (*Dorsch, 1981*)

### Métodos para obtener estimados de la validez de construcción

- a) Diferencia de grupo: Si la teoría de la que se deriva el constructo hipotético que el instrumento pretende medir estipula que las personas que tienen el rasgo que se mide difieren de las personas que no poseen este rasgo, el constructor del instrumento debe probar que en realidad existen diferencias de grupo entre los sujetos a los que se les aplica el instrumento. (Ejemplo MMPI)
- b) Matrices correlacionales: Ilustra de manera grafica la relación (o no relación) entre el nuevo instrumento y los demás instrumentos psicológicos usados en el proceso de validación (debe dar alta y positiva con otros instrumentos que midan la misma característica).
- c) Análisis de factores: Prevé un estimado de la carga de factores que el examen tiene en cada una de sus partes y en su totalidad. Mediante este análisis se puede identificar el factor común o los factores comunes que contribuyen a la varianza total de los resultados de un instrumento.
- d) Estudios de cambio a través del tiempo: Si la teoría psicológica que postula la existencia de un rasgo estipula que este es estable, la estabilidad de los resultados a través del tiempo arroja luz sobre la validez de construcción del instrumento. Si postula que el rasgo es uno que cambia a través del tiempo, la estabilidad de los resultados que se obtienen por medio de administraciones sucesivas del mismo instrumento a las mismas personas se usará como evidencia en contra de la validez de construcción lógica de dicho instrumento.

- e) Estudio sobre la estructura y la consistencia interna del instrumento de medición: Cuando la teoría que postula la existencia de una característica específica indica que ese rasgo es homogéneo, entonces los estudios de confiabilidad de consistencia interna de los instrumentos arrojan evidencia de la validez de construcción de ese instrumento. Si la teoría postula que el rasgo que se está midiendo es heterogéneo (compuesto), la consistencia interna de la prueba irá en contra de la validez de construcción lógica del mismo.
- f) Estudio de procesos: Para establecer índice de la validez de construcción de un instrumento, se debe demostrar que sus resultados correlacionan alta y positivamente con los resultados de otros instrumentos que midan lo mismo, pero además hay que demostrar que los resultados del instrumento que se está validando difiere significativamente de los resultados de otros instrumentos que midan otros rasgos de personalidad. Este proceso de validación proporciona estimados de lo que se conoce como validez convergente y validez divergente.

#### Otros Tipos de Validez

Validez de Aspecto: Se habla de “inspeccionar el contenido de una prueba, para ver si en verdad mide lo que se supone que tiene que medir” (Coolican, Métodos de Investigación y Estadística en Psicología, 1997, pág. 183). Este tipo de validez deja toda la responsabilidad de la *probable* medición del instrumento a quien lo diseña, ya que, si bien se busca que los reactivos cubran en cantidad y calidad suficiente el aspecto que interesa medir, no evita o controla un posible sesgo no intencionado de los reactivos que se presentan al hacerlos tendenciosos, demasiado complejos o mal estructurados, por carecer de algún otro elemento de control sobre la redacción de los mismos.

### Confiabilidad

Es el “grado en que un instrumento produce resultados consistentes y coherentes” (Hernández, Fernández, & Baptista, 2006, pág. 277)

Atiende a la pregunta ¿Hasta qué punto será consistente la ejecución de una persona si toma otra forma de la prueba psicológica (equivalencia) o si toma la misma forma en un momento diferente (estabilidad)?

#### Tipos de Confiabilidad

##### Confiabilidad Test-Retest:

Una medida posible de confiabilidad es la correlación existente entre medidas repetidas (o sea, entre una prueba y una reaplicación), a través del coeficiente de estabilidad.

Su paradigma se basa en la aplicación de la prueba, dejar transcurrir cierto tiempo y aplicar nuevamente. El intervalo puede variar de minutos hasta varios años.

Existen varias suposiciones en las que se basa la confiabilidad Test-retest:

- 1) La característica que mide el instrumento es estable en el tiempo.
- 2) Debemos suponer que no existe ningún efecto diferencial de olvido o práctica.

3) No debe producirse ningún aprendizaje diferencial entre las dos aplicaciones.

Cualquier variable que influya en la ejecución de una aplicación, pero no en la otra, reducirá la correlación. Cualquier error asociado a una aplicación específica de la prueba afectará a la confiabilidad del test-retest. Al igual que con las fluctuaciones cotidianas de ánimo, salud y motivación.

