

PSICOMETRÍA BÁSICA

Zulma Ramos Vega



AREANDINA

Fundación Universitaria del Área Andina

MIEMBRO DE LA RED

ILUMNO

Psicometría Básica
Zulma Ramos Vega
Bogotá D.C.

Fundación Universitaria del Área Andina. 2018

Catalogación en la fuente Fundación Universitaria del Área Andina (Bogotá).

PSICOMETRÍA BÁSICA

© Fundación Universitaria del Área Andina. Bogotá, septiembre de 2018
© Zulma Ramos Vega

ISBN (impreso): **978-958-5462-79-3**

Fundación Universitaria del Área Andina
Calle 70 No. 12-55, Bogotá, Colombia
Tel: +57 (1) 7424218 Ext. 1231
Correo electrónico: publicaciones@areandina.edu.co

Director editorial: Eduardo Mora Bejarano
Coordinador editorial: Camilo Andrés Cuéllar Mejía
Corrección de estilo y diagramación: Dirección Nacional de Operaciones Virtuales
Conversión de módulos virtuales: Katherine Medina

Todos los derechos reservados. Queda prohibida la reproducción total o parcial de esta obra y su tratamiento o transmisión por cualquier medio o método sin autorización escrita de la Fundación Universitaria del Área Andina y sus autores.

BANDERA INSTITUCIONAL

Pablo Oliveros Marmolejo †
Gustavo Eastman Vélez

Miembros Fundadores

Diego Molano Vega
Presidente del Consejo Superior y Asamblea General

José Leonardo Valencia Molano
Rector Nacional
Representante Legal

Martha Patricia Castellanos Saavedra
Vicerrectora Nacional Académica

Jorge Andrés Rubio Peña
Vicerrector Nacional de Crecimiento y Desarrollo

Tatiana Guzmán Granados
Vicerrectora Nacional de Experiencia Areandina

Edgar Orlando Cote Rojas
Rector – Seccional Pereira

Gelca Patricia Gutiérrez Barranco
Rectora – Sede Valledupar

María Angélica Pacheco Chica
Secretaria General

Eduardo Mora Bejarano
Director Nacional de Investigación

Camilo Andrés Cuéllar Mejía
Subdirector Nacional de Publicaciones

PSICOMETRÍA BÁSICA

Zulma Ramos Vega



AREANDINA

Fundación Universitaria del Área Andina

MIEMBRO DE LA RED

ILUMNO

EJE 1

Introducción	7
Desarrollo Temático	8
Bibliografía	30

EJE 2

Introducción	32
Desarrollo Temático	34
Bibliografía	50

EJE 3

Introducción	52
Desarrollo Temático	53
Bibliografía	77

EJE 4

Introducción	70
Desarrollo Temático	80
Bibliografía	97

PSICOMETRÍA BÁSICA

Zulma Ramos Vega

EJE 1

Conceptualicemos



Historia y antecedentes disciplinarios



Si bien, históricamente ha existido un interés por reconocer las características particulares de los individuos, existen registros de intentos de medición en antiguas dinastías chinas y sendos tratados filosóficos al respecto, no es sino hasta finales del siglo XIX, en Europa, que se inicia la exploración de las características particulares de los individuos de manera científica.



Lectura recomendada

Para profundizar acerca de este desarrollo histórico se puede consultar el capítulo 1 de:

Instrumentos de evaluación psicológica

González, F.

Como cualquier desarrollo científico, la psicometría se estructura alrededor de preguntas; la primera de ellas gira entorno a qué es lo que se debe medir; la segunda se refiere a cómo medirlo y la tercera, con qué medirlo, al encontrarse estas preguntas con hitos de la humanidad, facilitan el auge y desarrollo de esta importante rama de la psicología.

Muchos fueron los proyectos que buscaron las respuestas a estas preguntas y que contribuyeron a la constitución de la psicometría. Desde el interés por identificar las diferencias individuales; el estudio de los procesos psíquicos en el laboratorio como lo hacía Wundt (1879); el establecimiento de tipologías antropométricas, la introducción del concepto de test mental de Cattell (1890), la evaluación de las funciones psíquicas superiores a través de la escala Binet-Simon y los aportes matemáticos de Pearson y Spearman hasta el desarrollo de las técnicas proyectivas para el estudio de la personalidad, inspiradas por las teorías freudianas. Marcaron el auge y el desarrollo de los dos modelos de evaluación psicológica, el modelo proyectivo y el modelo psicométrico, dada la orientación de este módulo nos centraremos en el modelo psicométrico.



Video

La forma en que se estructura históricamente esta rama que aporta al carácter científico de la psicología, se puede comprender con la ayuda de la videocápsula.

<https://www.youtube.com/watch?v=r60licHuxXE&index=5&list=PLjjVfXgqEfhgEnQ2-KnTH-E8J3HcAj-pP>

El modelo psicométrico toma su más fuerte impulso con el precedente de la prueba de Binet, y ante la necesidad de clasificar y seleccionar a los mejores soldados para atender la Primera Guerra Mundial (1914-1918), toma fuerza la primera pregunta: ¿qué es lo que se debe medir? Los soldados ya se estaban seleccionando desde criterios de resistencia y salud física, pero esto no era suficiente para garantizar su óptimo desempeño en la

guerra. Una vez aclarado qué medir, era necesario plantear un procedimiento, cómo medir, y por supuesto con qué ejecutar este procedimiento. Desarrollaremos cada una de estas preguntas a continuación, identificando las respuestas y argumentaciones que los psicómetras de la época dieron y cómo estas, sientan las bases de la psicometría empleada en el mundo de hoy.

¿Qué medir?

La pregunta era contundente: si a partir de las características físicas y el estado de salud de un individuo no se lograba predecir con suficiente precisión su desempeño futuro, ¿qué otra cosa hacía falta? ¿qué se necesitaba comprender? La respuesta fue igualmente contundente: su estructura psicológica, pero esta seguía siendo compleja, abstracta, es decir, “difícil” de medir. Era necesario ir al detalle, ser más específicos. La respuesta fue, entonces, medir características específicas de los sujetos, sus rasgos y dimensiones que describan sus funciones intelectuales y/o su personalidad. Siendo así que el objeto de medición de la psicometría es el **atributo psicológico**.



Figura 1.
Fuente: shutterstock/675185620

Teniendo claro que el objeto de medición de la psicometría es el atributo psicológico y no las personas -como lo señaló el profesor José Muñiz en su conferencia del 26 de septiembre de 2013-, se procede a describir, delimitar y argumentar todo aquello que es y no es el atributo desde una fundamentación teórica e hipotética inferida a partir de la observación de las conductas, esto es el **constructo teórico** que sustentará las pruebas.



Constructo teórico

Sistema de proposiciones que se relacionan entre sí formalmente y se relacionan con los fenómenos observables por medio de reglas de correspondencia entre la teoría y la conducta. Ejemplos de ellos son: ansiedad, aptitud matemática, inteligencia, personalidad, etc.



Video

Para complementar vamos a ver la videocápsula sobre la videoconferencia del profesor José Muñiz, llevada a cabo en el Colegio Oficial de Psicólogos de Madrid el jueves 26 de septiembre de 2013 (recomendamos hasta el minuto 12:32).

<https://www.youtube.com/watch?v=LbXw62Q1Tpk>

Entre los atributos psicológicos que se han medido históricamente tenemos, la **personalidad**, entendida como un conjunto de rasgos o características comportamentales más o menos estables en el tiempo. Otro de los atributos que ha ocupado históricamente el quehacer de la psicometría, se refiere a las **aptitudes** o habilidades, contemplando entre ellas también a la **inteligencia**, el logro y el rendimiento, entendidos como la capacidad y estilo de respuesta ante las situaciones cotidianas, escolares y laborales.

Algunos de los constructos teóricos que los explican, son: Inteligencia Fluida - Inteligencia Cristalizada (Catell), Factor General - Factor G- (Spearman), Teoría Multifactorial (Thurstone), Inteligencias Múltiples (Gardner). Las **preferencias**, que agrupan también los intereses, actitudes, motivaciones e incluso los valores. La **autoeficacia percibida** (Bandura) descrita por las creencias sobre las propias capacidades para ejecutar acciones. Los **rasgos complejos** (Akerman y Beier) que agrupan intereses, autopercepciones y rasgos conductuales. De igual forma, en desarrollos contemporáneos se ha desarrollado material psicométrico para explorar atributos que dan cuenta de los diversos trastornos clínicos. En la figura 1, se puede apreciar un resumen de estos atributos, constructos y algunas de las pruebas de mayor uso, construidas para medirlos.



Lectura recomendada

Para profundizar acerca de este desarrollo histórico se puede consultar el capítulo 1 de:

Instrumentos de evaluación psicológica

González, F.

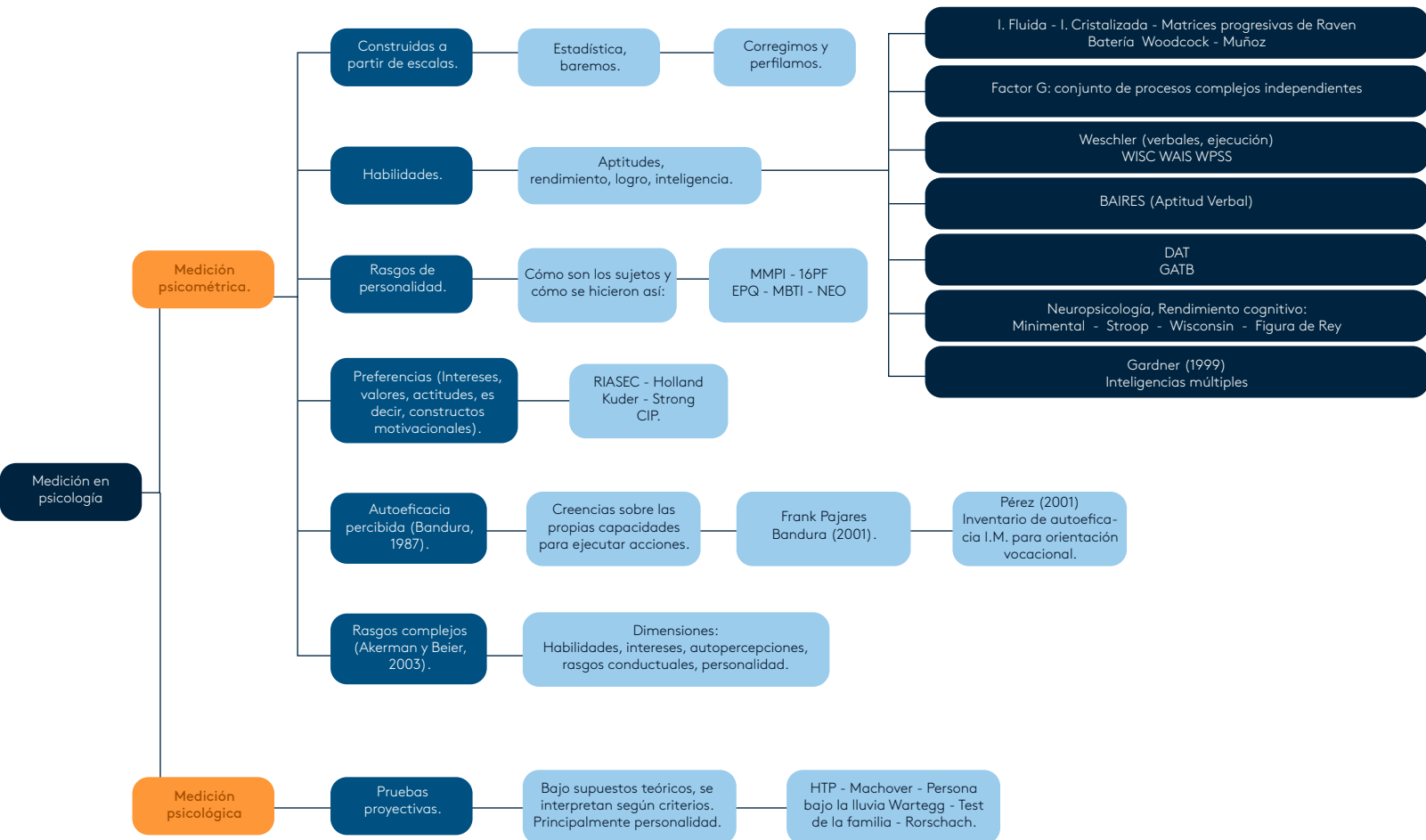




Figura 2. Mapa conceptual constructos y pruebas
Fuente: propia

Para explicar el objeto de estudio de la psicometría y establecer las condiciones y propiedades de las asignaciones numéricas a los atributos, se cuenta con tres **teorías de la medición**, la **Teoría Clásica**, la **Teoría Operacional** y la **Teoría Representacional**.

 Lectura recomendada

Psicometría. Capítulo 1.
Martínez y Hernández.

 Instrucción

Para aclarar su aporte, puede consultar la memonota sobre las teorías de los test.

La teoría clásica, se basa en el modelo lineal propuesto por Spearman en el que las variables se relacionan de forma directa, la puntuación observable de una persona en un test es una función -lineal- de dos componentes, su puntaje verdadero (inobservable, el atributo en tanto tal) y el **error de medición** implícito en la prueba. Desde esta teoría se estiman los posibles errores que se pueden inducir al usar un test para medir variables psicológicas con el objetivo de lograr la exactitud en la medida, es decir, que el resultado de la prueba de cuenta de manera muy cercana a lo que verdaderamente es el atributo.

La teoría operacional por su parte propone un proceso para reconocer el objeto de medición diferenciándolo del entorno construido por la experiencia empírica. Esta teoría busca llevar los constructos teóricos a conceptos observables para definirlos operacionalmente, especificando las actividades u operaciones necesarias para medir la variable. Bridman y Campbell son los principales impulsores de esta teoría.

La teoría representacional, impulsada por Stevens, argumenta que lo esencial son las propiedades de la escala, en donde el componente empírico y el formal se unen de manera comprobada y el papel de los números es representar el atributo, brindando información que permite la interpretación de los datos. Plantea un sistema relacional entre entidades, considerando que las entidades empíricas son aquellas variables identificables y que las entidades abstractas son los números. Usando así los números para representar relaciones de identidad o de diferencia entre los atributos.



Figura 3. Charles Spearman

Fuente: https://es.wikipedia.org/wiki/Charles_Spearman



Error de medición

Se representa y controla a partir de ajustes estadísticos y hace referencia a la influencia que pueda tener cualquier variable sin controlar sobre la puntuación obtenida en un test.



Figura 4.

Fuente: shutterstock/376738720

¿Cómo medir?

Después de dar respuesta a esta primera pregunta, inmediatamente surge una nueva, ya se sabe qué medir, ahora es preciso determinar los procedimientos. ¿Cómo traer ese contenido abstracto del interior de los sujetos a una realidad tangible, convertirlo en observable, es decir, empíricamente comprobable? En efecto, ellos también pensaron que la forma más efectiva era realizar **diseños experimentales**, simulaciones de escenarios en las que los sujetos tuvieran la oportunidad de manifestar a través de conductas, reacciones, actitudes y reportes verbales de sus atributos internos. Sin embargo, desarrollar este tipo de diseños experimentales supone un despliegue económico y temporal demasiado arduo y no permitiría la evaluación de gran cantidad de sujetos o la evaluación simultánea de estos.

Había que pensar en un procedimiento que diera resultados igualmente válidos y confiables, pero con mayor efectividad. Llegando así a plantear descripciones y/o preguntas que permitieran explorar los pensamientos, reacciones, actitudes de los sujetos evaluados en determinadas circunstancias. Sin embargo, era necesario asegurar que esos ítems o **reactivos** estaban contruidos para medir lo que se buscaba medir, es decir, que tuvieran **validez**. Así mismo, que las mediciones realizadas con ellos tuvieran **confiabilidad** al ser usados en momentos y sujetos diferentes.

De esta manera se fueron estableciendo procedimientos sistemáticos para la construcción de los test, así como criterios para la estructura de los ítems y de las respuestas, estos procedimientos y criterios se establecen y explican a través de los **métodos de escalamiento**. **Torgerson** (1958) señala tres métodos: el método centrado en el sujeto (sus atributos psicológicos), el método centrado en el estímulo y el método centrado en la respuesta.



Ítem o reactivo

Unidad básica que compone un test/prueba. Los ítems pueden ser redactados en forma de afirmaciones o preguntas y deben cumplir con criterios y escalas específicas que permitan indagar a través de ellos el atributo.



Validez

Una de las dos principales propiedades de los test, que garantiza estadísticamente que el test mide el atributo para el que fue construido y no otro.



Confiabilidad

Una de las dos principales propiedades de los test, que garantiza estadísticamente que los resultados obtenidos tras la aplicación de un test/prueba son consistentes y estables en el tiempo y en diferentes sujetos.



Métodos de escalamiento

Procedimientos y criterios que determinan la forma en que se construyen las escalas que dan estructura a la prueba.



Instrucción

En la infografía sobre estos métodos se puede profundizar sobre su aporte a la construcción de los test.

El **método centrado en el sujeto** establece la escala para ubicar a los atributos de las personas proporcionando un valor -escalar- para localizarlos en un continuo psicológico y es utilizada habitualmente para medidas de inteligencia. El **método centrado en el estímulo**, puede hacer referencia a objetos físicos, frases que expresan opiniones o actitudes de las personas respecto al atributo a **evaluar**, localizando los "objetos" en valores escalares del continuo psicológico, la persona al responder, cumple la función de "instrumento de medida", ejemplo de ello son las medidas psicofisiológicas de umbrales. El **método centrado en la respuesta**, tiene como objeto de medición tanto a la persona y su atributo como a los estímulos, localizando a los dos en el continuo psicológico, las pruebas de rendimiento y de conocimientos, son ejemplo de este método.



Evaluar

Procedimiento que permite comparar las características de un atributo psicológico, su relación y funcionamiento en comparación con variables externas.

¿Con qué medir?

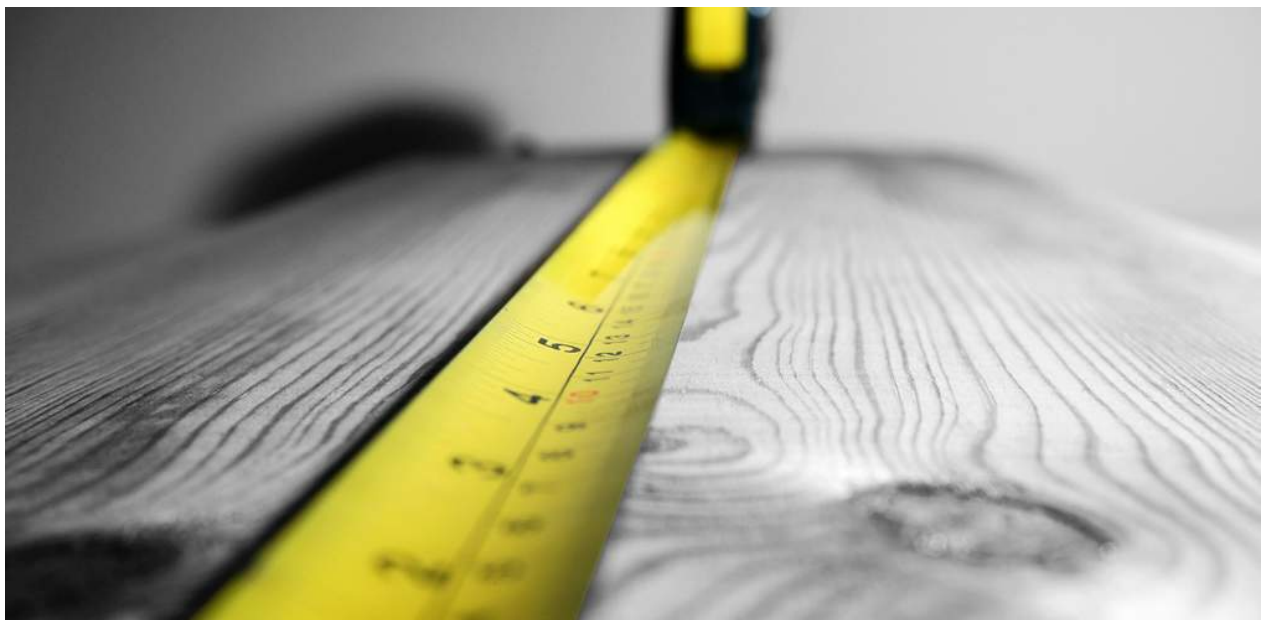


Figura 5.
Fuente: shutterstock/288081992

La tercera pregunta que surge después de aclarar el qué y el cómo, se orienta hacia el instrumento de medida, el test o prueba, herramienta que explora los atributos psicoló-

gicos de manera equiparable a un experimento, de hecho, se considera que una prueba es un arreglo experimental condensado en sus ítems, ya que tanto su construcción, como su aplicación y resultados cuentan con procedimientos sistemáticos y estandarizados.



Video

Una descripción de la forma en que históricamente se llega a esta conclusión se puede apreciar en la siguiente videocápsula.

<https://www.youtube.com/watch?v=wssJH34wDWY&list=PLjjVfXgqEfhgEnQ2-KnTH-E8J3HcAj-pP&index=6>

Teorías de los test




Los procedimientos de construcción de las pruebas están explicados y determinados por las **teorías de los Test**, cuyas propuestas y desarrollo datan de principios del siglo XX. Spearman propone la formulación matemática para estudiar las propiedades de medida de las puntuaciones asignadas por las pruebas para cuantificar el nivel de las personas en funciones psicológicas superiores, como en el caso del test de inteligencia de Binet. En 1950, la Teoría Clásica de los Test es descrita por Gulliksen en el libro *Theory of Mental Tests*, para la década de 1960 aparecen los libros, de Rash y de Lord y Novick, que describen los primeros desarrollos de la Teoría de Respuesta al Ítem, también conocida como teoría moderna, de igual forma, la propuesta de la Teoría de la Generalizabilidad surge como respuesta a las limitaciones y aspectos que no se lograban atender desde la teoría clásica.

Las tres teorías buscan, sin embargo, explicar el significado de las puntuaciones obtenidas por un sujeto tras la aplicación de un test, a continuación, veremos en detalle lo propuesto por cada una de ellas.

Teoría Clásica de los Test – TCT

Sus mayores exponentes fueron Spearman, Thurstone, Thorndike y Gulliksen, sus principales aportes son, la fiabilidad de medición o estimación del error, y la estimación de la validez o el contraste de las **inferencias** hechas a partir de la prueba. Esto se evidencia en su principal postulado, en el que se considera que la puntuación obtenida (observada) por un sujeto en un test, es resultado de su puntaje verdadero, es decir, la manifestación del atributo como tal, y el error de la medición.

 **Inferencia**
Predicción sobre el atributo que se mide con la aplicación de un test.

$$O = V + e$$

Puntación observada	Puntación verdadera	Error
---------------------	---------------------	-------

Figura 6. Principal postulado teoría clásica de los test
Fuente: propia.