#### Confiabilidad de Pruebas Paralelas:

Este tipo de estimación de confiabilidad se refiere a la consistencia sobre las formas de una prueba, es decir, se establece un coeficiente de equivalencia.

Las formas paralelas cubren el mismo contenido, utilizan los mismos tipos de reactivos y tienen un grado de dificultad igual. Se requieren al no ser factible ni conveniente la segunda aplicación de una misma prueba.

Para determinar la confiabilidad de formas equivalentes, se debe aplicar una de ellas (forma A) y, a continuación, con un periodo mínimo transcurrido, la segunda (forma B). La correlación de los resultados obtenidos en ambas formas da el coeficiente de equivalencia.

Para contrarrestar los efectos debidos al orden de aplicación, a la mitad del grupo se le aplica la forma A seguido por la B, y a la otra mitad la forma B y a continuación la A.

Ambas pruebas deben ser equivalentes en contenido, formato y longitud, además de dificultad (que tengan medias similares) y en variabilidad (desviaciones estándar similares).

La falta de consistencia en las calificaciones se puede atribuir a diferencias en el muestreo de reactivos, además de que no se pueden controlar del todo las fluctuaciones a corto plazo debidas al estado de ánimo del sujeto o a las diferencias de aplicación de las dos formas.

#### Confiabilidad por Mitades:

Requiere de dos mitades independientes, esta división se realiza generalmente utilizando los reactivos de número impar en una de las formas y pares en la otra. La correlación de esos dos resultados proporciona una estimación de la confiabilidad.

El problema que presenta, es que cada conjunto de calificaciones se basa en la mitad de los reactivos de la prueba de origen. Al ser más corta, la confiabilidad será más baja que la que se espera de una prueba en su longitud original.

La estimación de la confiabilidad por mitades es un coeficiente de equivalencia, ya que ambas formas (mitades) se aplican en esencia en forma simultánea y sólo las fluctuaciones a corto plazo afectan su confiabilidad.

#### Consistencia Interna:

Es la consistencia de la ejecución a través de los reactivos que constituyen la prueba.

## **Estandarización o normalización**

La estandarización es el elemento primordial a la que se aspira en la elaboración de un test. En psicometría se entiende como la “uniformidad de procedimientos en la aplicación de un test y en la determinación del puntaje correspondiente” (Galimberti, 2002, pág. 448), permite "la definición de puntuaciones significativas comparados con el rendimiento de un 'grupo de estandarización' al que se administró previamente el test" (Myers, 2011, pág. 319). Es decir, se procura la adecuada interpretación de la información con base a los resultados obtenidos de la aplicación de una prueba en una determinada población. Dichos resultados permiten visualizar la forma de expresión de un determinado fenómeno psicológico en la población objeto de análisis.

A decir de Coolican (2005, pág. 116), requiere de ser rigurosamente cuidadoso en los métodos de muestreo, ya que permitirá la interpretación de lo que ocurre con la población, de tal forma que se debe especificar con claridad las poblaciones para el establecimiento de *normas de prueba*, en sus diversas subcategorías de la población general. Para ello es importante no solamente observar los resultados obtenidos, sino además vigilar la adecuada aplicación y calificación del instrumento en cuestión, ya que de ello se derivará la objetividad de los resultados y su adecuado análisis e interpretación.

Cronbach (1972, citado por Aliaga, s.f., pág. 88) señala que la estandarización es el “proceso mediante el cual se establecen procedimientos unívocos para la aplicación, calificación e interpretación de un test psicométrico”. Es decir, no es posible la adecuada interpretación de una prueba cuando no se cumplen las condiciones de aplicación y uso de la misma de acuerdo a los lineamientos señalados por el autor. La interpretación a voluntad de quien la aplica sin consideración de las normas establecidas corrompe la estandarización, como consecuencia, las distorsiones e inadecuadas interpretaciones de los resultados obtenidos, son entonces meramente responsabilidad de quien administre la prueba.