Ahora bien, pensemos que los atributos psicológicos por su carácter abstracto, aunque puedan ser medidos, no se podrán conocer completamente, es decir, siempre habrá algo que se escape a la vista o algo que influya en el momento de la medición y por tanto en los resultados, recordemos también que un error en la medición representa variables sin control en el experimento y no precisamente algo que esté mal, en ese orden de ideas, expresado matemáticamente el postulado principal de la TCT, es el siguiente:

En esta teoría se considera que la puntuación verdadera, es el promedio aritmético de puntuaciones empíricas obtenidas en N aplicaciones -recuerde que el test es un experimento condensado- es como si midiéramos mil veces la inteligencia de Juan e hiciéramos el promedio de esos resultados, esa sería su puntuación verdadera, es una estimación, un muestreo de la conducta del sujeto.





Figura 7. Calificaciones con iconografía
Fuente: shutterstock/425190901

Esto supone que las diferencias sistemáticas entre las respuestas se dan debido a la variación en el atributo, las fuentes de error externas se controlan por métodos de **estandarización** o por **aleatoriedad**, para evitar la disminución en la validez externa. Así mismo, las diferencias no se podrán generalizar más allá de la estandarización de su error, esto supone también que las características del sujeto no se podrían separar del test.

Por tanto, el objetivo de quien construye un test bajo estos criterios es, reducir al mínimo el error, ya que cuanto menor sea este, la puntuación observada será más cercana a la puntuación verdadera. Como consecuencia (corolarios) de estos supuestos asociados con el postulado central de la TCT, se tienen seis corolarios:

1. $\int_e = 0$ si $N \rightarrow \infty$ Esto es, la media del error (\int_e) tiende a cero cuando la muestra (N) es grande o tiende a infinito (∞).
2. $\int_v = \int_o$ si $N \rightarrow \infty$ Esto es, la media de la puntuación verdadera (\int_v) es igual a la media de la puntuación observada (\int_o) cuando la muestra (N) es grande o tiende a infinito (∞).
3. En una muestra de sujetos medidos, se asume que:
 - \int_o en distribución normal. La media de las puntuaciones observadas (\int_o) asume una distribución de curva normal (campana de Gauss).
 - Con una probabilidad de .95, los límites de confianza están dados por: $\int_o \pm 1.96 \int$.


Estandarización
Procedimiento que establece las normas, criterios o tipificaciones estadísticas que permiten el uso de un test en diferentes contextos y momentos con características similares a las de su construcción, es decir, permite la generalización de los resultados de sus mediciones.


Aleatoriedad
Tipo de análisis estadístico que permite identificar si un grupo de datos tiene un patrón o se distribuye arbitrariamente.

4. $\sigma_o^2 = \sigma_v^2 + \sigma_e^2$ La varianza de las puntuaciones observadas (σ_o^2) es igual a la varianza de las puntuaciones verdaderas (σ_v^2) más la varianza del error (σ_e^2).
5. La correlación (r) entre puntuaciones observadas de formas paralelas ($r_{xx'}$) es igual a la varianza de las puntuaciones verdaderas (σ_v^2) dividida en la varianza de las puntuaciones observadas (σ_o^2).

$$r_{xx'} = \frac{\sigma_v^2}{\sigma_o^2}$$

$$\sigma_e^2 = \sigma_o^2 (1 - r_{xx'})$$

6. Conociendo el corolario anterior y al reemplazar el término conocido en la ecuación tendremos el último corolario:

$\sigma_e = \sigma_x \sqrt{1 - r_{xx'}}$ Esto es el error estándar de la medición, que surge de la relación entre la desviación estándar observada en el test σ_x (que es igual a σ_o) y el coeficiente de confiabilidad del test ($r_{xx'}$).

A partir de estos corolarios se entiende que la puntuación verdadera de un sujeto es un punto en un intervalo; ante el aumento de la confiabilidad de la prueba, el intervalo disminuye y la puntuación observada está más cercana a la puntuación verdadera; es necesario disminuir los errores de medida para aumentar el coeficiente de confiabilidad.

Varios de estos cálculos se pueden realizar incluso en programas como Excel, por ello, se recomienda empezar a desarrollar la actividad evaluativa, modificar y hacer sus propios ajustes y cálculo, esto facilitará la comprensión de los modelos de las teorías de los test.

Con todo esto, diversos autores han identificado al menos tres aspectos que suponen problemas, limitaciones o aspectos que la TCT no alcanza a responder, y es justo ante estos problemas que surgen propuestas de solución que dan cabida a la teoría de la generalizabilidad y a la teoría de respuesta al ítem.



Figura 8.
Fuente: shutterstock/397626016

La primera dificultad ante la que la TCT se queda corta en dar respuesta, es que un test no puede medir con la misma precisión a todos los sujetos, puesto que las características del test no se pueden separar de las del sujeto y la precisión de la medida dependerá del nivel del sujeto en el atributo medido.

La segunda limitación de la TCT se refiere a la invariabilidad respecto del instrumento utilizado, es decir, teniendo dos test diferentes que miden un mismo atributo, la medición será diferente si se usa uno u otro test, aunque esta medición sea realizada en el mismo sujeto.

Finalmente, la propiedad del test podrá variar según el sujeto evaluado, es decir, un conjunto de ítems del mismo test puede ser altamente difíciles para un sujeto, pero para otro pueden ser sencillos y esto dependerá más de las características de los sujetos que de las propiedades de los test.

Teoría de la Generalizabilidad

Se considera esta como una extensión de la TCT que emerge de las propuestas de Cronbach y Gleser (1972), principalmente los coeficientes de consistencia interna alfa y Kuder Richardson. Desde la TG se afirma que las mediciones de un mismo sujeto pueden variar por efectos sobre el atributo y no solo por el error de medición. Por tanto, el análisis y la garantía de las propiedades de un test se logran desde el estudio de las fuentes y tipos de error, estudio que se realiza usando el análisis de varianza (Anova) de Fisher.

Fisher consideraba que el concepto de confiabilidad de la TCT era limitado porque no tiene en cuenta todas las fuentes de error ni de donde surgen, esto implica

que la variación se puede dar más allá de la construcción del test y trascender al momento de la aplicación, entonces, cualquier variación en las condiciones de aplicación es fuente de varianza y puede afectar el puntaje obtenido por el sujeto. Esto ilustra la principal diferencia entre la TCT y la TG, para la segunda, existen universos de generalización, es decir, hay más de una clase de varianza de error; por ello es necesario definir el universo para identificar las fuentes de variación que producen este error.

Las fuentes de error se conocen como facetas, entendidas como las características del momento de medición que pueden cambiar y afectar los resultados de la prueba. Ejemplo de una faceta serían los ítems que componen la prueba. Ahora bien, en el Anova las **facetas** son los factores y sus efectos, mientras que las **interacciones** son las combinaciones con otras facetas y las **condiciones** son las posibles manifestaciones de una faceta.

El diseño más simple propuesto por esta teoría es el diseño de una faceta que cuenta con cuatro fuentes de variación, en un test de rendimiento estas fuentes serían:

- Las diferencias individuales de los sujetos.
- La dificultad de los ítems.
- Las interacciones entre la dificultad de los ítems y las diferencias individuales.
- Los errores aleatorios o fuentes de error no identificados.

Si se incluyera la hora de aplicación del test, por ejemplo, habría más fuentes de variación y un factor más de interacción:

- Las diferencias individuales de los sujetos.
- La dificultad de los ítems.
- La fecha del test.
- Las interacciones entre la dificultad de los ítems, las diferencias individuales y la fecha del test.
- Los errores aleatorios o fuentes de error no identificados.

Por lo general, los datos se organizan en un diseño factorial cruzado de Anova para separar las fuentes de variación. Estas fuentes de variación son las que contribuyen a la varianza del error y disminuyen la generalizabilidad de los puntajes observados. Para finalizar, en esta teoría se considera el **universo de condiciones de medición** como el conjunto de todas las facetas estudiadas, el total de los ítems que miden el constructo, se denomina **universo y población**, las personas de las que se extrae la **muestra** para la medición.

Teoría de Respuesta al ÍTEM

La TRI aporta una fundamentación probabilística del medir los constructos latentes, considera al ítem como la unidad básica de la medición y se plantea como respuesta al problema de la TCT para separar las propiedades de los test de las propiedades de los sujetos evaluado. Ante los procedimientos para controlar el error, la TRI, elige el ajuste estadístico sobre la estandarización y la aleatorización. En la TRI, el nivel del atributo se estima a par-

tir del patrón de respuesta y se denomina aptitud o rasgo psicológico, identificándolo con la letra θ en la TCT es el puntaje verdadero, a diferencia de la TCT, la relación entre el valor esperado y el rasgo puede ser una función de tipo exponencial.

Esto se traduce en funciones matemáticas que relacionan las probabilidades de una respuesta particular a un ítem con la aptitud general del sujeto bajo modelos estadísticos como Poisson, Ojiva Normal, Error Binomial, Rash o modelos logísticos de 1,2 o 3 parámetros, debido a los complejos cálculos que implican estos modelos, por lo general se emplean programas computarizados como Bigstep, Logistic, Bilog, por mencionar algunos. En su argumentación, la TRI presenta tres postulados básicos:

1. El resultado de un examinado en un ítem, puede ser explicado por un conjunto de factores, es decir rasgos latentes o aptitudes (Torninbeni, 2004, p. 32).
2. La relación entre la respuesta de un sujeto a un ítem y el rasgo latente puede describirse como una función monótonica creciente (Curva Característica del Ítem - CCI). Si aumenta la aptitud, la probabilidad de una respuesta correcta al ítem aumenta (Torninbeni, 2004, p. 32, 33).
3. Las estimaciones del rasgo latente obtenidas con diferentes ítems serán iguales, y las estimaciones de los parámetros de los ítems obtenidas en distintas muestras de examinados serán iguales. Es decir, los parámetros de rasgo e ítem, no varían (Torninbeni, 2004, p. 33).

Así mismo la TRI presenta dos postulados fundamentales:

1. La **unidimensionalidad** de rasgo latente, es decir que los ítems de un test deben medir solo un rasgo.
2. La **independencia**, las respuestas de un examinado a los ítems, son independientes, no existe relación entre las respuestas a uno u otro ítem. Es decir, las aptitudes especificadas en el modelo son los mismos factores que influyen sobre las respuestas a los ítems del test. Así pues, que la probabilidad de dar una respuesta a un conjunto de ítems es igual a producto de las probabilidades de respuesta asociadas con cada ítem. Si la probabilidad de que responda correctamente a uno de los ítems es de 0.5, entonces la probabilidad de que responda correctamente a los dos es de 0.25 ya que: $0.5 \times 0.5 = 0.25$.

A continuación, presentaremos los modelos estadísticos más utilizados:

En el Modelo de Rasch o logístico de un parámetro, la distribución logística se define como la siguiente función:

$$y = \frac{e^x}{1 + e^x}$$

La Curva Característica del Ítem – CCI está dada por la siguiente ecuación:

$$P_i\theta = \frac{e^{(\theta - b)}}{1 + e^{(\theta - b)}}$$

- $P_i\theta$ es la probabilidad de responder correctamente al ítem., esto se calcula para cada ítem por separado.
- b es el parámetro de dificultad del ítem.
- e es la base de logaritmos neperianos (2,718).

La función entonces, forma una curva en forma de S, con valores en el eje Y que oscilan entre 0 y 1 representando a probabilidad de respuesta y en el eje X valores que representan la aptitud. El **parámetro b de dificultad**, es el punto en la escala de la aptitud con probabilidad de respuesta correcta de 0.5, cuanto mayor sea b , más aptitud se requiere para responder correctamente. Los valores de b oscilan entre -2 y +2, así un ítem muy fácil se acercará a -2 y uno muy difícil a +2, un ejemplo de ello se puede apreciar en la figura 3.

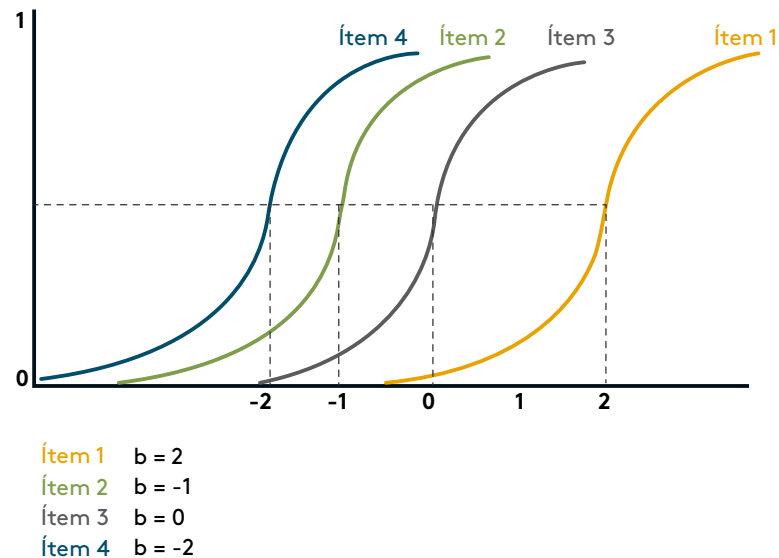


Figura 9. Ejemplo de la gráfica de función para modelo de un parámetro
Fuente: propia

El modelo de dos parámetros se basa en una distribución normal, Birbaum lo sustituyó por una función logística y está dado por la siguiente ecuación:

$$P_i(\theta) = \frac{e^{Da(\theta-b)}}{1+e^{Da(\theta-b)}}$$

b es el parámetro de dificultad del ítem.

D es un factor constante de valor arbitrario de 1,7 para ajustar la función logística a la ojiva normal con exactitud de 0,01.

a es el parámetro de discriminación, es la pendiente de la curva en el punto b.

Los ítems con pendiente mayor, son más útiles para separar a los examinados en diferentes niveles de aptitud que los de menor pendiente, al graficar la función es posible que las curvas se crucen.

El modelo de tres parámetros, introduce el parámetro de pseudo-azar que representa la probabilidad de que, ante un ítem de selección múltiple, un sujeto conteste correctamente sin tener la aptitud, es decir, adivinando. A continuación, se presenta la ecuación:

$$P(\theta) = c + (1-c) \frac{e^{Da(\theta-b)}}{1+e^{Da(\theta-b)}}$$

- b es el parámetro de dificultad del ítem.
- a es el parámetro de **discriminación**, es la pendiente de la curva en el punto b.
- c es el parámetro de pseudo-azar.



Discriminación

Poder, capacidad o propiedad de un ítem, que permite identificar y clasificar el nivel de manifestación del atributo de un sujeto con precisión.

Los parámetros a y b son invariantes ante la medición de un grupo de sujetos, es decir la dificultad y discriminación son propiedad del ítem y, por ende, del test mismo, no dependen de las características de los sujetos evaluados.

A este proceso de estimación de los parámetros se denomina **calibración de los ítems**, el método de estimación de los parámetros, se conoce como **método de máxima verosimilitud**, otra importante característica de los modelos de la TRI, es que exigen muestras grandes de sujetos, mayores a 300, para muestras menores se recomienda el uso del modelo de Rash (un parámetro).

El ajuste de estos modelos cumple con un cometido claro, obtener información de los test, a esto es a lo que se le conoce como **funciones de información**. La primera de ellas, el mejor test es aquel que proporciona la mayor cantidad de información acerca del rasgo o aptitud latente que se está midiendo. Así mismo, el valor de la función de información dependerá de la pendiente de la curva y del error estándar de medición, cuanto mayor es la pendiente y menor es la varianza de un ítem, mayor es la información, esta para un nivel de aptitud es directamente proporcional al poder discriminante del ítem, finalmente al comparar los parámetros que describen las CCI, se obtiene el funcionamiento Diferencial del Ítem (DIF).



Instrucción

Para ayudar en la comprensión de los postulados fundamentales de cada una de las teorías y los conceptos abordados en este eje, se recomienda realizar las actividades de aprendizaje 1 y 2.



Lectura recomendada

Ahora bien, veamos cuál es el procedimiento para crear una prueba. Para ello empezaremos por la lectura de

Psicometría Capítulo 1

Meneses.

Ahora bien, lo primero que se debe tener claro es el **atributo** que se quiere medir, posterior a ello se debe delimitar el **constructo** teórico bajo el que se argumentará la medición de ese atributo, una vez definido esto, se deberá **tipificar** la población y la muestra. Con esta información estructurada, se guía la construcción del universo de ítems que indagan el atributo en ese tipo de población, por lo general se construye un número mayor de ítems a los que compondrán el test. Los ítems se someterán a procedimientos que verifiquen su **validez** y **confiabilidad**.



Tipificar

Describir y delimitar las características de la población y muestra de sujetos a medir.

Uno de estos procedimientos es la **validación por jueces** que consiste en consultar con expertos en construcción de pruebas y/o en el atributo a medir, a partir del análisis de los jueces, los ítems serán ajustados y seleccionados para construir una prueba preliminar con la que se hará una aplicación piloto.

La prueba piloto consiste en aplicar una versión preliminar del instrumento a una muestra representativa que se elige de acuerdo con métodos de muestreo probabilísticos, los resultados obtenidos se sistematizan en programas como SPSS, r, o incluso Excel, a partir de esto se pueden realizar procedimientos estadísticos que permitan analizar la confiabilidad y validez de los ítems, eligiendo para la versión final del instrumento, los que mejores indicadores arrojen.

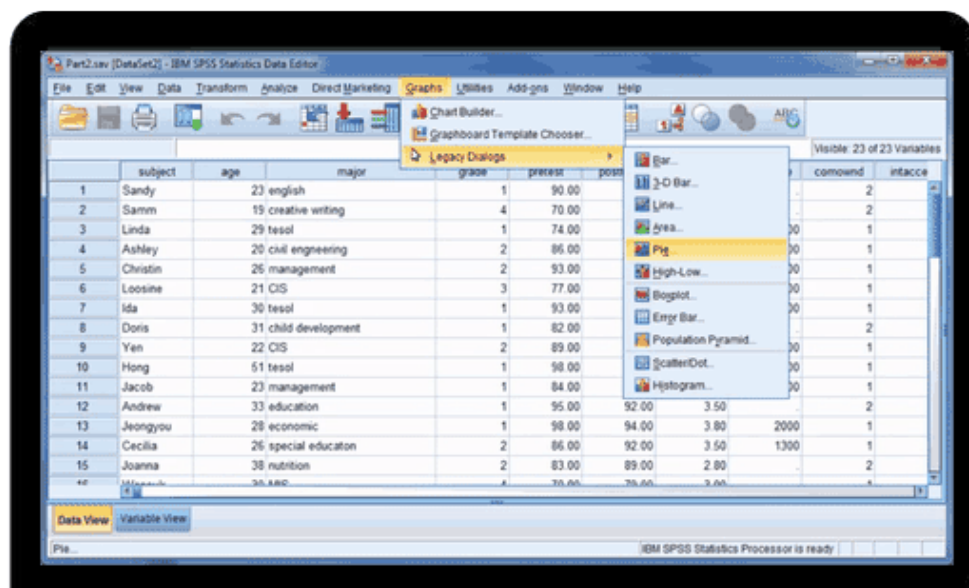


Figura 10. Paquete estadístico SPSS.

Fuente: https://www.google.com.co/search?q=spss&source=lnms&tbm=isch&sa=X&ved=0ahUKewiet-tHWkP3WAhWLSCYKHbu7CVkQ_AUICigB&biw=1366&bih=637#imgrc=0MTuItKO02s-tM:

Una vez lista la versión definitiva del instrumento se procede a realizar el proceso de estandarización. Tomando de la población una muestra representativa (si se hizo aplicación piloto, estas muestras deben ser diferentes), se procede a aplicar la prueba, los resultados obtenidos se sistematizan y se procede a análisis estadísticos que permiten crear las tablas de **bare-mación o normas**, que son las que permiten la generalización de los resultados, es decir, que un instrumento se pueda aplicar a otras poblaciones con características similares. Los baremos además son los que dan la pauta para **corregir** y **perfilar**, la prueba. Veámoslo con un ejemplo:

Se construirá un test que mida inteligencia argumentada desde el constructo de las Inteligencias Múltiples propuesto por Howard Gardner, el instrumento estará dirigido a estudiantes universitarios de Bogotá y queremos que sea un instrumento breve, así que tendrá en su versión final, 50 ítems. Tomando una muestra representativa de la población, ¡vaya que es grande! Se realiza la aplicación y se sistematizan las puntuaciones obtenidas. A la puntuación que obtiene un sujeto tras responder el test se le conoce como **Puntuación Directa (PD)** o puntuación bruta y en realidad esta no nos indica nada, solo que el evaluado respondió una cantidad determinada de ítems.



Corregir

Procedimiento en el que se transforman las puntuaciones directas obtenidas tras la aplicación de un test a puntuaciones baremadas o transformadas.



Perfilear

Procedimiento que se realiza tras la aplicación de un test y posterior a la corrección, en el que se grafican y describen (cuantitativa y cualitativa) las correlaciones de los resultados obtenidos.

Vamos a ver el caso de tres sujetos que respondieron nuestra prueba:

- Sujeto 1. Hombre de 24 años, estudiante de la Universidad Javeriana, noveno semestre de odontología, obtuvo una PD de 40.
- Sujeto 2. Mujer de 20 años, estudiante de la Universidad Nacional, tercer semestre de antropología, obtuvo una PD de 40.
- Sujeto 3. Mujer de 30 años, estudiante de la Fundación Universitaria del Área Andina, en jornada nocturna, quinto semestre de psicología, obtuvo una PD de 40.

Los tres sujetos obtuvieron la misma puntuación, sin embargo, ¿significará lo mismo? Es posible que, aunque la puntuación sea la misma, no haya sido frente a los mismos ítem, o que las diferentes variables mencionadas influyan en el resultado, es allí cuando se hace necesario comparar los resultados con su grupo de referencia.

Lo que se hace en la construcción de la prueba es tomar las puntuaciones obtenidas por las personas que cumplen con cada categoría que se necesita contrastar, por ejemplo, todos los estudiantes hombres con edades entre 20 y 24 años, estudiantes de universidades privadas, de últimos semestres de carreras de áreas de la salud, se realizan análisis estadísticos descriptivos y se transforman estas puntuaciones, obteniendo una tabla de baremación como la del siguiente ejemplo hipotético, en el que los valores están asignados de manera arbitraria.

Tabla de baremos para hombres con edades entre 20 y 24 años, estudiantes de universidades privadas, de últimos semestres de carreras de áreas de la salud.

PD	Puntuación t
10	2,5
20	2,6
30	2,7
40	2,8
50	3,1
	$\mu = 2,74$
	$ds = 0,23$
	$N = 500$

Tabla 1. Ejemplo hipotético de una tabla de baremos
Fuente: propia

Lo que se debe hacer es buscar en la tabla de baremos correspondiente a las características del evaluado la puntuación transformada a la que equivale su puntuación directa y compararla con su propio grupo normativo, un test construido bajo estos parámetros, es conocido como **test referidos a la norma**.

En la tabla anterior vemos que el N es igual a 500, eso quiere decir que de toda la población evaluada 500 de ellos cumplirían con las características del título de la tabla, el valor de la media es de 2,74 y de la desviación estándar es de 0,23. La puntuación transformada del evaluado es de 2,8, asumiendo una distribución normal tendremos que su puntaje se encuentra dentro del área de confianza, es decir en los parámetros esperados para las personas con esas mismas características y que está arriba del promedio.