En este sentido, 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<sup>3</sup> o tablas de normas. El establecimiento de las normas, depende del proceso de *estandarización*. De las formas de establecer esta normalización, las más comunes son:

### **Fractiles**

Son los “datos o puntajes que limitan por la derecha una proporción cualquiera de una distribución” (López, 2011, pág. 69), y permiten establecer las normas estimadas de una prueba, entre las más usuales se encuentran los percentiles o centiles, deciles y cuartiles. El rango percentilar, es el “porcentaje de casos que caen en o están debajo de un valor específico de  $x$ ” (Ritchey, 2002, pág. 59). Los cuartiles son “fractiles que identifican los valores de la puntuación que fracciona una distribución en cuatro grupos del mismo tamaño” (op. cit., pág. 60).

### **Estaninas**

Convierte cualquier conjunto de puntuaciones en una escala transformada que va de 1 a 9. La estandarización en estos términos busca como referencia una media de 5 y una desviación estándar de 2 (Kaplan & Saccuzzo, 2006, pág. 52)

---

<sup>3</sup> Baremo: Conjunto de normas establecidas convencionalmente para evaluar algo.

**Puntuación z**

La puntuación z “expresa una puntuación en función del número de unidades de desviación estándar que la puntuación cruda está por debajo o por encima de la media de la distribución” (Cohen & Swerdlik, 2001, pág. 118). Esta prueba permite la comparación de las puntuaciones crudas dentro y entre las pruebas. Hacer una comparación significativa, requiere de semejanza entre los grupos de comparación de los que se calculan las puntuaciones estándar. Las puntuaciones Z abarcan valores negativos y positivos y un punto cero intermedio

**Puntuación T**

Es un tipo de puntuación estandarizada, que tiene una media de 50 y una desviación estándar de 10. Este tipo de puntuaciones son comunes en las pruebas de personalidad (Gregory, 2001, pág. 80). El sistema es el mismo que las calificaciones estándar (calificaciones z), con excepción de las variaciones en la media y la desviación estándar (en calificaciones z es 0 y 1 respectivamente) (Kaplan & Saccuzzo, 2006, pág. 50).

## **Estructuración del Instrumento**

Retomando a Dietrich (1996, pág. 185) es necesario mencionar que en el cuestionario, es factible diferenciar dos elementos que lo conforman: la cabeza y el cuerpo.

### **1. Elementos de la cabeza:**

#### **1.1. Datos de Identificación**

##### **1.1.1. Institucionales**

Identificación del *ente*, entendiéndose como, grupo o institución, responsable del instrumento, lo que incluye: Logos institucionales (en su caso), y Nomenclatura de la institución.

##### **1.1.2. Personales**

Datos de identificación del examinado, esto de acuerdo a la finalidad del instrumento, pudiendo incluir desde el nombre, edad, género, religión, ocupación, escolaridad y, otros datos como la fecha de aplicación. No olvidar que *solo se solicitarán aquellos datos que sean pertinentes a la investigación a realizar*, por verbigracia, si se requiere identificar si una persona es soltera en una investigación, no se indagará si tiene novio(a) a menos que se requiera para los fines de la investigación. Es importante considerar que cualquier dato que no sea pertinente conlleva el rompimiento de principios éticos relacionados con la calidad de la investigación, así como la confidencialidad de la identidad del sujeto.

#### **1.2. Objetivo**

En ocasiones es necesario no presentar de manera clara y directa del todo la finalidad del instrumento para evitar una predisposición negativa del encuestado, por lo que se recurre al *encubrimiento del objetivo*, el cual ayuda para obtener información válida en temas embarazosos o amenazantes. El *encubrimiento* requiere de mucho cuidado al utilizarlo para no romper los principios éticos de honestidad en la investigación (Coolican, 2005, pág. 101). De esta forma se evita perjudicar o predisponer negativamente a los individuos que habrán de responder la prueba, por ejemplo: No hablar de *determinar si una mujer sufre de maltrato en su relación de pareja*, cuando se puede señalar como *determinar el tipo de relación de pareja*.

#### **1.3. Confidencialidad**

Cualquier tipo de instrumento en la actualidad, debe hacer patente que el manejo de la información se manejará de manera confidencial, ya que esto garantiza al examinado, que podrá expresarse de manera libre sin temor a represalias o señalamientos por las respuestas emitidas en el instrumento. En ocasiones, se requiere del manejo de un formato de *consentimiento informado* (más adelante se profundiza sobre el tema), que respalde incluso de manera legal a ambas partes en cuanto alcance y manejo de la información que se obtenga del instrumento en cuestión (Ver [Ética en la Medición](#)).