Ahora bien, en una prueba real, encontraremos al final del manual de prueba las normas o tablas de baremación que permitirán hacer la corrección de puntajes y la perfilación, es aquí en donde el psicólogo tiene su mayor aporte, ya que el análisis de estos resultados, emite un concepto argumentado de cómo se manifiesta el atributo del sujeto, cómo se relaciona con su contexto y que se debería hacer para que estas interacciones sean más funcionales, es decir, orientar sus acciones de intervención, a todo esto es a lo que se le conoce como **diagnóstico**.



Para profundizar acerca de los usos de las pruebas y cómo controlar su calidad y vigencias, es recomendable visitar la página web de la International Test Commission, en especial las guías (guidelines).

<https://www.intestcom.org/page/5>

Explorar las guías, principalmente las de uso de los test: https://www.intestcom.org/files/guideline_test_use.pdf

Y las de uso de test revisados, obsoletos: https://www.intestcom.org/files/guideline_test_disposal.pdf

Clasificación de los test

Históricamente se ha construido una gran variedad de pruebas y en un amplio número, por ello es necesario organizarlos y clasificarlos de alguna manera, para facilitar el acceso a ellos y orientar la decisión del psicólogo respecto a qué tipo de test se debe aplicar a un sujeto en determinada situación. A continuación, mencionaremos algunas de las más comunes.

- Según la finalidad de la medición, se puede decir que las pruebas apoyan la investigación, el diagnóstico, la clasificación y la selección.
- Según el atributo medido los test podrían agruparse por habilidades generales, inteligencia, personalidad, intereses, valores, motivaciones, pruebas psicomotoras y de rendimiento.
- De acuerdo con el objetivo de la medición, tenemos los test de ejecución máxima, que exploran el máximo rendimiento en las tareas de la prueba (test habilidades) y de ejecución típica, que aportan una clasificación tipológica o cualitativa del atributo (test personalidad).
- De acuerdo con la forma de administración del test, tenemos que puede ser individual, colectiva o de autoaplicación, en esta última bajo instrucciones específicas, el sujeto puede responder solo el test en ausencia del psicólogo y, posteriormente, retroalimentar los resultados.
- Según la forma de presentar las instrucciones o consignas para responder la prueba, se encuentran las ins-

trucciones escritas y las consignas verbales.

- De acuerdo con la libertad de ejecución, se tienen los test de poder, en los que el sujeto debe usar su capacidad al máximo, existe un tiempo estimado que en ocasiones se controla; los test de velocidad, que exigen responder lo más rápido posible, en ellos el tiempo es límite y se controla rigurosamente.
- Según el material empleado, tenemos pruebas de lápiz y papel, verbales, de ejecución, sistematizadas y mixtas.
- De acuerdo con la población a medir, se tiene: bebés, niños en edad preescolar y niños en edad escolar, adolescentes, adultos y los aplicables a dos o más etapas del desarrollo.
- Según la forma de corrección que puede ser manual o sistematizada.
- También conocemos como pruebas de **rendimiento óptimo** a aquellas que buscan medir la máxima ejecución del sujeto frente a los ítems descritas a través de los ítems, por otra parte, las pruebas de **rendimiento típico**, se enfocan en la medición de la manifestación completa del atributo, es decir, permiten conocer todas las características y manifestaciones del atributo en tanto tal.
- Por lo general, las pruebas de **rendimiento óptimo** se aplican para atributos como inteligencia, aptitudes, competencias, suelen estar construidas por ítems cuyas **escalas de respuesta** son **binarias** (sí-no, ver-

dadero-falso), de **elección múltiple** (una afirmación o pregunta, entre 3 y 5 alternativas de respuesta) o de **emparejamiento** (dos columnas de términos que se deben relacionar).

- Por su parte, las pruebas de **rendimiento típico** se aplican para atributos como la personalidad o la actitud, emplean por lo general ítems con **escalas de respuesta binaria** del tipo acuerdo-desacuerdo; las **categorías ordenadas**, en las que se presenta una afirmación ante la que el sujeto debe señalar su posición en un continuo que va desde el extremo más favorable hasta el menos favorable, pasando por la neutralidad (totalmente de acuerdo, de acuerdo, ni acuerdo ni desacuerdo, en desacuerdo totalmente en desacuerdo; también pueden usarse términos para determinar frecuencia -bastante/poco-); los **adjetivos bipolares** o **diferencial semántico**, en el que se presentan adjetivos opuestos frente a algún aspecto que el evaluado debe valorar (alegre-triste, cómodo-incómodo).
- Hemos mencionado a las pruebas **referidas a la norma**, aquellas cuyos resultados se comparan con los del grupo normativo, de otro lado, en las pruebas **referidas al criterio**, los resultados se comparan con un criterio determinado de manera arbitraria, identificando qué tan cercanos o lejanos están de este.

Aiken, L. (2003). *Test psicológicos y evaluación*. 11ª edición. Ciudad de México, México: Pearson Educación.

Anastasi, A, y Urbina, S. (1998). *Test psicológicos*. 7ª edición. Ciudad de México, México: Prentice Hall.

González, L., F.M. (2007). *Instrumentos de evaluación psicológica*. La Habana, Cuba: Ciencias Médicas. Recuperado de http://newpsi.bvs-psi.org.br/ebooks2010/en/Acervo_files/InstrumentosEvaluacionPsicologica.pdf

International Test Commission. (2005). International Guidelines on Test Adaptation. [www.intestcom.org]

International Test Commission. (2014). ITC Statement on the Use of Tests and Other Assessment Instruments for Research Purposes. V1.2 [www.intestcom.org]

International Test Commission. (2015). International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal.

International Test Commission. (2001). International Guidelines for Test Use. *International Journal of Testing*, 1(2), 93-114.

Martínez, A. M. R., y Hernández, L. M. V. (2014). *Psicometría*. Madrid, España: Difusora Larousse - Alianza Editorial.

Meneses, J. (2013). *Psicometría*. Capítulo 1. Barcelona, España: Editorial UOC.

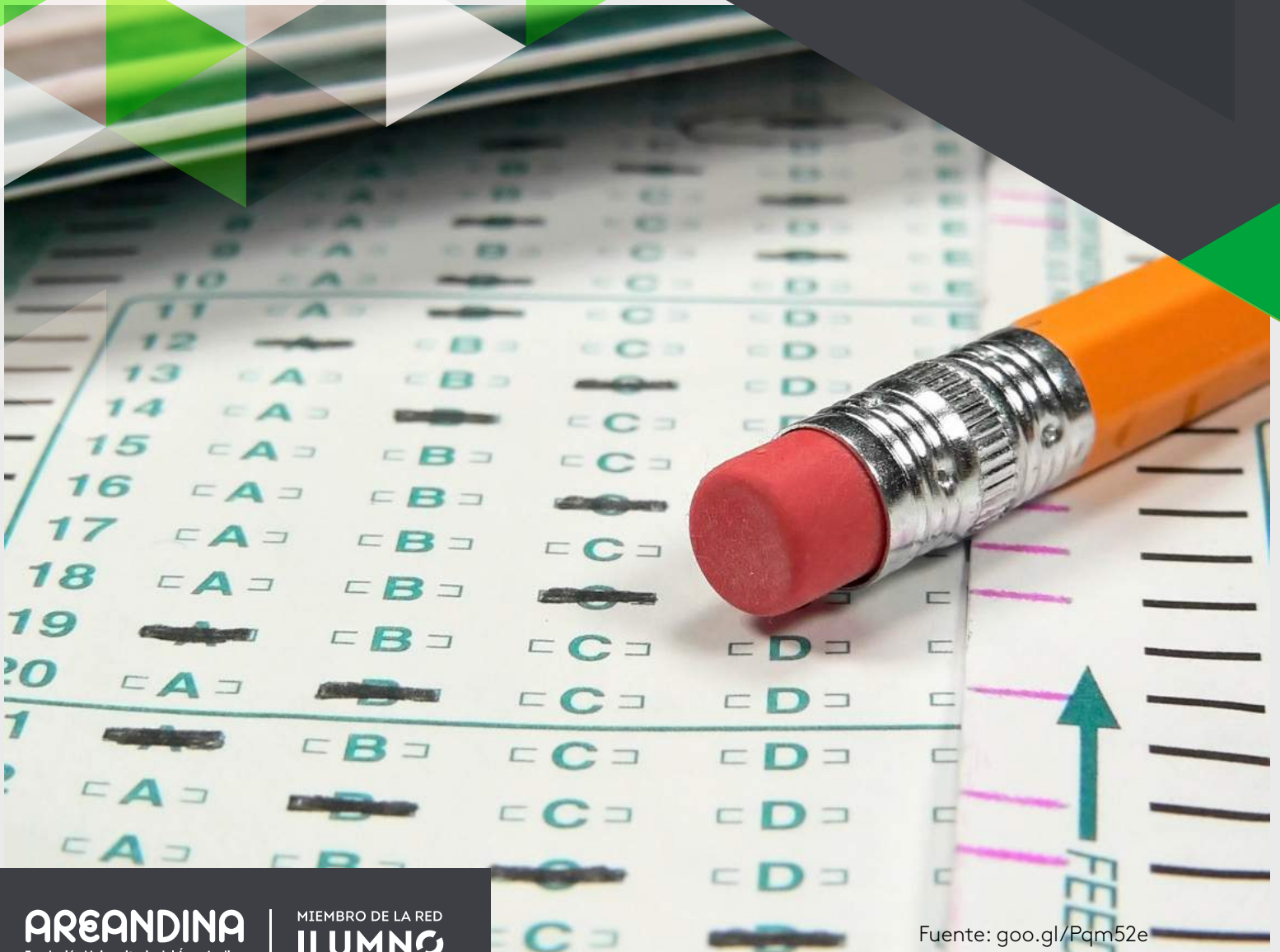
Tornimbeni S., et ál. (2004). *Introducción a los test psicológicos*. Buenos Aires, Argentina: Brujas.

PSICOMETRÍA BÁSICA

Zulma Ramos Vega

EJE 2

Analicemos la situación



Bienvenido(a)

Luego del recorrido histórico y conceptual, nos adentraremos en la funcionalidad de la psicometría, exploraremos los contextos y los retos que tiene para los profesionales del futuro.

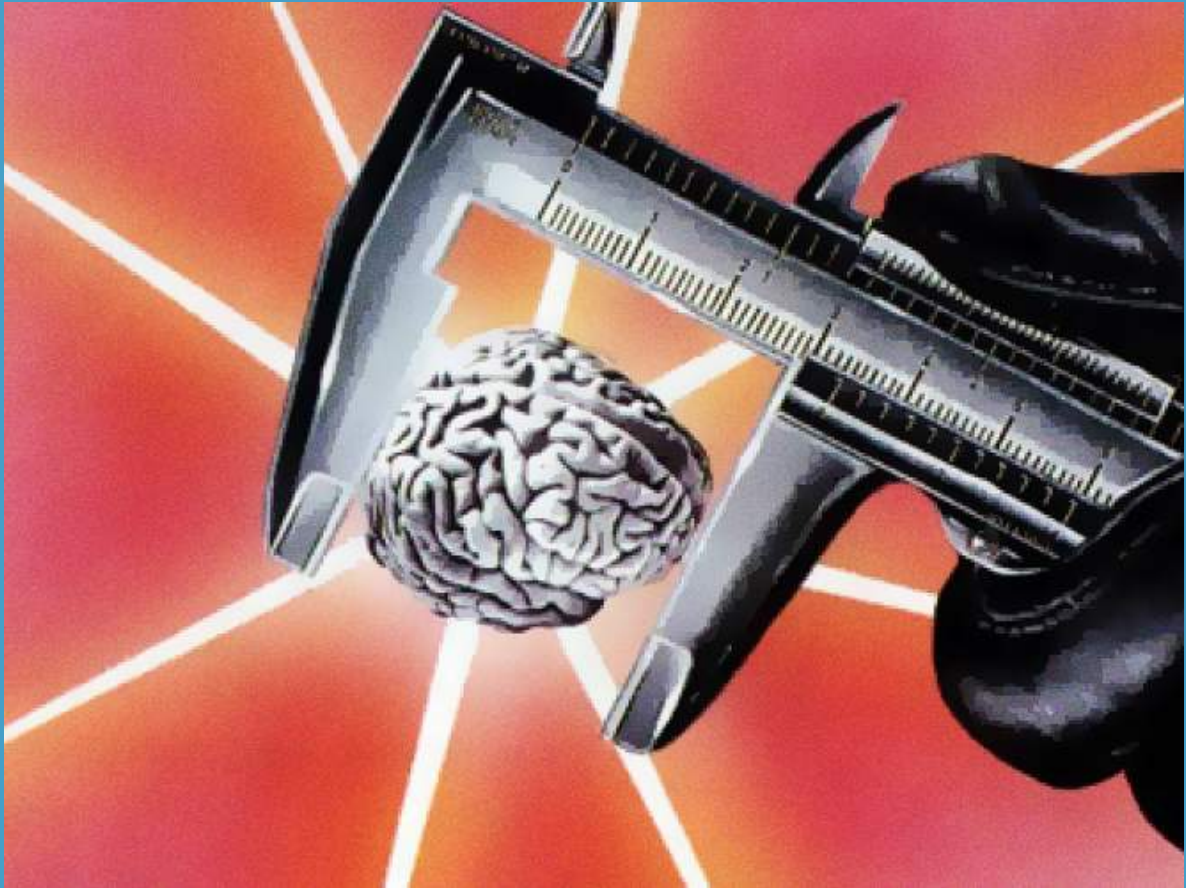


Figura 1.

Fuente: <https://puntmagic.com/blog/psicometria>

Durante estas dos semanas examinaremos los ámbitos de la psicología en relación con las pruebas psicométricas, cuáles son sus aportes y limitaciones, haremos un reconocimiento de la situación actual de la sociedad colombiana respecto a estas pruebas, aportando al desarrollo de nuestras estrategias para afrontar las necesidades actuales en la medición psicológica.

En la primera semana nos dedicaremos a explorar los campos aplicados de la psicología en Colombia, cuál es la labor del psicólogo en cada campo y, a partir de ello, identificar cuáles son los objetivos de la medición psicológica en estos campos.

Los test en los campos aplicados de la psicología



Para dar inicio a este eje, señalaremos las bondades de los test, es decir, cuáles son sus usos y los objetivos de medición y/o evaluación en los que se constituyen como una **herramienta de apoyo** ideal para el trabajo del psicólogo.

Recordemos que el interés de la psicometría y en consonancia con una conducta ética respetuosa de la condición humana, es medir y evaluar atributos psicológicos, no personas. Esto hace que nuestra mirada, nuestro actuar y nuestra forma de comunicar los resultados de la aplicación de una prueba, cambien.

Por ejemplo, no se mide, evalúa o diagnóstica a “la persona débil mental”, como se creía y reportaba en los inicios de la psicometría. Se mide uno de sus muchos atributos -características psicológicas-, la inteligencia, se evalúa la forma en que se presenta en comparación con las manifestaciones argumentadas teóricamente y comprobadas empíricamente, en otros sujetos de la misma edad, escolaridad y cultura (por mencionar solo algunas variables).



Herramienta de apoyo

Las pruebas psicométricas, se constituyen como una más de las múltiples evidencias con las que cuenta el psicólogo para medir, evaluar y diagnosticar la estructura y estado de los atributos psicológicos de un sujeto. Siempre debe estar acompañado de otras técnicas como la entrevista y la observación.



Figura 2.

Fuente: shutterstock/405344767

Se diagnostica, es decir, se emite un concepto profesional, en el cual se describe la forma en que se manifiesta el atributo de esa persona, las posibles causas y las características en el presente (**descripción**) y las probables consecuencias que puede tener en las diferentes áreas de su vida en el futuro (**predicción**). A partir de ello, el psicólogo plantea alternativas de acción, asesorías, toma de decisiones, acompañamiento o entrenamiento para una mejora en la calidad de vida de la persona y su contexto.



Figura 3.

Fuente: shutterstock/526087603

En la actualidad, y siguiendo las tendencias e intereses históricos señalados en el eje anterior, los objetivos de medición y evaluación psicológica que se ejerce sobre los **atributos psicológicos de las personas**, se orientan hacia la **selección** de candidatos a beneficios (becas, auxilios), programas de formación o empleos. La **clasificación** o **ubicación** en diferentes niveles de ejecución o formación escolar y/o laboral. El **asesoramiento** personal, terapéutico, vocacional, laboral y/o profesional. El **apoyo diagnóstico** en situaciones de salud psicológica y/o física para el establecimiento de **planes de acción** o **tratamiento**. **Evaluación** y **seguimiento** a planes de intervención terapéutica o formativa. Apoyo al desarrollo de **investigaciones** en cualquier ámbito en que se requiera.

Aunque el uso de material psicométrico para dar respuesta a los objetivos antes mencionados, es de uso extendido, esto no implica que sea el único elemento que apoya la toma de decisiones y guía el actuar frente a estos asuntos, es importante contrastar estos resultados con registros de autoobservaciones de los sujetos, observación sistemática del psicólogo, verificación de información a través de entrevistas con el sujeto, y dependiendo la situación, con personas que comparten su contexto.



Lectura recomendada

Para profundizar al respecto puede consultar la lectura del capítulo 1 de Aragón (2015).

Ahora bien, estos objetivos se pueden trazar y alcanzar en diversos ámbitos de aplicación de la psicología. De acuerdo con el Colegio Colombiano de Psicólogos, en el país existen 17 campos aplicados reconocidos por esta agremiación, en la tabla 1, se presentan los objetivos de la medición que pueden ser parte de la labor del psicólogo en cada campo aplicado.



Visitar página

Tenga en cuenta: puede consultar la información acerca de esta agremiación en la página web: Colegio Colombiano de Psicólogos.

Campo aplicado	Objetivos de la medición psicométrica
Análisis Experimental de la Conducta.	Desarrollo de investigaciones. Clasificación. Evaluación y seguimiento de intervenciones realizadas.
Desarrollo Humano/Ciclo de Vida.	Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento.
Evaluación, Medición y Estadística Aplicada.	Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones. Selección.
Epistemología e Historia de la Psicología.	Evaluación y seguimiento. Desarrollo de investigaciones.
Neurociencia y Psicobiología.	Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.

Procesos Psicológicos Básicos.	Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones. Selección.
Psicología Clínica.	Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología de la Salud.	Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología de la Sexualidad.	Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología de las Organizaciones y del Trabajo.	Selección. Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología del Consumidor.	Clasificación. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología del Deporte y del Ejercicio.	Selección. Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología Educativa.	Selección. Clasificación. Asesoramiento. Evaluación y seguimiento. Desarrollo de investigaciones.

Psicología Jurídica.	Selección. Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología Militar.	Selección. Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología Social, Ambiental y Comunitaria.	Clasificación. Asesoramiento. Apoyo diagnóstico. Evaluación y seguimiento. Desarrollo de investigaciones.
Psicología del Tránsito y Seguridad Vial.	Clasificación. Asesoramiento. Desarrollo de investigaciones.

Tabla 1. Campos aplicados y objetivos de medición
Fuente: propia



Instrucción

Con el fin de analizar de qué manera las pruebas psicométricas pueden aportar a un mejor desarrollo de la labor del psicólogo en los diferentes campos aplicados de la psicología en Colombia, les invitamos a participar en el foro de debate “Aplicabilidad de los test en los campos aplicados”.

Lo que demuestra que la aplicabilidad y los usos del material psicométrico trasciende el ámbito organizacional, clínico y educativo, siendo los primeros en los que se piensa al discutir estos usos. Es importante, entonces, que el psicólogo en la actualidad reflexione acerca de los procedimientos que puede acompañar en cada uno de los ámbitos, que analice de qué manera las pruebas psicométricas pueden aportar a un mejor desarrollo de su labor, y argumente la decisión de elegir usar material psicométrico y el tipo de prueba elegida.

Los test según el atributo medido

Se presentan a continuación algunos de los test disponibles en nuestro contexto según los atributos medidos. Conocer esta información, le permite al psicólogo **familiarizarse** con las posibilidades que ofrecen las pruebas y **tomar decisiones** respecto a los **objetivos** que espera alcanzar con la **medición** e identificar la **pertinencia** del uso de una u otra prueba en un campo aplicado. En la información subsiguiente, (tabla 2), se encuentran las siglas o nombre abreviado de la prueba, para una descripción específica del test y conocer su **ficha técnica**, puede consultar los catálogos de las editoriales que los comercializan y que se encuentran en el material de recursos para el aprendizaje.



Ficha técnica

Parte inicial de un test que describe datos esenciales para la elección de la prueba a aplicar. Este apartado presenta datos como población a la que va dirigido, tipo de material que lo compone, tiempo y tipo de aplicación.



Visitar página

Tenga en cuenta: puede consultarse en internet los catálogos de las editoriales en estas páginas.

- <https://web.teaediciones.com/Catalogo.aspx>
- <http://pseaconsultores.com/productos.php?closed=1>

Ahora bien, es pertinente recordar que el **psicólogo** es el **único** profesional facultado y autorizado legalmente para **aplicar** material psicométrico con fines de **medición, evaluación y diagnóstico psicológico**, para mayor precisión respecto a estas determinaciones puede consultar la **Ley 1090 de 2006** que reglamenta el ejercicio de la profesión de psicología y dicta el Código Deontológico y Bioético, esto debido al alto grado de sensibilidad de la información que allí se maneja respecto del evaluado.



Figura 4.
Fuente: propia



Lectura recomendada

Lectura complementaria

Ley 1090 de 2006

Por lo cual, los diversos organismos y asociaciones que alrededor del mundo se ocupan tanto de dar lineamientos de uso, como de comercializar las pruebas han generado un sistema de clasificación para el acceso y utilización de los test de acuerdo con la experticia del evaluador y las exigencias de la prueba.

Lo anterior implica que existen algunas pruebas diseñadas para explorar ciertos atributos que no son de exclusivo interés para la psicología, sino que también pueden ser estudiados por otros profesionales como pedagogos, orientadores, psicopedagogos, trabajadores sociales, terapeutas ocupacionales, de lenguaje, neuropsicólogos y psiquiatras desde los respectivos intereses de sus áreas.

De acuerdo con lo anterior, las clasificaciones establecidas para acceso al material, son de tres niveles, y se determinan con las letras A, B y C que se encontrarán señaladas en los catálogos y en ocasiones en las mismas pruebas. De acuerdo con Aiken (2003), estas se describen de la siguiente manera:

Las pruebas
clasificadas

con
A

Admiten el uso por parte de estudiantes de psicología supervisados y acompañados de un profesional especializado y algunos profesionales de otras áreas, acorde con las especificaciones de la prueba. Las exigencias básicas para el manejo de estas pruebas implican conocimientos de la forma de administración de las pruebas desde parámetros éticos.

Las pruebas
clasificadas

con
B

Admiten el uso por parte de psicólogos preferiblemente con niveles de especialización, estudiantes de psicología bajo supervisión y profesionales de área afines a la psicología, eso porque dentro de las exigencias básicas, se encuentra que se deben tener conocimientos en psicometría, en la corrección de pruebas, teorías de la medida, estadística y manejo ético de las mediciones.

Las pruebas
clasificadas

con
C

Son de uso exclusivo para psicólogos especializados, estudiantes de psicología bajo supervisión y profesionales de áreas afines a la psicología con formación de maestría o doctorado en psicología, debido a la complejidad de la aplicación de las pruebas y los datos obtenidos en este procedimiento, quien las aplique, debe tener entrenamiento y experticia en el uso, aplicación, corrección y perfilación de instrumentos estandarizados.

Figura 5.
Fuente: propia

Teniendo en cuenta lo anterior, revisemos la forma en que el psicólogo toma la decisión de aplicar o no material psicométrico para apoyar su labor. Lo primero que debe tener en cuenta es el atributo psicológico que desea medir (**¿qué?**), posterior a ello es necesario definir cuál es el objetivo de la medición (**¿para qué?**), contrastando estos datos, identificando las características particulares del evaluado (edad, escolaridad, estado de salud psicológica y física, otras más que sean pertinente según el caso) y realizando una cuidadosa revisión de las fichas técnicas de las pruebas elegirá el instrumento que más se adecúe para dar respuesta al objetivo de la medición (**¿con qué?**) teniendo en cuenta las particularidades del evaluado.



Lectura recomendada

Para profundizar sobre las consideraciones sobre esta toma de decisiones, consulte el capítulo 1 de Martínez (2014), el apéndice A y el capítulo 1 de Thomas (2015).

ACOSO ESCOLAR Y ADICCIONES	APTITUDES	CREATIVIDAD	ORIENTACIÓN VOCACIONAL	SIMULACIÓN	AUTOESTIMA Y AUTOCONCEPTO	PEDAGÓGICOS Y DE RENDIMIENTO	INTELIGENCIA EMOCIONAL
ADITEC AVE CONVES CYBERBULLYING FRIDA	AEI BAPAE BTI EFAI BAT-7 PMA TEA	CREA PIC-A PIC-J PIC-N TCI	CIPSA EXPLORA IPP-R KUDER-C	Estrategias de detección de la simulación SIMS TOMM PAI PAI-A MMP-2-RF	A-EP AF-5 EMMA LAEA	EPP FROSTIG GOT TEDI MATH TEMA-3 SOCIOMET	AECS CRI-A CTI MOLDES MSCEIT TECA

PSICOPATOLOGÍA Y FORENSE	TRASTORNOS DE ALIMENTACIÓN	APRENDIZAJE Y ESTUDIO	CALIDAD DE VIDA Y BIENESTAR	ADAPTACIÓN Y CONDUCTA	LECTOESCRITURA Y LENGUAJE	PERSONALIDAD	NEUROPSICOLOGÍA
CAQ	EDI-3 El enemigo en el espejo IMAGEN PICTA TSA	ACRA					
CTC		CEA					ANILLAS
EGEP		CHTE					BENDER
ESQUIZO-Q		DIE, 1, 2 Y 3	ABAS-II				BOSTON
LSB-50		IHE	CUBRECAVI			16 PF--APQ	BRIEF
MMPI-2		TAVEC	CESQT			16 PF-5	BRIEF-P
MMPI-2-RF		TAVECI	DERA			ASPA	CAMDEX-DS
MMPI-A		TOMAL	EBP			BFQ-NA	CAMDEX-R
PAI		AUTISMO				CEP	CÓRTEX
PAI-A		ADI-R		ABAS-II		CPQ	CUMANES
PCL-R		ADOS-2		ACE		CPS	CUMANIN
SIMS	SCQ		ACS		CTC	ENFEN	
ANSIEDAD, DEPRESIÓN Y ESTRÉS	IDIOMAS Y MEMORIA	INTELIGENCIA Y DESARROLLO	ATENCIÓN	A-D	DST-J	CUIDA	FDT
				AECS	ECL, 1 y 2	EPQ A-J	HARRIS
					ECONPLEC	EPQ-R	HPL
					EDAF	ESFA	LURIA DNA
CAEPO	ANOMIA	BAS-II	AGL			ESPA 29	LURIA INICIAL
CAS	CÓRTEX	Curriculo Carolina	AULA				MFF-20
CDI	El baúl de los recuerdos	DP-3	BASC				MMSE
	MAI	EOD	CARAS-R				PORTEUS
		EPP	CSAT-R				REGIA
		MATRICES					
		Merrill Palmer-R					

CDS				BAS, 1, 2 Y 3		
CECAD		RIAS		BASC		REY
CET-DE		RIST		CACIA	ITPA	SCIP-S
CRI-A		TOMAL	EDAH	CAPI-A	Logopedia interactiva	SDMT
EAE	MVR	TONI-2	D2	CLIFTON	LOLE	STROOP
EGEP	MY	PRIMARIA-R	DIVISA	EHS	MACARTHUR	TAVEC
IDER	TAEL	EPA-2	FI- R	FRIDA	PEABODY	TAVECI
IECI	TAELIS	FACTOR G, 1, 2 y 3	MFF-20	IAC	PLON-R	TESEN
ISRA	TAVEC	FACTOR G-R	PIAAR-R	PSYMAS	PROESC	TIDA
JSS	TAVECI	MATRICES	TO-1	SENA	PROLEC-R	TRVB
STAI	TOMAL	PRIMARIA-R	TP-R	SIS	PROLEC-SE	WCST
STAIC		TIG		SOC	REGIA	TOMAL
STAXI-2		ABAS-II		SOCIOMET	TALE	
STXI-NA				SPECI		
				TAMADUL		
				TAMAI		

Tabla 2. Test según el atributo psicológico medido
Fuente: propia a partir de la consulta de catálogos de pruebas

Los test y la ética



Como vemos, el material psicométrico es un material altamente complejo y delicado, lo que supone un manejo cuidadoso, responsable y sobre todo ético. Así, las organizaciones o instituciones poseedoras de pruebas psicológicas las deben tener bajo custodia especial, es decir, no todos pueden tener acceso a este material, por lo general, los **manuales** y las **plantillas** de corrección no deben extraerse del lugar, por derechos de autor, el material no debe ser copiado ni fotografiado, es decir, siempre se deben usar las hojas o **protocolos de respuesta** originales.

Debido a la información que tras la aplicación de una prueba, se puede obtener de una persona; este se considera un procedimiento invasivo, aunque no “invadimos” su cuerpo físico, si nos adentramos en su “intimidad”, “invadimos” su cuerpo psicológico, por tanto, para la aplicación de pruebas psicométricas, es necesario realizar un **consentimiento informado**.

El consentimiento informado es, a su vez, un procedimiento y un documento. Como procedimiento, es una conversación previa en la que se le expone al evaluado las razones por las que se va a realizar la medición, el tipo de material a utilizar, qué se va a hacer con la información obtenida tras la aplicación del test, se responden las inquietudes del evaluado y se le aclara que, de no desear continuar con la aplicación, puede hacerlo sin problema.

Como documento, es un formato en el que se consigna de manera breve y clara la información de la conversación, en él se manifiesta explícitamente que el evaluado ha sido informado y que acepta realizar la prueba voluntariamente, este documento debe reposar en el historial o carpeta del evaluado bajo custodia del profesional o la institución a la que representa, es importante que se aclaren también los riesgos de la evaluación y que el evaluado tendrá el reporte de sus resultados.

Una vez aplicado el test, la información obtenida también debe obedecer a criterios éticos, es decir, en primer lugar, la forma en que se comunican los resultados al evaluado y a su familia en algunos casos, debe evitar etiquetas, expresiones patologizantes (p. ej.: “usted es depresivo”) o versiones susceptibles de ser tergiversadas.

Recordemos que se miden y evalúan atributos psicológicos, no personas; así, es diferente decir: “su hijo tiene retardo mental”, “su hijo es retrasado mental”, a decir: “el ritmo de comprensión de la información en su hijo es diferente y requiere más tiempo de trabajo, otras formas de explicación”. El lenguaje utilizado y la asesoría que brindamos los psicólogos es fundamental para evitar un daño mayor a los sujetos, así mismo evita el temor y prejuicio de las personas hacia las pruebas psicológicas.

Manual de prueba

Uno de los elementos más importantes que componen una prueba psicométrica, en él se consigna toda la información referente al test, su sustento teórico y estadístico, las instrucciones de aplicación y corrección de la prueba y los baremos.

Plantillas de corrección

Láminas o cuadernillos que se utilizan para la verificación de las respuestas y para el procedimiento de corrección del test; algunos test cuentan con ellas, otros usan corrección sistematizada, por software especializado, a través de pines on-line.

Protocolo (hoja) de respuesta

Uno de los elementos que componen las pruebas, en él el evaluado diligencia sus datos y consigna las respuestas a los reactivos del test, en algunas pruebas, en esta hoja se incluye la gráfica del perfil del atributo evaluado.

Consentimiento informado

Procedimiento previo a la aplicación de una prueba psicométrica, a la vez es un documento en el que guarda constancia de que el evaluado fue informado del tiempo de test aplicado, los objetivos de la medición y el uso que se le dará a los resultados.

De igual forma, la información obtenida tras la aplicación de un test, es confidencial, esta no debe ser revelada a terceros y los informes en que se consigne esta información, deben permanecer también bajo custodia especial, la forma en que se transmite esta información, cuando se requiere, debe también evitar las etiquetas, tergiversaciones y expresiones patologizantes. Es importante que revise el código deontológico para verificar en qué casos se puede revelar el secreto profesional y a quienes comunicar esta información.

Los test en la actualidad

Somos conscientes de la influencia que el contexto ejerce en el desarrollo de los seres humanos, por eso es importante reconsiderar las características cambiantes de la sociedad, las exigencias y tendencias actuales a las que los seres humanos nos vemos expuestos, esto en conjunto con otros diversos factores que influyen en la manera como se configura nuestro ser psicológico.

En las últimas décadas, se han dado diversos cambios en nuestra sociedad, el desarrollo de las tecnologías de la información y la comunicación, su uso extendido y casi obligado, la importancia que se le ha atribuido a la capacidad adquisitiva y al reconocimiento público, las concepciones que sobre el concepto de ser exitoso, se han creado, los modelos sociales y publicitarios que se nos presentan para ser seguidos.

Lo anterior hace que los intereses y acciones de las personas cambien, así mismo se empiezan a modificar sus necesidades y las situaciones que les generan incomodidad o sufrimiento, es decir, los posibles trastornos que se puedan desencadenar, veamos con detenimiento estas situaciones.



Figura 6.
Fuente: Shutterstock/593149802

En primer lugar, mencionamos el desarrollo de las tecnologías de la información y la comunicación (TIC), esto supone dos importantes retos para la medición psicométrica. El primer reto, está en el uso de estas tecnologías, un uso que puede convertirse en disfuncional o inadecuado, además de desencadenar situaciones problemáticas en la salud física y psicológica de las personas. Por ejemplo, las alteraciones de sueño por el uso de las TIC, la modificación o disminución de la interacción social directa (persona-persona), dependencia o adicción a las TIC, esto solo por mencionar algunas.

El segundo reto, muy delicado, es el libre acceso a la información y el uso indiscriminado e indebido de material psicométrico. En la actualidad, cualquier persona que sepa ejecutar una búsqueda en internet puede “autodiagnosticarse”, acceder a “test” rápidos en internet, incluso se pueden encontrar por internet segmentos de pruebas a manera de “juegos para estimular tu cerebro” o “para conocerte mejor”, “tips para responder pruebas psicológicas”.

Así pues, ¿para qué consultar al psicólogo, si la prueba se encuentra en internet?, esto es una alerta para los psicólogos en la actualidad, es necesario estar al tanto de las nuevas necesidades de las personas para atenderlas con oportunidad y efectividad. Por otro lado, es fundamental posicionar la labor del psicólogo en nuestra sociedad y el verdadero sentido de la medición psicológica.

Así como hacer un uso ético del material psicométrico y generar conciencia en las personas acerca de lo delicado que puede ser creer en todo lo que se encuentra en internet, y de etiquetarse haciendo autodiagnóstico con pruebas de dudosa procedencia.



Figura 7.
Fuente: propia

En este mismo sentido, el uso extendido y casi obligado de las TIC también contribuye al desarrollo de nuevas situaciones psicológicas que deben hacer volcar nuestra mirada hacia allí, prácticas como el **cyberbullying**, temores como la **nomofobia**, se constituyen hoy como nuevos objetos de medición y atención psicológica. Lo que nos lleva a buscar respuesta a dos preguntas: ¿qué material psicométrico se ha desarrollado en los últimos años?, ¿las pruebas psicométricas existentes son suficientes y pertinentes para las necesidades de la actualidad?

Otro aspecto de la actualidad que exige atención y desarrollos por parte de la psicología y de la psicometría, respectivamente, es la importancia que se le ha atribuido a la capacidad adquisitiva y al reconocimiento público, esto ha llevado a que la sociedad se sobreexija en términos de productividad laboral y profesional, no solo en la generación de ingresos económicos que le permitan adquirir bienes y servicios que aporten estatus, sino también a hacer pública su vida privada para que el otro le felicite y reconozca por sus logros. Tenemos como ejemplo de lo anterior el Síndrome de **Burnout**.

Por último, las concepciones sobre el concepto de ser exitoso, los modelos sociales y publicitarios que se presentan para ser seguidos, han generado también modificación en las conductas, prácticas, necesidades y configuración de posibles trastornos psicológicos, así mismo de estrategias para sobrellevar estos trastornos, es importante que el psicólogo empiece a cuestionarse al respecto, se mantenga al tanto y atento a estos cambios y, sobretodo, que trabaje para aportar a la construcción de nuevos conocimientos y herramientas para atenderlos.



Instrucción

Un breve ejemplo de lo aquí expuesto, se muestra en el caso simulado, cuyo video invitamos a ver con perspectiva crítica y reflexionando sobre las preguntas planteadas para ello.

Como se puede apreciar, el panorama presente y futuro de la psicología y en especial de la medición psicométrica supone un sinnúmero de retos para el profesional, aquí se han mencionado los más notorios, no quiere decir que sean los únicos, sin embargo, son aquellos mínimos frente a los que se debe tener respuesta.



Cyberbullying

Ciberacoso en español, hace referencia al uso de TIC para amedrentar, acosar o generar burlas en contra de una o varias personas.

Nomofobia

Temor excesivo a estar sin celular.

Síndrome de Burnout

Tipo de estrés laboral caracterizado por un conjunto de signos y síntomas progresivo de agotamiento físico y psicológico.



Lectura recomendada

Para profundizar en este tema, leer los capítulos 4 y 5 de González (2007).

Cuestionarse acerca del aporte de las pruebas psicométricas en el contexto actual es una importante materia que no debe dejarse desligada de la responsabilidad y conducta ética, para apoyar al descubrimiento de esos retos y al desarrollo de herramientas para responder a estos retos, y a manera de cierre de este eje, se recomienda atender a los videos de la conferencia Presente y futuro de los test psicológicos del profesor José Muñiz.



Video

- <https://www.youtube.com/watch?v=TqnQWGmylks>
- <https://www.youtube.com/watch?v=V2SPLwLokKY>
- <https://www.youtube.com/watch?v=ME5Du7n1vzw>
- <https://www.youtube.com/watch?v=4HjBqwVBKxc>
- <https://www.youtube.com/watch?v=LbXw62Q1Tpk>



Instrucción

Te invitamos a desarrollar la actividad de aprendizaje 2 para la aplicación de los conceptos centrales de este eje.

Aiken, L. (2003). *Test psicológicos y evaluación*. (11ª edición). Ciudad de México, México: Pearson Educación.

Anastasi, A. y Urbina, S. (1998). *Test psicológicos*. (7ª edición). Ciudad de México, México: Prentice Hall.

Aragón, L. (2015). *Evaluación psicológica: historia fundamentos teórico-conceptuales y psicometría*. (2ª. ed.). Ciudad de México, México: Editorial El Manual Moderno.

González, L., F. (2007). *Instrumentos de evaluación psicológica*. La Habana, Cuba: Ciencias Médicas. Capítulo 4 y 5. Recuperado de http://newpsi.bvs-psi.org.br/ebooks2010/en/Acervo_files/InstrumentosEvaluacionPsicologica.pdf

International Test Commission. (2015). *International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal*. Londres, Reino Unido: International Test Commission.

Martínez, A. M. R., y Hernández, L. M. V. (2014). *Psicometría*. Madrid, España: Difusora Larousse - Alianza Editorial.

Meneses, J. (2013). *Psicometría*. Barcelona, España: Editorial UOC.

Thomas, P. H. (2015). *Pruebas psicológicas: una introducción práctica* (2ª. ed.). Ciudad de México, México: Editorial El Manual Moderno.

Tornimbeni, S., et ál. (2004). *Introducción a los test psicológicos*. Buenos Aires, Argentina: Brujas.

PSICOMETRÍA BÁSICA

Zulma Ramos Vega

EJE 3

Pongamos en práctica



Bienvenido(a)

Habiendo hecho el recorrido histórico y conceptual, y el abordaje de la funcionalidad de la psicometría desde la perspectiva analítica de los contextos y los retos de los profesionales del futuro, pondremos en práctica el sustento estadístico en las propiedades de los test y en los procedimientos de corrección y perfilación.

Durante estas dos semanas nos adentraremos en los usos prácticos de la estadística como fundamento de la psicometría y en el correcto uso de las pruebas, teniendo en cuenta sus componentes, el control de variables y las bases para ejecutar los procedimientos estadísticos que dan la garantía y verificación de las propiedades de los test.



En la primera semana, abordaremos una de las propiedades fundamentales de los test, la confiabilidad, cómo garantizarla y verificarla en los diferentes momentos de la medición psicométrica, además, conoceremos las dimensiones de la confiabilidad y los métodos estadísticos para calcularla.

En la segunda semana, exploraremos otra de las dos propiedades de los test: la validez, conoceremos las fuentes de evidencia de validez, los procedimientos estadísticos para ello y su utilidad práctica, así mismo, se explorarán las diferentes puntuaciones transformadas, sus significados y ajustes. Estos dos criterios estadísticos no solamente guían la elección de una prueba sobre otra, sino que, al construir un test, son la garantía de su aporte y valor para la medición psicológica.

A través de los recursos de aprendizaje y las actividades, se aplicarán dichos procedimientos, llevándonos a reflexionar sobre la pregunta central de este eje: ¿cómo construir un test válido y confiable y garantizar una elección acertada a la hora de realizar una medición psicológica eficaz?

Es recomendable que dediquemos unas horas diarias a la lectura del material propuesto para este eje, tomar apuntes, subrayar los aspectos esenciales, desarrollar esquemas, realizar las actividades de refuerzo, esto facilitará la comprensión del tema, consultar el glosario aportará claridad y reforzará la comprensión de los aspectos más importantes del módulo.

¡Adelante!

Vamos a comenzar.

Confiabilidad



Un importante aspecto para maximizar el aprovechamiento de la información que pueda ser obtenida tras la aplicación de un test, es comprender lo que significan e informan las propiedades de este y sus correspondientes valores, en los apartados que se describirán a continuación, se presentarán las propiedades de los test, la confiabilidad y la validez, sus ajustes estadísticos y las puntuaciones transformadas y su uso en la psicometría.



Video

Un canal recomendado en YouTube para reforzar este eje, es: Psicometría y Estadística – proclapar.com

<https://www.youtube.com/channel/UCKYcwxW6XYrw3c66ulXr9rA>



Lectura recomendada

Antes de comenzar, los invitamos a realizar la siguiente lectura.

El error estándar de medida y la puntuación verdadera de los tests psicológicos: algunas recomendaciones prácticas.
Págs. 117 a 123

Gempp, R.

Esta propiedad sustenta la **exactitud** de las mediciones realizadas con un test, así mismo, cuando presenta un nivel significativo, garantiza la **precisión** de la medida, es decir, que la puntuación obtenida y transformada, representa de la manera más cercana posible al atributo medido.

De igual forma, representa la **consistencia** de las mediciones, esto es: asumiendo que, si el test es aplicado en diversas ocasiones en un mismo sujeto o en una población, se obtendrían rangos de puntuaciones estables, coherentes.

Sin embargo, es importante tener en cuenta que la medición psicológica es compleja por cuanto los “objetos” medidos presentan mayor probabilidad de variabilidad, es decir, se debe recordar que más allá del test, las mediciones se ven influenciadas por las diferencias individuales de los sujetos: sus experiencias, motivaciones, cansancio o disposición, ansiedad, estados emocionales y/o fisiológicos, entre otros.

Debido a ello, de aplicarse en ocasiones repetidas un mismo test a un mismo sujeto, no se obtendrían puntuaciones perfectamente consistentes, esto es a lo que llamamos, error de la medición, como lo vimos en los ejes previos, recordemos entonces que cuanto mayor sea la confiabilidad, menor será el error, y es justo esto lo que se busca al construir una prueba, reducir el error, identificando el origen (fuentes) y la magnitud.

Ahora bien, existen varias características del error, importantes para diferenciar. En primer lugar, el **error** siempre es **impredecible** y **aleatorio**, sin embargo, existen también **errores sistemáticos**, aquellos que afectan por igual a todos los elementos observados, siendo **constante**, o pueden afectar de manera diferente a algunos elementos, debido a "interferencias" determinadas, en cuyo caso, se habla de **sesgo**.

El error **aleatorio**, por su parte, es menos predecible y existe mayor dificultad para controlarlo, como, por ejemplo, los factores internos del evaluado -diferencias individuales-, o acciones técnicas en el procedimiento de corrección, por ejemplo.

Confiabilidad en el momento de construcción del test

La confiabilidad se puede garantizar o afectar en diversos momentos y procesos de la medición. La confiabilidad se puede ver afectada en la construcción del instrumento, por ello es importante desarrollar de manera correcta los ítems, su redacción, la elección cuidadosa y la ubicación dentro del test (balanceo de ítems).

Una de las fuentes de error de mayor peso en esta fase, es el muestreo de contenido, la cantidad, calidad y precisión de la información relacionada con el constructo, si la información es tendenciosa, es decir, no contempla todos los factores que componen el atributo, sino uno solo de ellos, esto disminuirá la confiabilidad, como también lo hará una baja cantidad de ítems, estos se determinan de acuerdo con la complejidad del constructo y con su definición operacional.

El **escalamiento de la respuesta** es otro factor fundamental para garantizar confiabilidad, en la medición de algunos atributos psicológicos es poco recomendable el uso de ítems de **reconocimiento**, es decir, aquellos en el que el sujeto debe elegir la respuesta, cuanto menores sean las alternativas a elegir, más se reduce la confiabilidad en la medición, porque la probabilidad de que el evaluado "adivine" la respuesta correcta, es mayor, esto se puede contrarrestar incluyendo una mayor cantidad de alternativas de respuesta. Por ejemplo, en una prueba de aptitud verbal, esto se puede evidenciar de la siguiente manera:



*Exiguo es una palabra grave y por tanto debe llevar tilde.
Verdadero () Falso ()*

**Para aumentar la confiabilidad de la medida en este ítem,
se podría transformar de la siguiente manera:**

*Respecto a las reglas ortográficas, la palabra "exiguo":
Es grave pero no se marca el acento.
Es aguda y se debe marcar el acento.
Es grave, pero le falta una letra "h".
Es aguda y está bien escrita.*

Figura 2.
Fuente: propia

La **estabilidad** o no de las puntuaciones de cierto tipo de **atributos medidos** es otra fuente frecuente de error. Se espera que atributos como la **inteligencia** o la **personalidad**, al ser medidos en diferentes momentos en un mismo sujeto, presenten resultados sin **variaciones** demasiado amplias, es decir, se espera estabilidad en las puntuaciones y por supuesto en la manifestación del atributo.