#### **1.4. Instrucciones**

Son el conjunto de las reglas para un fin, las cuales consideran la forma de expresión, dependiendo el tipo de población a la que se dirige, y el conocimiento claro y pleno de la prueba.

Se recomienda reconocer y observar las diversas actividades que se habrán de realizar, si es posible tratar de ensayar el cómo *contestar* previamente la prueba, para resolver las dudas que se presenten, verificar que se esté resolviendo adecuadamente la prueba.

Evitar obviar la información y el entendimiento de la misma y buscar la forma de expresión apropiada y libre de interpretaciones inadecuadas

**1.4.1. Para el examinador**

1.4.1.1. Antes del examen: Presentación personal, condiciones del lugar (iluminación, disponibilidad del espacio y distribución dentro del mismo, ventilación, ruido y otros distractores), materiales, conocimiento de la prueba, explicación adecuada de las instrucciones.

1.4.1.2. Durante del examen: Verificar se entiendan bien las instrucciones, resolver dudas, mantener el orden en el área de trabajo, evitar distractores.

1.4.1.3. Después del examen: Agradecer la participación, recoger y verificar se entreguen todos los materiales.

**1.4.2. Para el examinando:**

Deberán ser clara y precisa, en relación a:

1.4.2.1. Las actividades a realizar

1.4.2.2. La disciplina a guardar

1.4.2.3. La resolución de dudas

**1.5. Agradecimiento por la participación**

Aunque es un meramente un detalle de cortesía, es necesario reconocer y agradecer a quienes participen al responder a la prueba, incluso hay ocasiones en que se da algún pequeño obsequio que de manera simbólica demuestre el reconocimiento por parte de quienes aplican un instrumento, hacia quienes amablemente conceden un poco de su tiempo para contestarlo.

**2. Elementos del cuerpo**

Aunque pareciera simple, la conformación y presentación de los mismos deberá hacerse de tal forma en que sea entendible y de fácil manejo para quienes habrán de responderlo, pero de igual forma deberá contemplar la facilidad en el acceso a los datos por parte de quien se encargará de calificar las respuestas inmersas en la misma.

Por lo que considera:

**2.1. Protocolo de respuestas**

Deberá permitir conjuntar la información de manera rápida y precisa y de igual forma, su pronta captura para la base de datos. Comprende los datos de identificación de los examinados involucrados en el pilotaje (ver *Elementos de la cabeza (1.1.2. Personales)*), y el área para sus respuestas, en cuanto a éste último, una de las formas más sencillas se presenta a continuación:

1 ( )	21 ( )	41 ( )	61 ( )	81 ( )
2 ( )	22 ( )	42 ( )	62 ( )	82 ( )
3 ( )	23 ( )	43 ( )	63 ( )	83 ( )

4 ( )	24 ( )	44 ( )	64 ( )	84 ( )
5 ( )	25 ( )	45 ( )	65 ( )	85 ( )
6 ( )	26 ( )	46 ( )	66 ( )	86 ( )
7 ( )	27 ( )	47 ( )	67 ( )	87 ( )
8 ( )	28 ( )	48 ( )	68 ( )	88 ( )
9 ( )	29 ( )	49 ( )	69 ( )	89 ( )
10 ( )	30 ( )	50 ( )	70 ( )	90 ( )
11 ( )	31 ( )	51 ( )	71 ( )	91 ( )
12 ( )	32 ( )	52 ( )	72 ( )	92 ( )
13 ( )	33 ( )	53 ( )	73 ( )	93 ( )
14 ( )	34 ( )	54 ( )	74 ( )	94 ( )
15 ( )	35 ( )	55 ( )	75 ( )	95 ( )
16 ( )	36 ( )	56 ( )	76 ( )	96 ( )
17 ( )	37 ( )	57 ( )	77 ( )	97 ( )
18 ( )	38 ( )	58 ( )	78 ( )	98 ( )
19 ( )	39 ( )	59 ( )	79 ( )	99 ( )
20 ( )	40 ( )	60 ( )	80 ( )	100 ( )

## 2.2. Cuaderno de reactivos

Se conforma con todas las preguntas o enunciados que forman parte del instrumento, y que por su estructura se pueden presentar como un *cuaderno independiente*, lo cual a su vez, facilita la aplicación del instrumento y la elaboración de una base de datos, ejemplo:

- 
- 1 Su principal utilidad es encontrar información que distinga a los grupos de trabajo por cualidades que no son susceptibles de ser medidas  
a) Nivel Nominal      p) Nivel Ordinal      w) Nivel de Razón      x) Nivel Intervalar
  - 2 Básicamente, permite establecer límites entre las características de un grupo, señalando límites entre lo aceptable o no, entre lo alto y lo bajo, etc. sin importar que tanta sea la diferencia entre los individuos que entren en cada segmento  
a) Nivel Nominal      p) Nivel Ordinal      w) Nivel de Razón      x) Nivel Intervalar
  - 3 El proceso de medición conlleva el poder interpretarla y dar un sentido de bajo o alto, bueno o malo, etc. con base a las reglas previamente establecidas y parámetros de comparación.  
a) Evaluación      p) Medición      w) Constructo      x) Calificación
  - 4 Se puede entender como la descripción de datos en términos de números  
a) Evaluación      p) Calificación      w) Medición      x) Constructo
  - 5 Asigna a las personas a categorías cualitativamente distintas, y pretende determinar si existen dos personas miembros de la misma categoría o clase  
a) Nivel Nominal      p) Nivel Ordinal      w) Nivel de Razón      x) Nivel Intervalar
  - 6 Asignación de números a objetos o eventos, de acuerdo a reglas explícitas  
a) Constructo      p) Evaluación      w) Calificación      x) Medición

### Ficha Técnica

Es la integración de todos los componentes técnicos del instrumento, derivados del proceso de desarrollo metodológico, así como de sus características teóricas, permitiendo comprender de forma rápida su utilidad y alcance en los procesos de medición psicométrica, considera de manera enunciativo y no limitativa, principalmente las siguiente características:

1) Nombre del instrumento	
2) Nombre del autor o los autores	
3) Objetivo	
4) Sustento teórico	
5) Población a la que se dirige	
6) Características del Instrumento	
a) Clasificación de acuerdo al objetivo:	<ul style="list-style-type: none"> <li>- De ejecución máxima o potencia</li> <li>- De ejecución típica</li> </ul>
b) Tipo de respuesta:	<ul style="list-style-type: none"> <li>- Pruebas objetivas</li> <li>- Pruebas subjetivas</li> </ul>
c) Forma de administración de la prueba:	<ul style="list-style-type: none"> <li>- Individual</li> <li>- Colectiva</li> <li>- Autoadministrada</li> </ul>
d) Libertad de ejecución:	<ul style="list-style-type: none"> <li>- Prueba de poder</li> <li>- Prueba de velocidad</li> <li>- Prueba de libre ejecución</li> </ul>
e) Forma de dar las instrucciones:	<ul style="list-style-type: none"> <li>- Oral</li> <li>- Escrita</li> <li>- A través de ejemplos</li> </ul>
f) Material empleado:	<ul style="list-style-type: none"> <li>- Lápiz y papel</li> <li>- Verbal</li> <li>- Manipulable</li> <li>- Aparatos Especiales</li> </ul>
g) Forma de calificación:	<ul style="list-style-type: none"> <li>- Contando puntos</li> <li>- Interpretación</li> </ul>
h) Límites de edad	<ul style="list-style-type: none"> <li>- Infantes o bebés</li> <li>- Niños preescolares</li> <li>- Niños escolares</li> <li>- Adolescentes</li> <li>- Adultos</li> <li>- Senectos</li> </ul>
i) Modo de expresión de la respuesta	<ul style="list-style-type: none"> <li>- Verbal</li> <li>- Manipulativo</li> <li>- Impreso</li> <li>- Gráfico</li> </ul>
j) Tiempo aproximado:	<ul style="list-style-type: none"> <li>- De aplicación</li> <li>- De calificación</li> </ul>