Figura 3.
Fuente: shutterstock/469156640

Lo esperado no sería que en un periodo “corto” -por ejemplo, de un año-, al realizar una medición sobre la personalidad de un sujeto, esta pasa tajantemente de un rango de puntuación a otro opuesto, muy elevado o muy bajo.

Por el contrario, con atributos como las actitudes, se espera una fluctuación amplia y en menor tiempo, debido a que este atributo es mucho más susceptible a las variables del contexto.

Confiabilidad en el momento de aplicación del test

Para garantizar la confiabilidad de las mediciones, es importante ejercer un estricto control de las variables psicométricas, esto es, realizar un procedimiento de aplicación de la prueba evitando que variables externas afecten al evaluado y por ende a los resultados.

Es importante garantizar un **lenguaje** claro para el evaluado, no impositivo y con términos comunes para él durante las **instrucciones** previas a la aplicación del test. El **ambiente** debe estar **controlado** en términos de iluminación, ruido e interrupciones, se debe garantizar disponibilidad de materiales como lápices, borrador, tajalápiz, en caso de que la prueba contenga **material manipulativo**, este debe estar organizado y de fácil acceso para ser entregado en el momento pertinente, no debe estar a disposición del evaluado durante toda la aplicación, solo en el momento preciso de aplicar esta **subprueba**.

En el caso de aplicaciones colectivas (a un grupo de sujetos), se debe garantizar homogeneidad en las condiciones de medición para todos los sujetos, es recomendable contar con la presencia de **coevaluadores**, de manera tal que, ante dudas o preguntas, los evaluados no se dispersen y obtengan aclaraciones, respuestas y atención pronta.

En el caso de **pruebas de rendimiento** o de **ejecución máxima**, es importante hacer control estricto del tiempo acorde con las instrucciones consignadas en el manual de **prueba**, se debe informar con oportunidad al evaluado, es decir, durante las instrucciones previas, con cuanto tiempo cuenta para responder y que se le indicará cuando éste



Material manipulativo

Elementos físicos que componen una prueba y con los que el sujeto debe interactuar para ejecutar una tarea determinada.



Subprueba

Parte de un test que contiene una tarea específica para evaluar uno de los componentes o características del atributo medido.



Coevaluado

Persona que acompaña un proceso de aplicación de pruebas colectivo con el fin de atender de manera rápida las inquietudes y necesidades de los evaluados.

se cumpla y deba detenerse. El tiempo señalado en este tipo de prueba nunca debe excederse.

En pruebas de **ejecución típica**, se cuenta con un tiempo estimado, sin embargo, se permite que el evaluado emplee el necesario para completar la totalidad de las respuestas de la prueba.



Figura 4.

Fuente: shutterstock/668259205

Finalmente, se deben tener en cuenta aquellos factores y/o cambios internos en el evaluado, ya que también influyen en los resultados de la prueba y pueden inducir error en la medición, sin embargo, estas variables son más difíciles de controlar por parte del evaluador.

Confiabilidad en el momento de corregir el instrumento

Cuando se realiza corrección manual de la prueba es posible afectar la confiabilidad si se erra al contar las respuestas de los sujetos a los ítems, o se marca accidentalmente una respuesta no indicada por el sujeto. Una forma de controlar estos errores y garantizar la confiabilidad es con el uso de correcciones sistematizadas o electrónicas.

Dimensiones de la confiabilidad

Aunque estas características de la confiabilidad están relacionadas con los objetivos de la medición, se pueden reconocer, en general, tres de ellas, se resumen también en la tabla 1.

DIMENSIÓN DE CONFIABILIDAD	DESCRIPCIÓN BREVE	MÉTODO ESTADÍSTICO PARA CALCULAR	PROCEDIMIENTO ESTADÍSTICO (coeficiente)
Estabilidad	Puntuaciones libres de error por variaciones temporales.	Test-retest. Formas equivalentes.	Coefficiente de estabilidad. Pearson, Spearman.
Consistencia interna	Cada parte del test (ítems) mide el mismo atributo.	Formas equivalentes. Partición en mitades. Análisis de varianza de ítems.	Coefficiente de equivalencia. Pearson, Spearman. Fórmula de corrección Spearman-Brown. Alfa de Cronbach o K20.
Confiabilidad interexaminadores	Resultados de medición independientes del examinador.	Acuerdo entre jueces.	Índice Kappa. Y de Yule. Q de Kendall.

Tabla 1. Dimensiones de confiabilidad y métodos estadísticos para calcularla
Fuente: propia

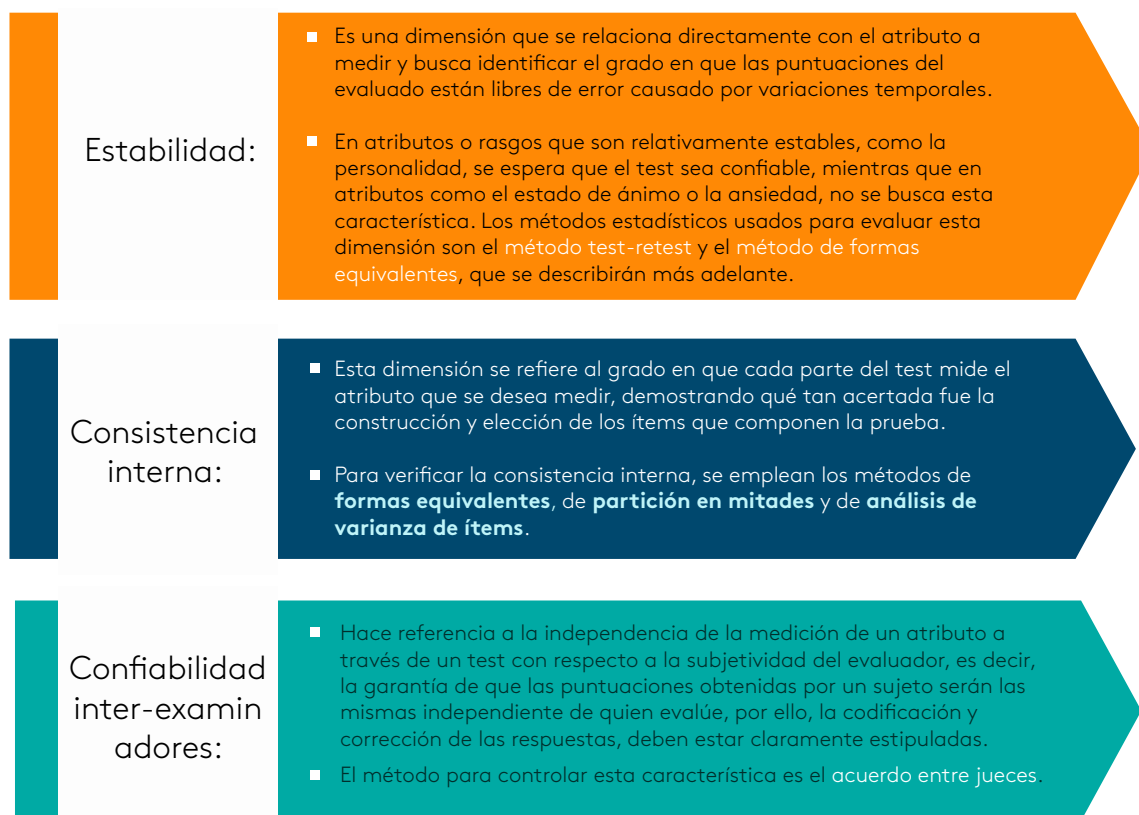


Figura 5.
Fuente: propia

A continuación, se revisarán los métodos mencionados para el control de la confiabilidad, estos se pueden ejecutar en dos momentos, durante el desarrollo del instrumento/test o en la fase experimental y durante el procesamiento estadístico de las respuestas obtenidas tras la aplicación de la prueba.

Método test-retest

Recordemos que este método evalúa la estabilidad de las puntuaciones en el tiempo. Por tanto, el procedimiento a ejecutar de este método consiste en dos aplicaciones de una prueba a la misma muestra de sujetos en un lapso intermedio, lo que busca este procedimiento estadístico es el cálculo de la correlación entre las puntuaciones obtenidas en las dos aplicaciones.

El tiempo que debe transcurrir entre las dos aplicaciones se determina teniendo en cuenta el atributo a medir y las características de los evaluados. Debe tenerse en cuenta que si el lapso es demasiado corto, para ciertos atributos como las habilidades pueden presentarse efectos de práctica y memoria, lo que induciría a una correlación falsamente alta. Sin embargo, al establecer un lapso demasiado prolongado, y de encontrarse variaciones demasiado amplias entre las puntuaciones, esto puede deberse a cambios reales de los sujetos, más no a debilidades propias del instrumento.

De acuerdo con Anastasi (1998) grado de correspondencia o relación entre los dos grupos de puntuaciones obtenidas se puede expresar estadísticamente en términos de un coeficiente de correlación (r), el más utilizado es el producto-momento de Pearson. Sin embargo, dependiendo el tipo de variable medida (nominal, intervalar, ordinal), determina el coeficiente a utilizar.

Este coeficiente presenta valores entre -1 y 1:

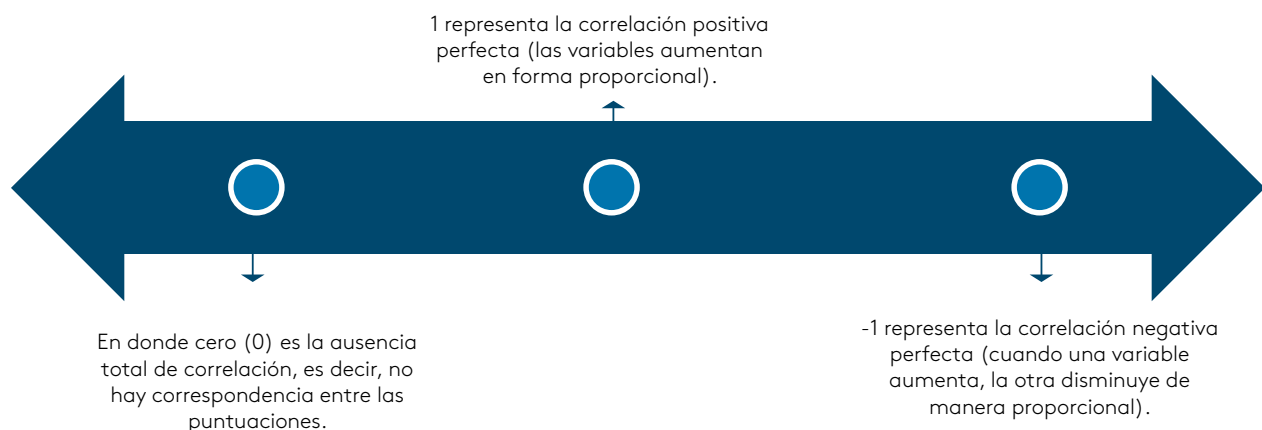


Figura 6.
Fuente: propia

Cuando se usa el coeficiente de correlación para estimar la estabilidad de una prueba, este se llama **coeficiente de estabilidad**. Para el coeficiente de Pearson, la fórmula es la siguiente:

$$r = \frac{N \sum x.y - (\sum x) (\sum y)}{\sqrt{[N \cdot \sum x^2 - (\sum x)^2] [N \cdot \sum y^2 - (\sum y)^2]}}$$

Método de formas equivalentes (alternativas o paralelas)

Este método puede ser utilizado para evaluar tanto la consistencia interna como la estabilidad de un conjunto de puntuaciones. El procedimiento consiste en la aplicación de dos formas equivalentes de la prueba a un mismo grupo de evaluados con una diferencia temporal entre las aplicaciones, las puntuaciones obtenidas, se correlacionan. El coeficiente de confiabilidad en este método se denomina: coeficiente de equivalencia.

Evaluar la confiabilidad de una prueba con este método se controlan errores en la muestra de ítems, las condiciones físicas y mentales de los examinados, la situación ambiental, el efecto de distintos evaluadores y de diversos formatos, como lo aclara Torninbeni (2004).

Para emplear este método con efectividad, es necesario verificar que las formas son realmente equivalentes, teniendo en cuenta los siguientes aspectos:

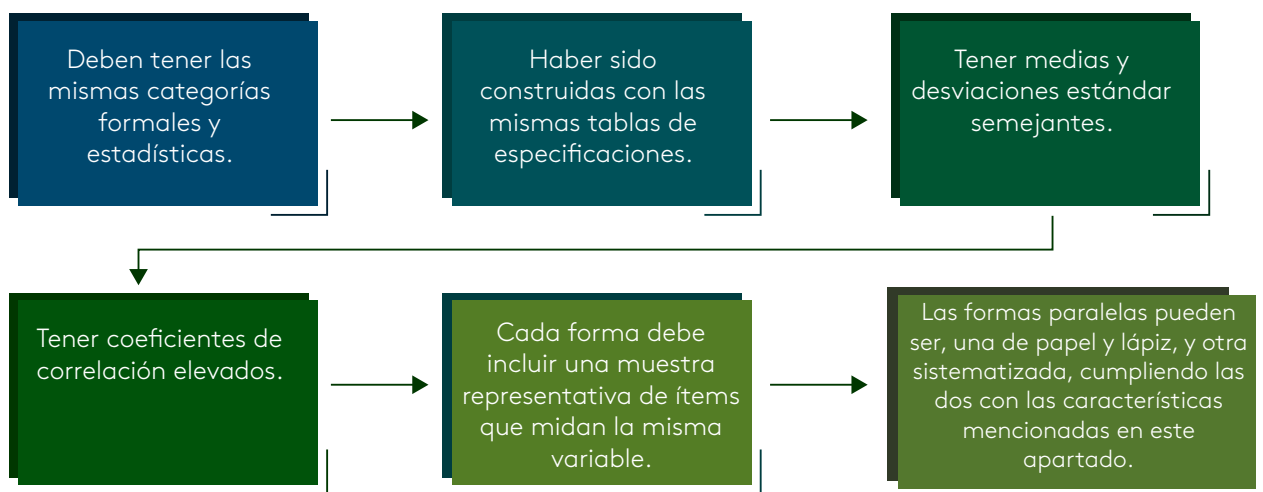


Figura 7.
Fuente: propia

Método de partición en mitades

Este método permite establecer la consistencia interna de la prueba que, según Tor-nimbeni (2004), es el grado en que las partes del test miden la misma variable. En el procedimiento se aplica una sola forma del test, en una única ocasión a un mismo grupo de sujetos, se divide el test en dos mitades comparables, obteniendo dos puntuaciones para cada sujeto, que se correlacionan por coeficiente.

Usualmente se dividen los ítems en pares e impares, cuando estos poseen el mismo grado de dificultad, se pueden dividir en la primera mitad y segunda mitad del test; otro procedimiento empleado con mayor rigor es el apareamiento de todos los ítems con un criterio estadístico determinado arbitrariamente, luego se asignan al azar los datos de cada ítem a una respectiva mitad.

Debido a que el coeficiente obtenido expresa la confiabilidad para una sola de las mitades, se hace necesario el uso de la fórmula de corrección Spearman-Brown, dado por:

$$r_n = \frac{n \cdot r}{1 + (n-1)r}$$

r_n es el coeficiente de confiabilidad estimado para la prueba completa.

r es el coeficiente de correlación entre las dos mitades.

n es el número de veces que se acorta el test (dos mitades, $n = 2$).



Instrucción

Un ejemplo de ello se puede apreciar en la base de datos en la pestaña de coeficiente de confiabilidad.

Método de análisis de varianza de ítems

Este método también permite conocer la consistencia interna del instrumento, se realiza una sola aplicación a una muestra de sujetos, obteniendo la estimación del grado de covarianza de los ítems a través del Alfa de Cronbach o la fórmula de Kuder Richardson 20 (K20). A continuación se describe la fórmula del coeficiente alfa de Cronbach:

$$\alpha = \frac{k}{k-1} \cdot \left(1 - \frac{(\sum s_i^2)}{s_t^2}\right)$$

k es el número de ítems de la prueba.

$\sum s_i^2$ es la sumatoria de la varianza de cada ítem.

s_t^2 es la varianza total de las puntuaciones del test.

Este coeficiente puede emplearse en pruebas con **ítems de selección múltiple** y **dicotómicos**, aunque con estos últimos se emplea también la fórmula K20. En el caso de test de velocidad o tiempo limitado, es recomendable usar métodos como test-retest o formas equivalentes.



Instrucción

Un ejemplo de este cálculo, se puede apreciar en la base de datos en la pestaña de análisis de varianza de ítems.



Ítems de selección múltiple

Tipo de reactivos o preguntas que tiene varias opciones de respuesta, generalmente entre tres y cinco.



Ítems dicotómicos

Tipo de reactivos o preguntas que tiene dos opciones de respuesta, por ejemplo, falso-verdadero, sí-no.

Método de acuerdo entre examinadores (jueces)

La dimensión de confiabilidad que se trabaja por este método es la confiabilidad entre examinadores. Una vez aplicada la prueba a una muestra de sujetos, los resultados se entregan a un grupo de jueces (dependiendo la disponibilidad, entre 3 y 5 jueces, para test cortos), quienes de manera independiente evaluarán estos resultados, las valoraciones de los jueces se correlacionan entre sí para verificar el grado de acuerdo. Los jueces emplean una escala para clasificar las respuestas de los sujetos a los ítems y a partir de ello, para estas estimaciones se emplean los coeficientes:

- Phi de Cramer, e índice Kappa para escalas nominales.
- Índice Kappa modificado, Q de Kendall, Y de Yule para escalas ordinales.

El índice Kappa, está determinado por la siguiente fórmula:

$$K = \frac{F_c - F_a}{N - F_a}$$

F_c es la frecuencia de coincidencias, es decir, la cantidad de veces en los que las clasificaciones de los jueces coinciden.

F_a es la frecuencia de azar, es decir, la cantidad de veces en que se espera que las clasificaciones de los jueces coincidan por azar, se obtiene sumando los productos de los subtotales de cada categoría sobre el número de casos ($F_a = \frac{\sum (\text{subtotal categoría 1} \times \text{subtotal categoría 2})}{N \text{ casos}}$).

N es el número total de casos evaluados por los jueces.

El coeficiente obtenido se puede entender de la siguiente forma:

$K = 1$ es la máxima concordancia.

$K = 0$ es la concordancia esperada a causa del azar.

En la siguiente tabla se puede apreciar la valoración del índice Kappa, de acuerdo con Tornimbeni (2004).

VALOR DE K	FUERZA DE LA CONCORDANCIA
• 0.20	Pobre.
0.21–0.40	Débil.
0.41–0.60	Moderada.
0.61–0.80	Buena.
0.81–1.00	Muy buena.

Tabla 2. Valoración del índice Kappa
Fuente: Tornimbeni (2004)

Para finalizar, el coeficiente de confiabilidad permite construir y elegir una prueba adecuada para la medición de un atributo determinado, sin embargo, para corregir las puntuaciones y perfilar es necesario conocer otros estadísticos como el **error de la medición** y el **intervalo de confianza**, para recordar estos parámetros se puede revisar el material del eje 1, acerca de las teorías de los test.



Video

La información de este apartado puede ser complementada con la video-cápsula sobre confiabilidad y sus métodos.

<https://www.youtube.com/watch?v=XHG7hFQKvJE>



Lectura recomendada

Así mismo, la lectura del artículo de Gempp (2009) ayudará con la comprensión.

El error estándar de medida y la puntuación verdadera de los tests psicológicos: algunas recomendaciones prácticas.

Gempp, R.

Págs 123 a 129

Validez

Esta propiedad de los test hace referencia a la capacidad del test para dar cumplimiento al propósito para el que fue construido; es decir, que el test para medir inteligencia mide inteligencia y no otros atributos cercanos.

La validez permite garantizar la adecuación del test, su significación y la utilidad de las inferencias y/o predicciones hechas a partir de las puntuaciones obtenidas, en ese orden de ideas, su estimación se basa en las evidencias empíricas que sustentan los anteriores factores. Estas evidencias se pueden clasificar en tres tipos: validez de contenido, validez de constructo y validez de criterio.

Cuando en el proceso de construcción de un test se verifica la validez del test, se puede afirmar sobre este que: está bien definido en términos de una variedad de observaciones y que se correlaciona con otros constructos de interés (Tornimbeni, 2004), veamos entonces, las principales fuentes de evidencia de validez de los test.



Instrucción

Para un ejercicio aplicado de estos conceptos se recomienda realizar la actividad de aprendizaje "Ficha de análisis de pruebas".

Evidencia basada en el proceso de respuesta al test

Esta evidencia está dada por la coherencia entre el constructo medido y el tipo de ejecución o respuesta dada por los sujetos. Por ejemplo, en una prueba que mide la inteligencia, es importante verificar que el evaluado realmente responde a criterios que determinan la inteligencia y no la memoria. Esto puede aclarar en qué medida las capacidades o variables accesorias relacionadas con el constructo pueden influir diferencialmente en las respuestas de los sujetos.

Evidencia basada en el contenido del test

Esta se demuestra cuando los resultados obtenidos con el contenido de un test, son producto de una muestra representativa de un conjunto de situaciones que dan cuenta de los criterios del constructo y atributo que se quiere conocer.

En la construcción del test es importante elaborar un plan detallado que incluya el contenido y los objetivos que describen el atributo a medir. Esto se logra, según Martínez (1995), con el procedimiento de validación de contenido que tiene cuatro etapas:

1. Definición del universo de situaciones a medir (contenidos, dimensiones de la variable o atributo).
2. Identificación de expertos en el tema.
3. Juicio de estos expertos acerca de la representatividad y relevancia de los contenidos, a esto se le conoce también como validación por jueces.
4. Procedimiento de resumen de los datos resultantes.

Para desarrollar la validación por jueces es conveniente construir un formato uniforme de valoración de los ítems, en el que los jueces consignarán su evaluación para el posterior análisis de estas valoraciones por parte del constructor de la prueba.

En este formato es conveniente incluir aspectos formales en términos de redacción y sintaxis de los ítems, su pertinencia y relevancia, así como su ubicación en el test, un ejemplo de ello se presenta en la matriz de la tabla 3, en la que se evalúan ítems de una prueba para medir estado de ánimo.

Ítem	Pertinencia	Relevancia	Ubicación	Aspectos formales	Observaciones
Frecuentemente me siento cansado.					
Las situaciones de la vida me ponen triste.					
Cuando estoy feliz me siento bien.					

En estos espacios los jueces consignarán sus valoraciones de acuerdo con las instrucciones dadas por el constructor, pueden ser valoraciones de 1 a 5 para cada criterio.

Tabla 3. Ejemplo de formato para validación por jueces
Fuente: propia

Evidencia basada en la estructura interna del test

Estos análisis determinan el grado en el que las relaciones entre los ítems y los componentes del test (factores, escalas, subtests) dan cuenta del atributo medido. El análisis factorial es uno de los procedimientos estadísticos más usados para verificar esta evidencia, el análisis factorial permite estudiar las intercorrelaciones entre los datos, es decir, si se aplican 4 tests de personalidad a una muestra de 100 sujetos, este análisis permite conocer las correlaciones entre las pruebas, lo que permite identificar si hay o no factores comunes medidos por las pruebas y que tan fuertemente se relacionan. De esta manera, el constructor de la prueba, podrá decidir las categorías, factores o escalas que deben componer el test.

Evidencia de la relación del test con variables externas

Evidencia convergente-discriminante

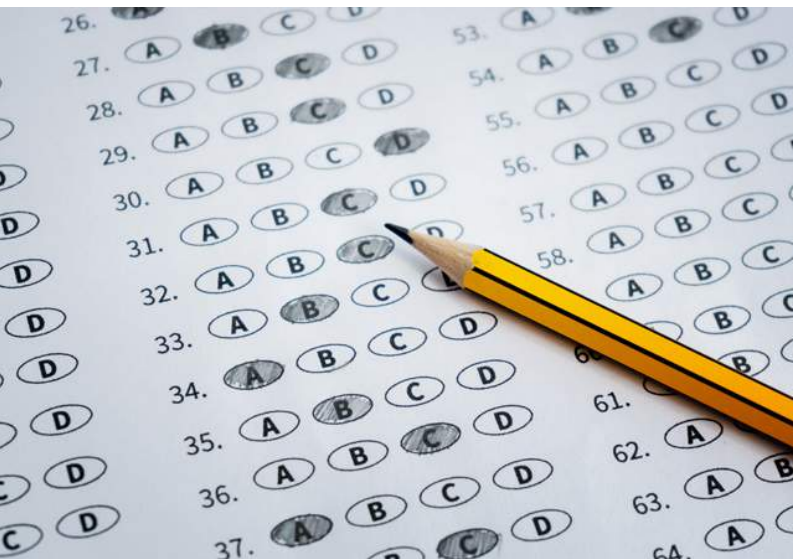


Figura 8.
Fuente: shutterstock/524144329

Esta evidencia busca contrastar los resultados obtenidos a partir del test construido con los resultados de otros test que miden el mismo atributo y a su vez con pruebas que midan atributos diferentes. Lo esperado es que las correlaciones con los test que miden el mismo atributo, sean altas, mientras que las correlaciones con los test que miden otros atributos, se espera, sean bajas. Al primer caso, se le denomina como **validez convergente**, y al segundo, de **discriminación**, es decir, que el test si tiene el poder de diferencia lo que es y no es evidencia del atributo medido.

Evidencia de las relaciones test-criterio

Esta evidencia está dada por la efectividad con que se puede predecir el comportamiento o la ejecución de un sujeto en la vida real, a partir de los resultados obtenidos. La correlación se analiza entre las puntuaciones obtenidas y criterios ya determinados que indican el nivel y/o presencia del atributo, por ejemplo, la puntuación obtenida por un sujeto en una prueba de orientación vocacional sobre su ejecución en pensamiento matemático y lo que se espera de una persona que se desempeñe en la ingeniería.

Los estadísticos comúnmente utilizados para este tipo de análisis son: el coeficiente producto momento de Pearson, la correlación biserial, el coeficiente de rangos de Spearman, sin embargo, su elección dependerá del tipo de variable medida y de la escala empleada ordinal, intervalar). El diseño que Murat (1985), citado por Tornimbeni (2004) recomienda para este tipo de análisis, consiste en:

- Selección de una muestra de sujetos representativa.

Generalización de la validez

- Se evalúa a los sujetos por el test y por el criterio a predecir (en el ejemplo descrito antes, desempeño esperado en ingeniería).
- Con los resultados se estima el coeficiente de correlación. Si el coeficiente es significativamente diferente de cero, el test resulta un predictor efectivo del criterio, de lo contrario, carecerá de **validez de criterio**.

En ese orden de ideas, la evaluación del criterio puede ser simultánea, en cuyo caso se denomina **validez concurrente**, o posterior a la evaluación por medio del test, en cuyo caso se llama **validez predictiva**. (Murat, 1985 en Tornimbeni, 2004, p. 89).

Entre los factores que afectan estos análisis, se tienen: las diferencias de grupo (sexo, edad, rasgos de personalidad, escolaridad) para lo cual se recomienda emplear muestras poblacionales lo más heterogéneas posible para establecer las comparaciones y la extensión de la prueba que afecta las varianzas de las puntuaciones (a mayor cantidad de ítems, mayores varianzas). En este sentido, los coeficientes significativos serán aquellos diferentes de cero y con valores superiores a 0.30.



Figura 9.

Fuente: shutterstock/414488914

Los procedimientos empleados para este fin son principalmente dos. La *evidencia incremental*, que indica cuánto aumenta la precisión de las predicciones logradas por el instrumento en comparación con otros. Para ello se emplean análisis de regresión múltiple, que permite **ponderar** la contribución independiente de cada predictor al criterio determinado. Otro de los procedimientos es el *metaanálisis*, que permite integrar y combinar los resultados de diversos estudios empíricos, generando métricas comunes y comprobación estadística del aporte de diversos estudios para apoyar o refutar hipótesis. Estos estudios tienen sus inicios en la década de 1980 y en la actualidad son reconocidos por las normas APA como argumentos de validación sintética (Tornimbeni, 2004, p. 92).



Ponderar

Identificar, asignar el valor aportado por cada dimensión de una variable. Por ejemplo, en las notas de los estudiantes, los valores ponderados: examen 20%, talleres 20%, proyecto de aula 60%.

Validez y utilidad práctica de los test

Según Tornimbeni (2004), tanto Thorndike como Cronbach afirman que la utilidad predictiva de un test es un criterio adecuado para determinar el valor práctico de una prueba entendida como una herramienta para la toma de decisiones, sin embargo, no es suficiente e influye en los propósitos u objetivos del uso de los test en la medición psicológica, para recordar estos, puede consultar el material del eje 2. Cuando el objetivo es seleccionar a un sujeto a partir de un criterio de predicción, se debe verificar que la prueba realmente cumpla con ello antes de aplicarla.



Instrucción

Para fortalecer la comprensión de esta propiedad, se recomienda consultar los documentos de evaluación del inventario BAI, BASIII y Matrices. En este punto puedes revisar el recurso nube de palabras.



Lectura recomendada

Así mismo, profundizar en la lectura del capítulo 3 de Meneses (2013)

Psicometría.

Meneses.

Capítulo 3. Páginas 142 a 153

Puntuaciones transformadas

En el eje 1, abordamos las etapas de construcción de una prueba y se ha mencionado como parte de ello el proceso de **baremación**, en este apartado revisaremos en detalle los procedimientos que permiten la construcción de las normas de corrección o baremos de las pruebas, que permiten la comparación de resultados individuales con los resultados obtenidos por un grupo de referencia o grupo normativo.

Recordemos los pasos del proceso de baremación: en primer lugar, se debe seleccionar una muestra representativa de la población a evaluar, a esta muestra se le aplica la versión final del test, con lo que se obtienen las puntuaciones observadas de los sujetos; posterior a ello se debe hacer la transformación de estas puntuaciones directas a puntuaciones derivadas.

El primer paso, la elección de la **muestra representativa**, cuenta con unos procedimientos de elección específicos, que permiten hacer un control riguroso de las variables de experimentación. Esto dará garantía de que los baremos son locales y actualizados y que corresponden con la población a evaluar. A la hora de elegir aplicar un test, se debe verificar que las características del evaluado guarden la mayor coincidencia posible con las normas estipuladas en los baremos. Las características más empleadas son: sexo, edad, nivel de escolaridad, nivel socioeconómico, profesión/ocupación, afiliación institucional, condiciones de salud psicológica, entre otras que son definidas por el constructor del test de acuerdo con el atributo a medir.

Las técnicas para seleccionar muestras de estandarización pueden ser probabilísticas y no probabilísticas. Las segundas, por lo general atienden a características determinadas por el investigador y las necesidades de su decisión, estas son poco empleadas en la psicometría.

Las técnicas probabilísticas garantizan la misma posibilidad de participación a todos los sujetos, se determinan bajo características poblacionales específicas, con un tamaño de muestra determinado y por lo general con procedimientos de selección aleatoria. Para profundizar sobre este tema puede consultar manuales de metodología de la investigación.

Entre las técnicas estadísticas de muestreo tenemos:

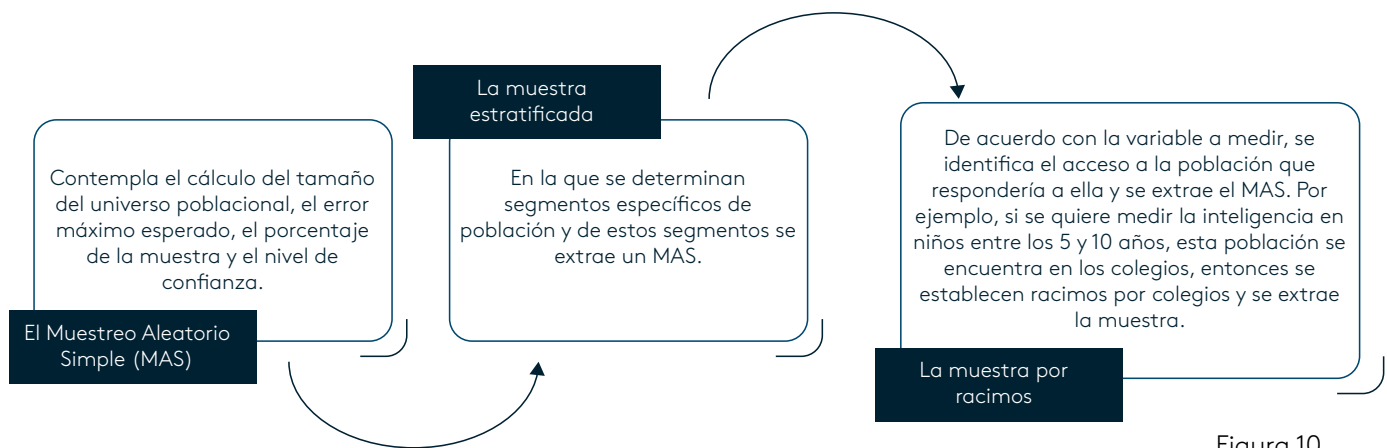


Figura 10.
Fuente: propia

Una vez seleccionada la muestra, realizada la aplicación del test y recolectadas las puntuaciones, se procede a transformarlas, a continuación, se describirán los métodos estadísticos diseñados para ello.

Métodos de transformación de puntuaciones

Por lo general, se emplean dos clases de transformaciones, las transformaciones lineales y las transformaciones no lineales.

Transformaciones lineales

Las transformaciones lineales, no alteran la distribución original de las puntuaciones y según Murat, 1985 en Tornimbeni (2004), se efectúa aplicando la ecuación:

$$X' = m.x + k$$

X' es la puntuación derivada o transformada.

x es la puntuación original.

m es una constante distinta de cero.

k es una constante mayor, menor o igual a cero.

Por lo general estas puntuaciones se organizan en una distribución de frecuencias y se expresan gráficamente a través de una curva normal o campana de Gauss. Las transformaciones lineales más conocidas son las puntuaciones standard, típicas o z , el cociente intelectual y el coeficiente intelectual de desviación.

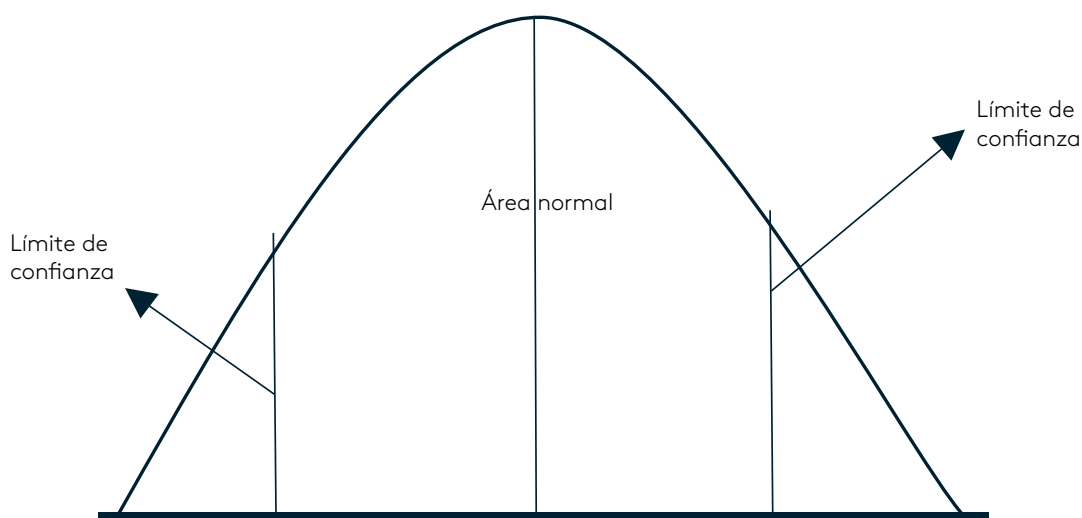


Figura 11. Curva normal o campana de Gauss
Fuente: propia

σ es la desviación estándar (distancia promedio de un dato con respecto a la media).

μ es la media o promedio aritmético (puntuación más central de la distribución).

Los límites de confianza se establecen a una desviación estándar de la media.

Puntuaciones estándar o puntuaciones z

Al transformar las puntuaciones directas en puntuaciones z, se obtiene una distribución que tiene la misma forma, pero con una media y una desviación estándar diferentes a las de las puntuaciones directas o naturales. Teóricamente la media de las puntuaciones z es igual a cero y la desviación estándar es de 1. Las equivalencias en puntuaciones z se determinan por la fórmula:

$$z_i = \frac{x_i - M}{S_x}$$

z_i es la puntuación típica de un sujeto (i).

x_i es la puntuación original del sujeto (puntuación directa).

M es la media de las puntuaciones del grupo de referencia.

S_x es la desviación típica de la distribución.

Esto implica que las puntuaciones típicas o estándar representan la distancia de la puntuación del individuo a la media en función de la desviación típica. Las puntuaciones z pueden ser positivas o negativas y oscilan entre -3,00 y +3,00. Para evitar dificultades en el informe y manejo de los datos, se corrigen los puntajes negativos con una **segunda transformación lineal**, multiplicando cada puntaje en z por una desviación estándar prefijada y a ese resultado se suma una media con valor prefijado, esto es, la fórmula:

$$z' = z.k + m$$

z' es la puntuación transformada.

z es la puntuación típica correspondiente a un sujeto determinado.

k es la desviación estándar prefijada por el constructor.

m es la media constante establecida por el constructor.

De esta manera, la nueva distribución de puntuaciones conserva la forma de la distribución de puntuaciones original, se modifica solo el valor de la media y de la desviación estándar, para comprender lo que sucede, es como si la distribución de las puntuaciones se moviera hacia un lado, en la figura 12, la distribución original se representa con la línea azul y la transformada con línea roja.

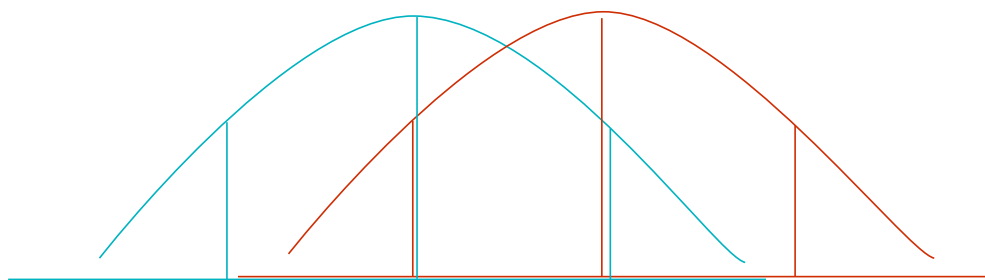


Figura 12. Ejemplo de la forma de la distribución de puntuaciones transformadas
Fuente: propia

En resumen, para transformar las puntuaciones directas en puntuaciones z o z' se siguen los siguientes pasos:

1. Se calcula la media (M) y la desviación estándar (s) de las puntuaciones directas.
2. Se calcula la desviación de cada puntaje ($x_i - M$).
3. Se calcula el valor z .
4. Se fijan los valores M' y S' para eliminar los valores negativos y disminuir los decimales y se aplica la ecuación: $z' = z.k + m$.

Un ejemplo de estas transformaciones se puede apreciar en la base de datos en la pestaña "transformación de puntuaciones z ".

Coeficiente de desviación

Esta puntuación fue desarrollada para reemplazar la versión del Cociente Intelectual (CI) planteada por la escala de Binet-Simon en 1908, que agrupaba los ítems en niveles de edad y se asumió que, si un niño respondía correctamente a los ítems de cierto nivel, eso determinaba su *Edad Mental* (EM) y que esta podía ser diferente a la *Edad Cronológica* (EC), la fórmula del CI es la siguiente:

$$CI = \frac{EM \cdot 100}{EC}$$

Este cálculo supone dificultades para argumentar y generalizar con suficiencia los resultados obtenidos con la medición, por lo que a partir de la Escala Weschler se desarrolla el coeficiente intelectual de desviación que son puntuaciones típicas derivadas (z') en lugar de cocientes. En esta escala, para la segunda transformación se asume una media de 100 y una desviación estándar de 15.

Transformaciones no lineales

Estas transformaciones asumen una distribución *ad-hoc*, es decir, diseñada únicamente para este propósito, esta distribución altera la forma de la distribución de los puntajes directos, y son fáciles de manejar para personas sin conocimientos estadísticos, las distribuciones no lineales más utilizadas son los percentiles y las puntuaciones estándar normalizadas.

Percentiles

Estas puntuaciones se expresan en términos del porcentaje de personas que quedan por debajo de una puntuación directa determinada. Un percentil representa un punto en la escala de medición original (directa) que divide el total de puntuaciones en dos: el percentil 30 (P.30) es aquel por debajo del que se registraron el 30 % total de las puntuaciones y por encima del que se encuentra el 70 % de las observaciones. Los percentiles se empiezan a contar por abajo, así un percentil bajo indica una posición baja de la puntuación del evaluado. Su cálculo se determina por la siguiente fórmula, para la que se requiere establecer la tabla de frecuencias de las puntuaciones para conocer los límites inferior y superior del rango:

$$P_x = \text{lim.inf} + i \cdot \frac{(N \cdot P_x/100 - f_a)}{f_p}$$

P_x es el percentil a calcular.

lim.inf es el límite exacto inferior del intervalo que contiene P_x .

f_a es la frecuencia acumulada hasta el límite, o frecuencia por debajo de la clase que contiene P_x .

f_p es la frecuencia en el intervalo que contiene P_x o frecuencia propia.

i es la amplitud del intervalo.

N número total de casos.

Para el cálculo de los percentiles se debe construir una tabla de frecuencias de respuesta, de acuerdo con la cantidad de datos, esta puede ser o no agrupada en intervalos. Un ejemplo de estas transformaciones se puede apreciar en la base de datos en la pestaña "transformación de puntuaciones P".

Puntuaciones estándar normalizadas (zn)

La transformación de puntuaciones directas a puntuaciones estándar normalizadas, se inicia con el cálculo de rangos percentiles de las puntuaciones originales (directas). Posteriormente se convierten los rangos percentiles en proporciones, por ejemplo el percentil 88,5 se convierte en 0,885, luego en la *tabla de áreas bajo curva normal*, se ubica la puntuación z debajo de la que se encuentra la proporción dada.

Entre las puntuaciones estándar normalizadas se encuentran las puntuaciones "T", con una media de 50 y una desviación estándar de 10, es de las más utilizadas en la construcción de baremos de las pruebas psicométricas; las puntuaciones estaninas con media de 5 y desviación estándar de 2, son poco utilizadas debido a la complejidad que representan.

Para estimar las puntuaciones T, se deben seguir los pasos:

1. Convertir la puntuación original en percentil.
2. Transformar el percentil en proporción.

3. Ubicar la puntuación z en la tabla de áreas bajo curva normal.
4. Calcular la puntuación T con la fórmula:

$$z' = z.k + m$$

z' es la puntuación T .

z es la puntuación típica correspondiente a un sujeto determinado.

k es la desviación estándar igual a 10.

m es la media igual a 50.

En la actualidad para ejecutar estas transformaciones se cuenta con software especializados que agilizan el proceso de las operaciones, sin embargo, es importante comprender los significados, procesos y representaciones de estos procedimientos estadísticos, ya que de no comprenderlos se dificulta la comprensión y uso de los datos arrojados por el programa.



Instrucción

Un ejemplo de estas transformaciones se puede apreciar en la base de datos en la pestaña "transformación de puntuaciones zn".



Lectura recomendada

Como estrategia de fortalecimiento de esta temática, se recomienda realizar la lectura del capítulo 4 de Aragón (2015)

Evaluación psicológica: historia fundamentos teórico-conceptuales y psicometría

Borja

Capítulo 4, páginas 60 a 72.



Instrucción

Finalmente invitamos al desarrollo de la prueba objetiva sobre los temas abordados en este eje.

- Aiken, L. (2003). *Test psicológicos y evaluación*. (11ª edición). Ciudad de México, México: Pearson Educación.
- Anastasi, A. y Urbina, S. (1998). *Test psicológicos*. (7ª edición). Ciudad de México, México: Prentice Hall.
- Aragón, L. (2015). *Evaluación psicológica: historia fundamentos teórico-conceptuales y psicometría*. (2a. ed.). Ciudad de México, México: Editorial El Manual Moderno.
- Consejo General de Colegios Oficiales de Psicólogos. (s.f.). *Evaluación de la batería BAS-II*. Madrid, España.
- Consejo General de Colegios Oficiales de Psicólogos. (s.f.). *Evaluación del inventario BAI*. Madrid, España.
- Gempp, R. (2006). El error estándar de medida y la puntuación verdadera de los tests psicológicos: algunas recomendaciones prácticas. *Red Terapia Psicológica*, 2(24), p. 117-130.
- González, L., F. (2007). *Instrumentos de evaluación psicológica*. La Habana, Cuba: Ciencias Médicas. Capítulo 2. Recuperado de http://newpsi.bvs-psi.org.br/ebooks2010/en/Acervo_files/InstrumentosEvaluacionPsicologica.pdf
- Martínez, A. M. R., y Hernández, L. M. V. (2014). *Psicometría*. Madrid, España: Difusora Larousse - Alianza Editorial.
- Martínez, R. (1995). *Psicometría: Teoría de los test psicológicos y educativos*. Madrid, España: Síntesis.
- Meneses, J. (2013). *Psicometría*. Barcelona, España: Editorial UOC.
- Sánchez, F., Santamaría, P., y Abad, F. (2015). *Matrices. Test de inteligencia general*. Madrid, España: TEA Ediciones.
- Thomas, P. H. (2015). *Pruebas psicológicas: una introducción práctica* (2a. ed.). Ciudad de México, México: Editorial El Manual Moderno.
- Tornimbeni, S., et ál. (2004). *Introducción a los test psicológicos*. Buenos Aires, Argentina: Brujas.