## Referencias

- Aiken, L. R. (2003). *Tests Psicológicos y Evaluación* (Undécima ed.). México: Pearson Educación.
- Aliaga, J. (s.f.). *Psicometria: Tests Psicométricos, Confiabilidad y Validez*. Recuperado el 20 de Octubre de 2015, de Universidad Nacional Abierta y a Distancia:  
<http://datateca.unad.edu.co/contenidos/401517/1U2LibroEAPAliaga.pdf>
- Aron, A., & Aron, E. N. (2001). *Estadística para Psicología* (Segunda ed.). Argentina: Pearson Education S.A.
- Brown, F. G. (1980). *Principios de la Medición en Psicología y Educación*. México: El Manual Moderno.
- Cohen, R. J., & Swerdlik, M. E. (2001). *Pruebas y Evaluación Psicológica. Introducción a las pruebas y a la medición* (Cuarta ed.). México: McGraw-Hill.
- Coolica, H. (2005). *Métodos de Investigación y Estadística en Psicología* (3a. ed.). México: El Manual Moderno.
- Coolican, H. (1997). *Métodos de Investigación y Estadística en Psicología* (2a. ed.). México: El Manual Moderno.
- Coolican, H. (2005). *Métodos de Investigación y Estadística en Psicología* (3a. ed.). México: El Manual Moderno.
- CRECEDES. (s.f.). *Puntuaciones Z*. Recuperado el 22 de octubre de 2015, de Crecimiento y Desarrollo (CRECEDES): <http://www.galeno.sld.cu/crecedes/articulos/p46.htm>
- Dieterich, H. (1996). *Nueva guía para la Investigación Científica*. México: Planeta Mexicana.
- Dorsch, F. (1981). *Diccionario de Psicología*. España: Herder.
- Galimberti, U. (2002). *Diccionario de Psicología*. México: Siglo XXI.
- García, F. (2004). *El Cuestionario: Recomendaciones metodológicas para el diseño de un cuestionario*. México: Limusa.
- Gomez, R. (10 de Mayo de 2012). *Desempeño del Docente en la Educación Superior*. Obtenido de Universidad Gran Mariscal de Ayacucho: <http://ugma2012.blogspot.mx/2012/05/perfil-del-docente-como-evaluador-por.html>
- Gregory, R. J. (2001). *Evaluación Psicológica. Historia, principios y aplicaciones*. México: Manual Moderno.
- Gutierrez, J. (Ed.). (2007). *La investigación Social del Turismo: Perspectivas y Aplicaciones*. España: Thomson Editores.
- Hernández, R., Fernández, C., & Baptista, P. (2006). *Metodología de la Investigación* (4a ed.). México: McGraw-Hill.

- Hernández, R., Fernández, C., & Baptista, P. (2010). *Metodología de la Investigación* (5a. ed.). México: McGraw-Hill.
- Kaplan, R. M., & Saccuzzo, D. P. (2006). *Pruebas Psicológicas* (Sexta ed.). México: Thomson.
- Kerlinger, F. N. (2002). *Investigación del Comportamiento. Métodos de Investigación en Ciencias Sociales* (4a. ed.). México: McGraw-Hill.
- López, A. (2007). *EL Método en Ciencias del Comportamiento*. México: Universidad Autónoma del Estado de México.
- López, A. (2011). *Estadística Descriptiva en Ciencias del Comportamiento*. México: Universidad Autónoma del Estado de México.
- López, A. (2014). *Estadística Inferencial en Ciencias del Comportamiento*. México: Bonobos.
- Mikulic, I. M. (s.f.). *Construcción y Adaptación de Pruebas Psicológicas*. Recuperado el 4 de Septiembre de 2015, de Universidad de Buenos Aires, Facultad de Psicología:  
[http://23118.psi.uba.ar/academica/carrerasdegrado/psicologia/informacion\\_adicional/obligatorias/059\\_psicometricas1/tecnicas\\_psicometricas/archivos/f2.pdf](http://23118.psi.uba.ar/academica/carrerasdegrado/psicologia/informacion_adicional/obligatorias/059_psicometricas1/tecnicas_psicometricas/archivos/f2.pdf)
- Millman, J. D.-H. (1997). *Manual para la Evaluación del Profesorado*. España: La Muralla.
- Myers, D. (2011). *Psicología* (Novena ed.). España: Médica Panamericana.
- Padua, J. (1996). *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. México: Fondo de Cultura Económica.
- Ritchey, F. J. (2002). *Estadística para las Ciencias Sociales*. México: McGraw-Hill.
- Rueda, M., & Diaz- Barriga, F. (2004). *La Evaluación de la Docencia en la Universidad: perspectiva desde la investigación y la intervención profesional*. México: Plaza y Valdés, S.A. de C.V.
- Siegel, S., & Castellan, N. J. (2005). *Estadística No Paramétrica* (4a. ed.). México: Trillas.