PSICOMETRÍA BÁSICA

Zulma Ramos Vega

EJE 4

Propongamos

Bienvenido(a)

Habiendo hecho el recorrido histórico y conceptual, y el abordaje de la funcionalidad de la psicometría desde la perspectiva analítica de los contextos y los retos de los profesionales del futuro para poner en práctica el sustento estadístico de las propiedades de los test, nos adentraremos en los procedimientos de aplicación, corrección y perfilación de las pruebas.

Durante estas dos semanas haremos énfasis en la elección y el correcto uso de las pruebas, teniendo en cuenta sus componentes, el control de variables en la aplicación, cómo ejecutar los procedimientos de corrección, perfilación y, a partir de ello, cómo orientar las decisiones de intervención o asesoría psicológica.

En la primera semana abordaremos el proceso de elección de un test, los aspectos que guían esta decisión, entre ellos las propiedades de la prueba, características de aplicación, adaptación y actualización, los diferentes materiales que componen los test y cómo manejarlos.

En la segunda semana revisaremos el control de variables para la aplicación de los test, qué hacer con las puntuaciones obtenidas, cómo comprender y aplicar esta información en el ejercicio del psicólogo, estos son los procedimientos de corrección, perfilación e informe, y el verdadero aporte del psicólogo desde la emisión de un concepto integral que oriente las decisiones sobre las acciones y recomendaciones para el evaluado.

A través de los recursos de aprendizaje y las actividades, se aplicarán dichos procedimientos, llevándonos a reflexionar sobre la pregunta central de este eje: ¿De qué forma debería orientar el psicólogo sus decisiones frente al uso del material psicométrico y frente a los resultados obtenidos con este?

Es recomendable que dediquemos unas horas diarias a la lectura del material propuesto para este eje, tomar apuntes, subrayar los aspectos esenciales, desarrollar esquemas, realizar las actividades de refuerzo, esto facilitará la comprensión del tema, consultar el glosario aportará claridad y reforzará la comprensión de los aspectos más importantes del módulo.

¡Adelante!

Vamos a comenzar.

Elección de uso de un test



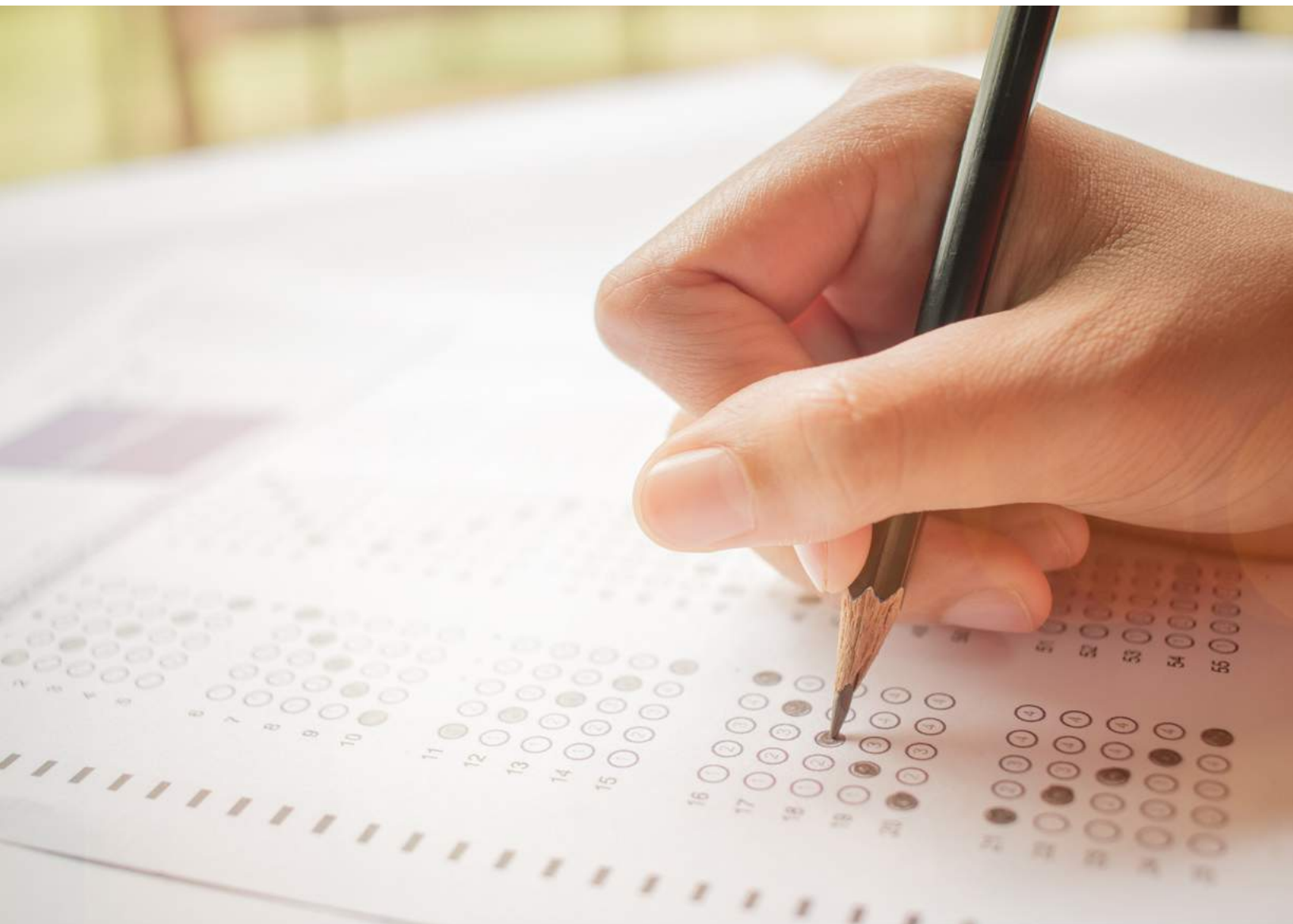


Figura 1.

Fuente: Shutterstock/662982442

Uno de los aspectos más importantes acerca de los test es su uso, es decir, las implicaciones de su aplicación, las recomendaciones y las decisiones que surgen sustentadas en estos datos. Es por ello que para el psicólogo no basta con saber aplicar un test, es necesario ser precisos, éticos, responsables y conscientes desde la misma elección del test hasta la emisión de su informe. A continuación, describiremos una guía de pasos a tener en cuenta para la elección de aplicación de una prueba.

La aplicación de una prueba psicológica implica mucho más que el procedimiento técnico de aplicación, es pertinente tener en cuenta una serie de criterios que permiten una medición profesional óptima, y que realmente aporte a los evaluados. Contemplando las necesidades de evaluación, los objetivos, las características psicológicas y sociodemográficas del evaluado, el contexto y las características mismas de la prueba.

En primer lugar, se debe tener claro el objetivo de la medición, es decir, el para qué (puede remitirse a la tabla 1 del eje 2). Los objetivos de medición, guiarán la toma de decisiones frente a los procedimientos orientados a evaluar los atributos psicológicos, seguido de la selección de la prueba adecuada para tal fin. Con esta información se procede a verificar las pruebas disponibles que miden dicho atributo, cabe aclarar, que en este punto es necesario identificar el constructo bajo el que están construidas las pruebas y contrastar con el objetivo de la medición, para que todo el proceso lleve el mismo camino argumentativo teórico.



Lectura recomendada

Con este fin se puede consultar el capítulo 2 de:

Pruebas psicológicas: una introducción práctica.

Thomas, P. H.

Una vez teniendo claridad de las pruebas que permitirían cumplir con el objetivo de la medición, se procede a hacer lectura y revisión de sus respectivas fichas técnicas, en las que se debe tener en cuenta:

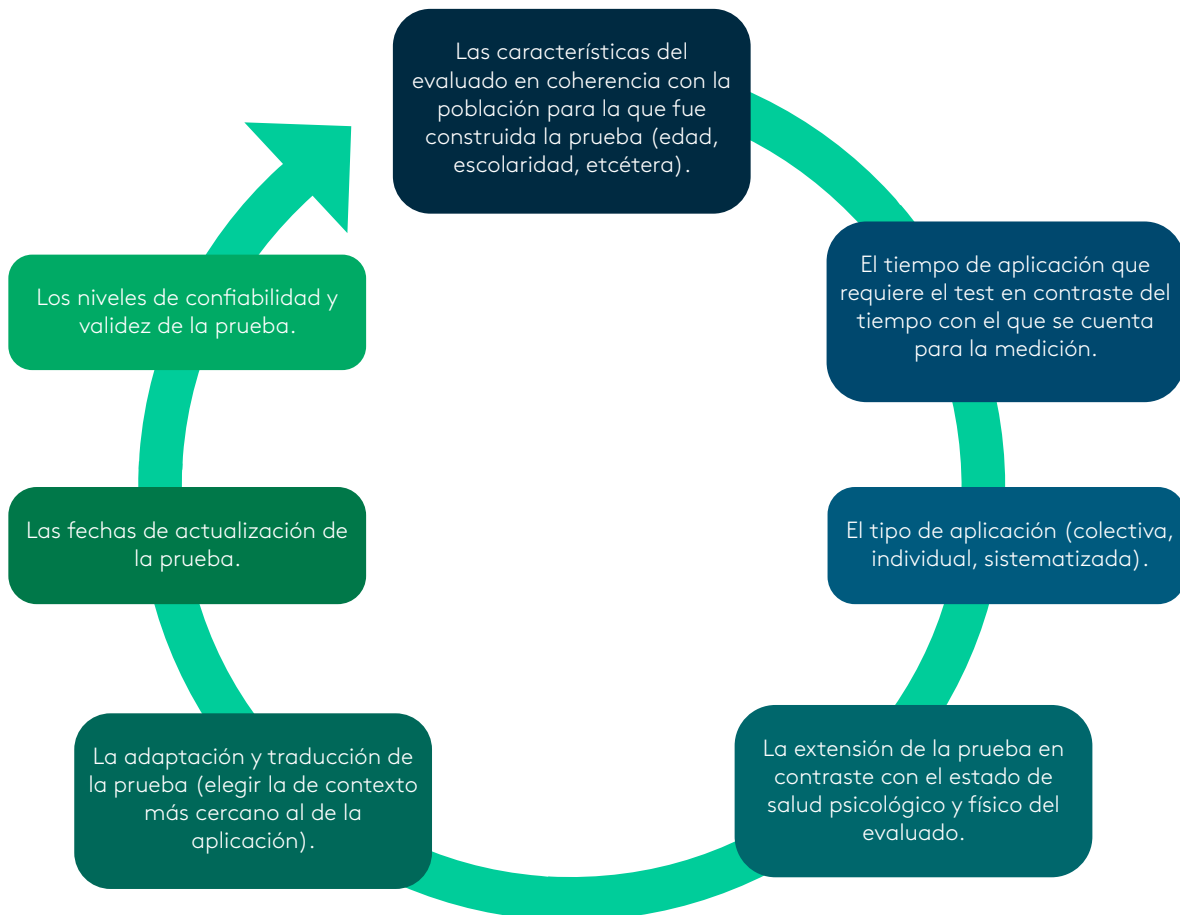


Figura 2.
Fuente: propia

En la siguiente tabla se resumen los pasos sugeridos y las diversas alternativas que se pueden contemplar en cada uno de ellos:

Pasos	Alternativas a contemplar en cada caso
Aclarar objetivo de medición.	Selección. Clasificación. Asesoría. Apoyo diagnóstico. Establecimiento de planes de acción o tratamiento. Evaluación y seguimiento. Investigación.
Atributo a medir.	Adicciones. Acoso. Aptitudes. Actitudes. Creatividad. Vocación/Intereses. Simulación. Autoestima/Autoconcepto. Rendimiento. Inteligencia. Inteligencia emocional. Psicopatología. Conducta alimentaria. Aprendizaje. Calidad de vida y bienestar. Adaptación y conducta. Lenguaje/Lectoescritura. Personalidad. Indicadores neuropsicológicos. Ansiedad. Depresión. Estrés. Memoria. Desarrollo. Atención.

Pruebas disponibles según atributo y objetivo de medición.	Catálogos de editoriales. Bases de datos. Laboratorios de psicometría. Pruebas institucionales.
Revisión de fichas técnicas de pruebas.	a) Las características del evaluado (edad, escolaridad, etc.). b) El tiempo de aplicación. c) El tipo de aplicación (colectiva, individual, sistematizada). d) La extensión de la prueba. e) La adaptación y traducción de la prueba. f) Actualización de la prueba. g) Confiabilidad y validez de la prueba.

Tabla 1. Pasos y alternativas para la elección de la prueba adecuada
Fuente: propia

Así mismo, es pertinente tener en cuenta las implicaciones éticas de la aplicación de material psicométrico, analizar las consecuencias y beneficios a corto y largo plazo de la información obtenida en las pruebas y el nivel de afectación de las acciones tomadas a partir de ello.



Instrucción

Para fortalecer las competencias en este procedimiento se recomienda desarrollar el siguiente caso simulado.

Adaptación de los test

En su mayoría, los test han sido creados en contextos culturales y científicos distintos al colombiano, lo cual implica que las normas establecidas para su análisis no son generalizables, en la mayoría de los casos también han sido contruidos en idiomas diversos y las expresiones que componen los ítems también son diferentes. Esto supone que su aplicación en contextos diferentes cuente con sesgo.

De presentarse sin control, estos sesgos acarrearán resultados erróneos y efectos muy graves sobre la vida de las personas evaluadas. La necesidad de adaptar los test no surge solamente de las implicaciones éticas, también surge como respuesta a los fenómenos migratorios y las dinámicas de la globalización, cuyo crecimiento en la actualidad es acelerado. Así mismo, la adaptación de los test contribuye a la validación empírica de las teorías y el control de las diferentes fuentes de sesgo identificadas por Van de Vijver y Tanzer (1997).



Sesgo

Estadísticamente implica la diferencia entre la puntuación esperada y la puntuación verdadera, así mismo puede considerarse como el curso que toma la medición en una dirección específica.

Sesgo de constructo

Este sesgo hace referencia a que el constructo medido no tiene el mismo significado o sentido en todos los grupos culturales, es decir, algunas conductas son valoradas de manera diferente por lo que es posible que al usar un test en una cultura diferente a la que fue base para su construcción, la diferencia en las puntuaciones se debe a la diversidad cultural y no corresponden a la verdadera medida del atributo.

Sesgo metodológico

Este sesgo se presenta en tres formas, el sesgo de muestra, el sesgo en el instrumento y el sesgo de administración.

- El *sesgo de muestra* hace referencia a que en muestras diferentes no existe posibilidad de comparar mediciones, por lo que es necesario homogenizar las muestras.
- El *sesgo de instrumento* proviene de las características de la prueba, los estímulos que contienen los ítems pueden ser más familiares para unos sujetos que para otros, esto incluye también el idioma y el proceso de traducción.
- El *sesgo de administración* incluye la comunicación de instrucciones para el desarrollo del test, el tipo de palabras empleadas, la forma de explicación del aplicador, si la traducción no es clara, el aplicador puede presentarlas a su consideración.

Sesgo del ítem

Este sesgo se da cuando el contenido del ítem puede tener diferente significado de una cultura a otra, así mismo la deseabilidad social, es decir, que el evaluado responda lo que cree que se espera de él y no lo que realmente considera.

Métodos de adaptación de los test

La adaptación cultural de una prueba es mucho más que la simple traducción de un idioma a otro, no solo se contemplan las palabras a utilizar sino también las variables culturales que puedan influir. Tornimbeni (2004), ejemplifica esto de la siguiente manera:

”

La traducción del ítem del NEO PI-R, I wouldn't enjoy vacationing in Las Vegas por No disfrutaría tomando vacaciones en Las Vegas, es correcta, sin embargo, el significado para un evaluado de Estados Unidos y un evaluado colombiano, sea diferente (Tornimbeni, 2004, p. 148).

Van de Vijver y Leung (1997) señalan tres niveles de adaptación de los test, la primera de ellas, la *aplicación* que consiste en la traducción de la prueba de un idioma al otro sin contemplar la equivalencia del constructo. La segunda, es la *adaptación*, en esta se incluye la transformación, adición, o eliminación de algunos ítems de la prueba original. La tercera, es el *ensamble* (*assembly*) es una modificación profunda que casi logra la transformación a un nuevo instrumento.



Lectura recomendada

La *International Test Commission*, presenta las pautas para la adaptación de los test de una cultura a otra, agrupándolas en contexto, desarrollo del instrumento y adaptación, y administración e interpretación, de acuerdo con Tornimbeni (2004), son las que se pueden observar en las páginas 2 a 3 del documento *International Test Commission Guidelines for Translating and Adapting Test de la International Test Commission*.

Técnicas de traducción

De igual forma, existen técnicas de traducción para una adecuada adaptación de los test. Los métodos más comunes son la traducción directa y la traducción inversa. En la *traducción directa*, un grupo de traductores hacen la traducción del idioma original al idioma nuevo, luego otro grupo de traductores revisa la equivalencia de las dos versiones. En la *traducción inversa*, un grupo de traductores realiza una traducción desde el idioma original al nuevo idioma, otro grupo de traductores toma

el test traducido en el nuevo idioma y lo traduce al idioma original. Las versiones original y retraducida, se comparan determinando su equivalencia.

Una vez verificada la traducción, se ejecutan diseños experimentales para verificar la equivalencia de los test. La administración de la prueba en el idioma original y de la prueba traducida a sujetos bilingües, puede haber variaciones en la aplicación para controlar algunos sesgo inherentes a las características de los sujetos evaluados. El segundo, consiste en la administración de versión original y la traducción inversa a sujetos monolingües en idioma original, y el tercero, consiste en la administración de la versión original a sujetos que hablan el idioma original y la versión traducida a un grupo que hable el idioma al que ha sido traducida la prueba.

La obtención de los datos tras cualquiera de los diseños, se analizarán estadísticamente a partir de la Teoría de Respuesta al Ítem, con Funcionamiento Diferencial de Ítems (FDI), y a partir de Teoría Clásica, con métodos como: Delta de Angoff y Ford, suma de chi-cuadrada, Mantel y Haenszel.




Instrucción

Para reforzar este tema se sugiere desarrollar la actividad de aprendizaje 2 Mapa adaptación de los test.

Las pruebas, sus elementos

Antes de continuar, veamos una videocápsula en donde se describe y ejemplifica los procedimientos de aplicación de un test psicométrico, muestra los componentes de una prueba y su correcto uso.

 **Video**

Se sugiere ver al menos la primera parte de la aplicación de la Escala Weschler, video realizado por la Universidad Nacional de Mar del Plata (Argentina).

<https://www.youtube.com/watch?v=MrTUKFOhyUs>

https://www.youtube.com/watch?v=tT1B_4K4na8&t=1s

Todas las pruebas psicológicas cuentan con un **manual de prueba**. En algunos casos el manual de prueba se divide en dos documentos diferentes, el *manual técnico* y el *manual de aplicación y corrección*. La información que contiene este documento va desde la ficha técnica de la prueba, la descripción del constructo y su fundamentación teórica, la fundamentación metodológica y estadística de la estandarización de la prueba y sus grados de validez y confiabilidad, cuenta con una sección de descripción de la prueba, sus sub-pruebas y componentes, así como las instrucciones específicas de aplicación, corrección y perfilación, al final del manual se presentan las tablas de baremación.



Figura 3. Contenidos del manual de prueba
Fuente: propia

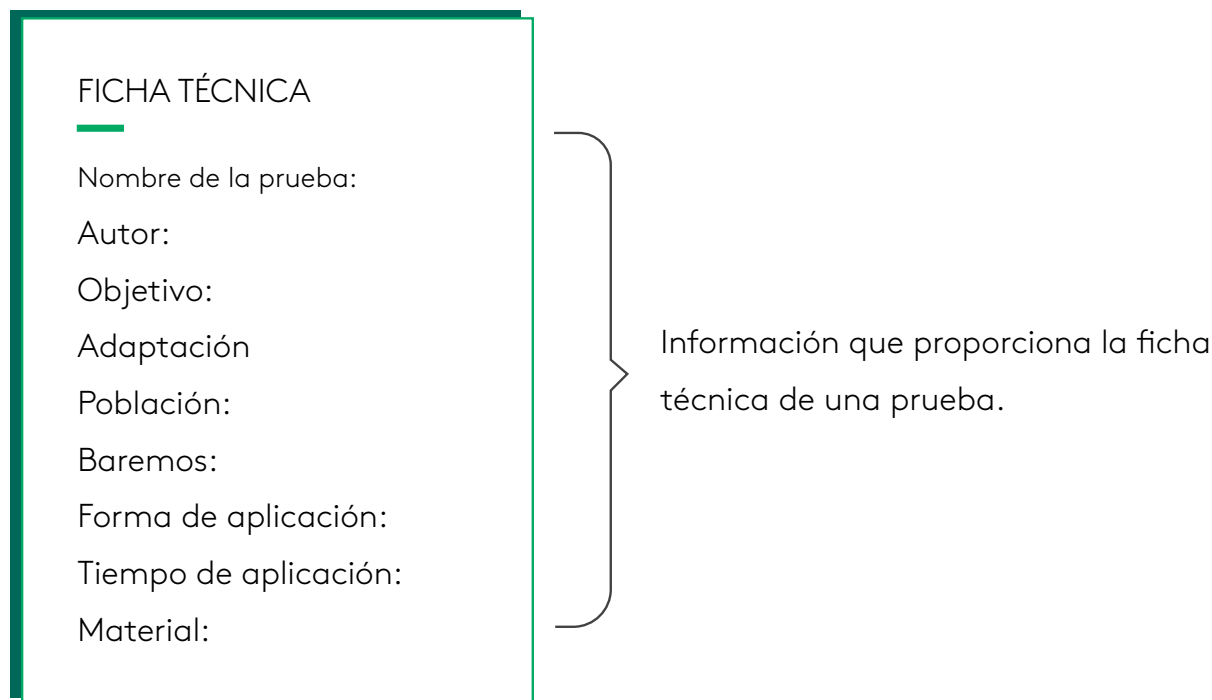


Figura 4. Contenido de la ficha técnica de un manual de prueba
Fuente: propia

Otro de los componentes de una prueba es el protocolo de aplicación o cuadernillo de prueba en el que, por lo general, se consignan las instrucciones dirigidas al evaluado para responder la prueba, así mismo se presentan entre 1 y 3 ejemplos del formato y procedimiento de respuesta para verificar que el evaluado tenga clara la forma en que debe responder. Así mismo, contiene los ítems que componen la prueba. En algunas pruebas, el cuadernillo de respuestas y el de aplicación se encuentran en el mismo documento.

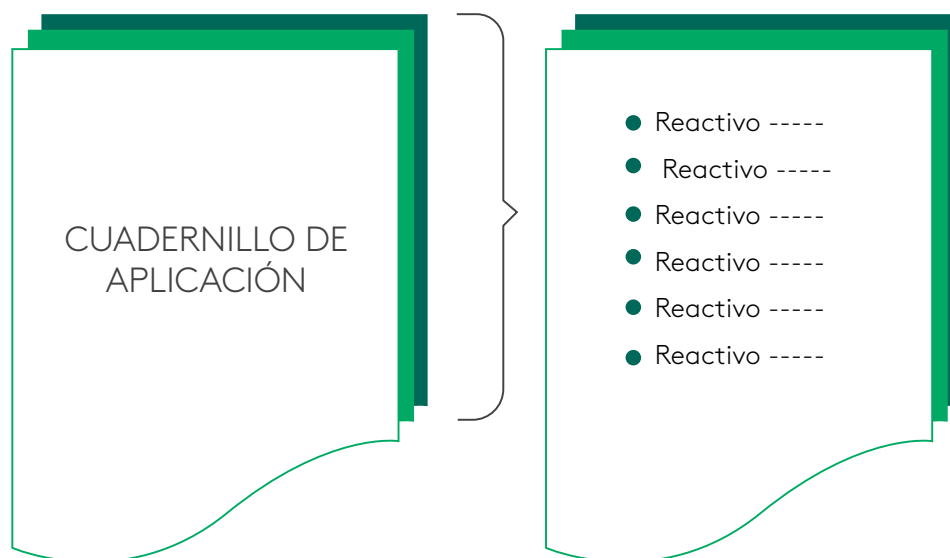


Figura 5. Contenido del protocolo o cuadernillo de aplicación de una prueba
Fuente: propia

El protocolo de respuestas, también conocido como hoja de respuestas, es un documento en el que el sujeto debe consignar sus respuestas, generalmente con lápices N° 2, para corrección automatizada, algunas pruebas tienen papel químico que marca las respuestas en una plantilla incluida, otras poseen unas plantillas de corrección.

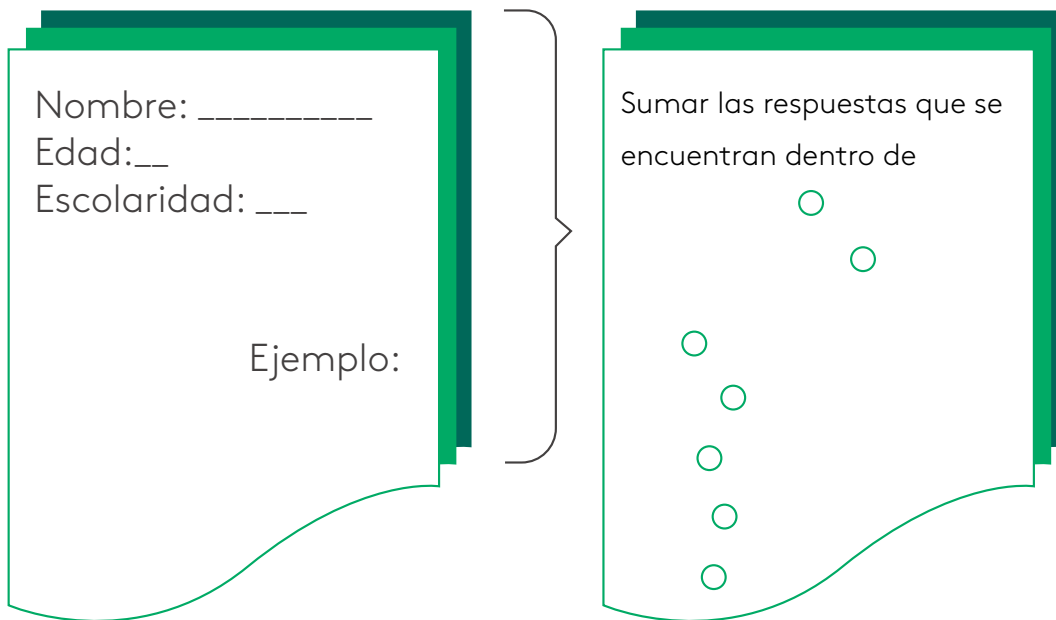


Figura 6. Ejemplo de una hoja de respuestas con plantilla de corrección incluida
Fuente: propia

Algunas pruebas contienen material manipulativo, que es aquel con el cual el evaluado interactúa según las instrucciones del evaluador para valoración de algunas funciones ejecutivas o práxicas, pueden ser cubos, rompecabezas o libretas de estímulos.

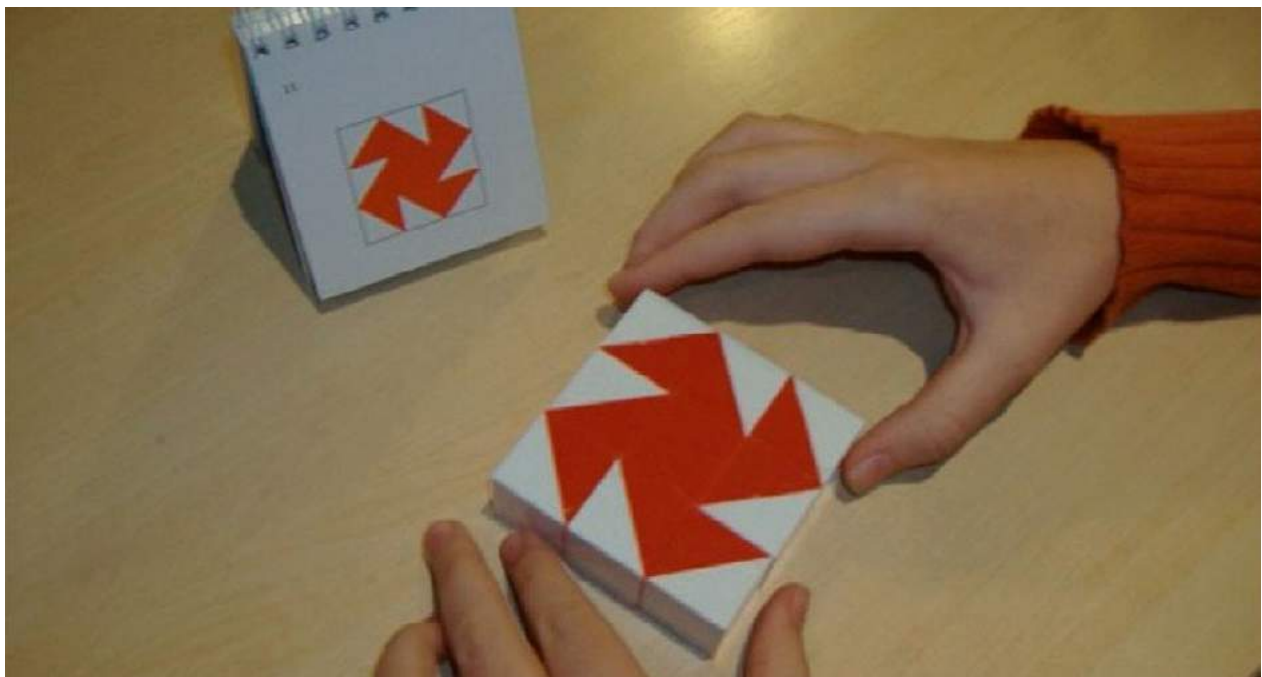


Figura 7. Ejemplo de material manipulativo de una prueba
Fuente: goo.gl/tcENj6

Las pruebas de corrección manual, en su mayoría contienen plantillas de corrección, por lo general son unos acetatos marcados con las claves de respuesta, estos se traspone-
nen sobre la hoja de respuesta, siguiendo las marcas indicadoras, en general las pruebas psicométricas se construyen en formatos de tamaño no usuales para preservar su autenticidad, los derechos de autor y evitar falsificaciones. Otras pruebas tienen formas de corrección a partir de lectores ópticos, a través de sistematizaciones en programas especializados, o correcciones on-line a través de códigos PIN.

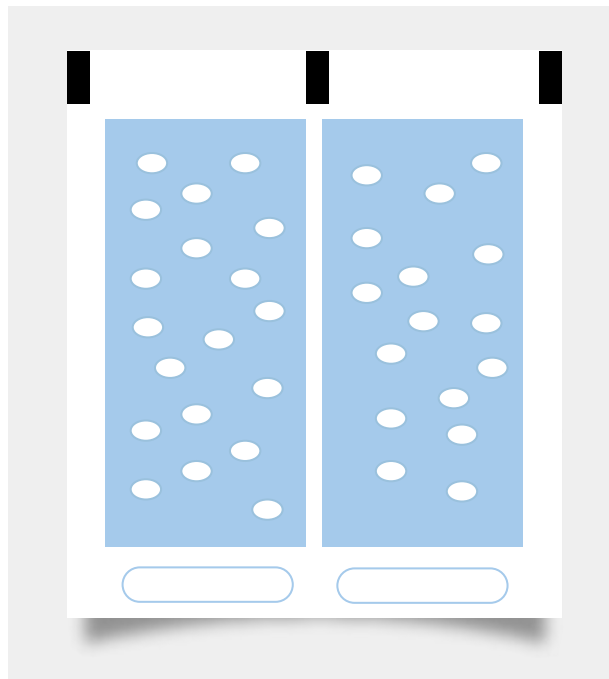


Figura 8. Ejemplo de una plantilla de corrección manual.
Fuente: Propia

Algunas pruebas poseen gráficas de perfilación incluidas en las mismas hojas de respuesta o en documentos que vienen por separado en estos, se consignan las puntuaciones transformadas, se grafican las relaciones entre los diversos factores que componen la prueba y que dan cuenta del atributo medido y la puntuación del evaluado respecto a ello.



Instrucción

Para profundizar y conocer estos elementos de pruebas reales, recomendamos consultar las zonas activas 1, 2 y 3.

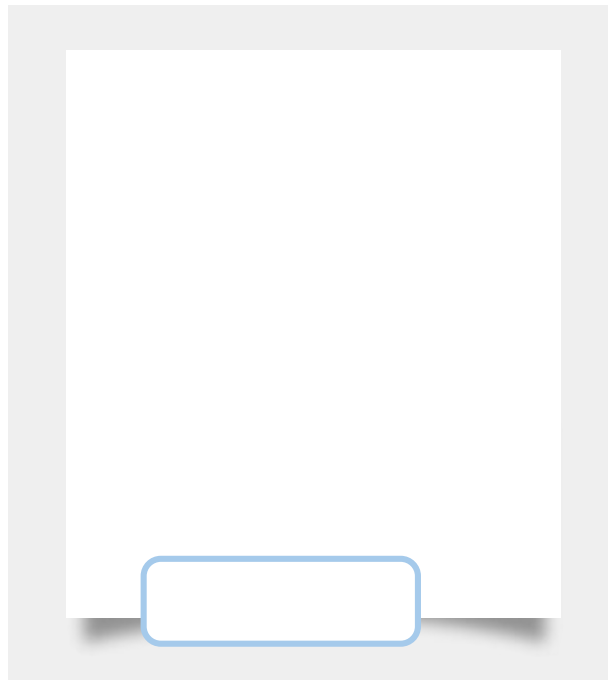


Figura 9. Ejemplo de una hoja de perfilación
Fuente: propia

Procedimiento de aplicación

Como se ha mencionado en los ejes previos, una prueba psicométrica se equipara a un arreglo experimental clásico, por lo que el procedimiento de aplicación debe cumplir con parámetros rigurosos para garantizar la confiabilidad y validez de los resultados de la aplicación.



Figura 10.
Fuente: <https://goo.gl/uBdzmM>

Control de variables

Uno de los principales aspectos a tener en cuenta para la aplicación de material psicométrico, es el control de variables psicométricas. Es decir, organizar, verificar y controlar cualquier factor que pueda ejercer algún tipo de influencia sobre el evaluado y por ende alterar las puntuaciones resultantes de su ejecución. Aunque el ambiente de evaluación debe ser neutral, también es recomendable adecuarlo a las características del evaluado para facilitar el *rapport* o confianza como, por ejemplo, los niños.

Esto incluye la preparación del **ambiente** de evaluación en términos de control de ruido e interrupciones, iluminación adecuada, **espacio** y mesas suficientes, se debe entregar al menos dos lápices, borrador y tajalápiz a cada evaluado; en caso de **aplicación colectiva** se debe garantizar el acompañamiento de co-evaluadores para dar respuesta oportuna a las dudas de los evaluados.

El **material** debe estar preparado previamente y asequible en una mesa auxiliar para presentarlo en el momento pertinente, principalmente el material manipulativo para las pruebas que cuentan con él. Los protocolos de respuesta y de aplicación se deben ubicar en la mesa en dirección al evaluado, si este es diestro los lápices deben estar de ese lado, justo después de la hoja de respuestas, y esta, ubicada enseguida del protocolo o cuadernillo de aplicación.

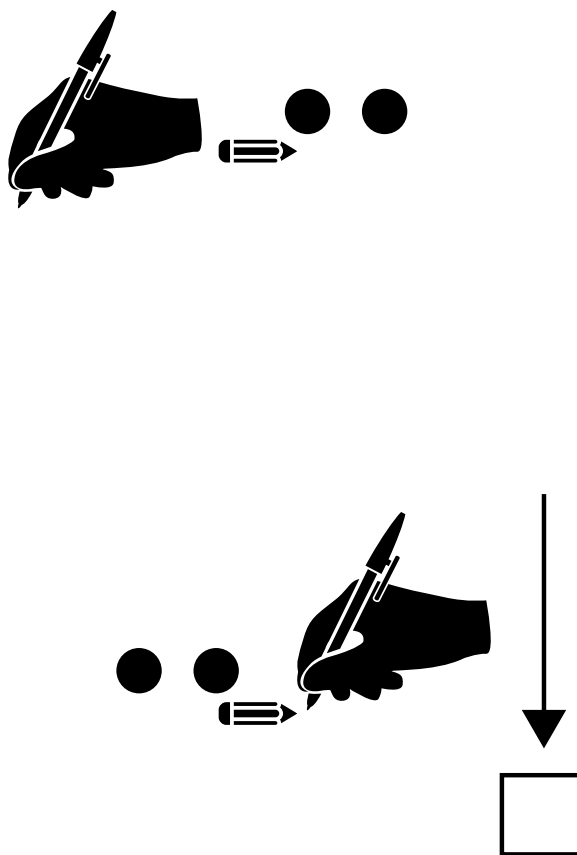


Figura 11. Ejemplo disposición del material en la mesa de acuerdo con la dominancia manual del evaluado
Fuente: propia

El trato con el evaluado debe ser respetuoso y amable, evitando tutearlo, es importante aclarar que la presencia del evaluador es para acompañar y guiar al sujeto evaluado, no para vigilar ni tampoco para indicar si lo está haciendo bien o mal, el lenguaje verbal y no verbal debe ser muy cuidadoso, tranquilo, sin dar muestras de aprobación o desaprobación.

Si bien cada prueba contiene en sus manuales las **instrucciones**, en general es importante tener en cuenta que el lenguaje debe ser neutral y claro para el sujeto, se debe indicar la forma de responder la prueba, es decir, la mecánica de respuesta, para esto se cuenta con los ejemplos que trae la prueba misma y que se pueden realizar en conjunto con el evaluado para verificar su comprensión.

Es importante recordarle al evaluado la cantidad de preguntas a responder, la importancia de **respuestas sinceras** y en caso de contar con una prueba de rendimiento con **tiempo límite**, esto se debe indicar con claridad al sujeto desde el inicio y se debe cronometrar dando cumplimiento estricto al parámetro temporal señalado.

Es recomendable que el psicólogo esté presente en todo momento de la aplicación, de manera discreta, debe estar atento a cualquier indicador relevante para la medición en el sujeto. También debe estar dispuesto a aclarar cualquier duda que surja en el evaluado sobre la mecánica de la respuesta, nunca sus respuestas o intervenciones deben inducir alguna respuesta.

Cuando la prueba indica un tiempo límite de ejecución, el aplicador o evaluador debe detener el cronómetro e indicar al evaluado que se ha terminado el tiempo, inmediatamente debe proceder a retirar la hoja de respuestas y posterior a ello el protocolo de aplicación.

Cuando la prueba no tiene un tiempo límite, sino un estimado, se debe esperar a que el evaluado termine, se retira el protocolo de respuestas y finalmente el protocolo de respuestas, verificando con una rápida inspección visual que el sujeto haya respondido todos los ítems de manera correcta, de no ser así, se buscará el ítem faltante en el protocolo de aplicación y se le solicita completarlo.

Para finalizar la sesión de aplicación de material psicométrico, se recomienda indagar por el estado del evaluado y cómo se sintió, reiterándole el objetivo de la aplicación y la utilidad de los datos obtenidos para su proceso. Se le indica que en la sesión siguiente se hará la devolución de resultados, en donde se le comunicará qué información se obtuvo y cómo se procederá a partir de ello.

Procedimiento de corrección



Figura 12.

Fuente: shutterstock/588502703

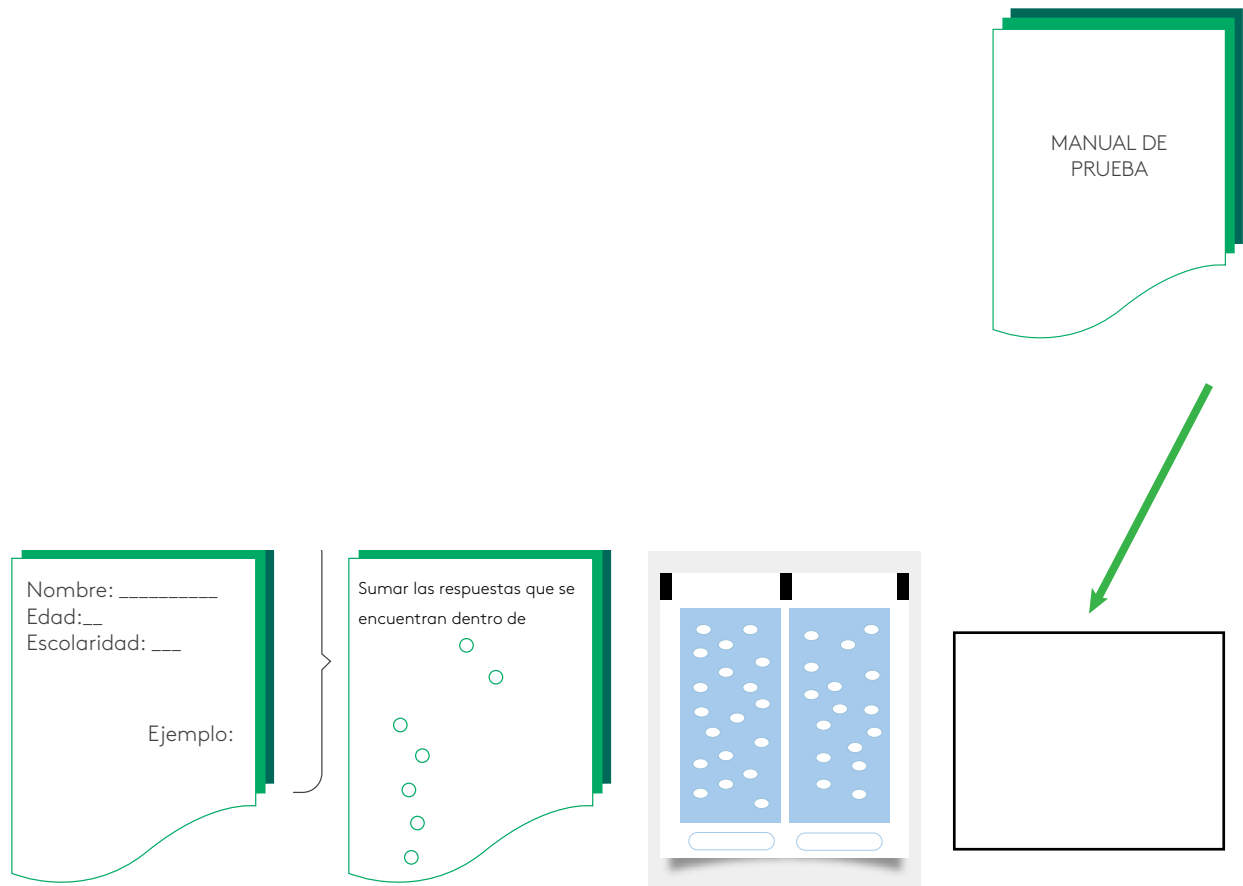
Cuando se habla de procedimiento de corrección, se está haciendo referencia a la transformación de las **puntuaciones directas** obtenidas tras la aplicación del test a **puntajes transformados** -para recordar estas puntuaciones, se puede consultar el material del eje 3.

Se le llama corrección porque se asume que la puntuación obtenida implica la influencia de variables no controladas (error de medición), por lo que se emplean los diversos ajustes estadísticos para controlar o disminuir dicha influencia del error.

Este es un procedimiento que por lo general el psicólogo realiza en ausencia del evaluado. Lo primero que se hace es la extracción de la puntuación directa, recordemos que en algunas pruebas se hace de manera manual con ayuda de plantillas.

Teniendo la **puntuación directa**, se consideran las características del evaluado para buscar en el **manual de prueba** la tabla de **baremos** correspondiente con dichas características, una vez identificada la tabla de baremos correspondiente, se procede a ubicar en ella la puntuación directa del sujeto y su equivalente en **puntuación transformada**, esta se va a analizar en contraste con las medidas de tendencia central (media, desviación estándar) que aparecen en esa misma tabla.

Con esta información se procede, si la prueba lo exige a realizar la gráfica del perfil e indicar lo que implica esta puntuación para el evaluado, por lo general en el manual de prueba se encuentran las descripciones que guían este análisis.



Baremos para mujeres universitarias solteras

PD	Puntuaciones T	.P Percentiles
10	59	80
20	65	90
30	67	95
	$\mu = 50$ $ds = 10$ $N = 180$	

Figura 13. Ejemplo del paso a paso del proceso de corrección de puntajes de una prueba
Fuente: propia

Informe de aplicación de prueba

En este documento se encuentra el verdadero aporte del psicólogo al usar herramientas de medición, en él se presenta el significado de los resultados obtenidos tras la aplicación de un test psicológico.

Contiene, en primer lugar, los datos del evaluado (datos sociodemográficos y aquellos pertinentes según el contexto de aplicación), presenta el objetivo de la aplicación y una breve presentación del instrumento usado, argumentando su elección. Se puede también describir el contexto de aplicación y las variables controladas o aquellas intervinientes.

Posteriormente se presentan los resultados cuantitativos, las normas de interpretación y gráficos de perfil, así mismo, debe contener los resultados cualitativos, la descripción del significado de las puntuaciones baremadas o estandarizadas, obtenidas para finalizar las recomendaciones o plan a seguir con el evaluado.



Instrucción

Ejemplos de informes de aplicación y perfilación pueden ser trabajados en la actividad de refuerzo, taller de informe de aplicación de prueba.



Lectura recomendada

Para profundizar los temas de este eje puede consultar las siguientes lecturas:

Psicometría (p. 27 y 28).

Martínez A. y Hernández L.

Evaluación psicológica: historia fundamentos teórico-conceptuales y psicometría (p. 83 a 86 y 102 a 103).

Aragón, L.

Instrumentos evaluación psicológica (p. 165 a 167 y 219 a 229).

González L.

- Aiken, L. (2003). *Test psicológicos y evaluación*. (11ª edición). Ciudad de México, México: Pearson Educación.
- Anastasi, A. y Urbina, S. (1998). *Test psicológicos*. (7ª edición). Ciudad de México, México: Prentice Hall.
- Aragón, L. (2015). *Evaluación psicológica: historia fundamentos teórico-conceptuales y psicometría*. (2a. ed.). Ciudad de México, México: Editorial El Manual Moderno.
- González, L., F. (2007). *Instrumentos de evaluación psicológica*. La Habana, Cuba: Ciencias Médicas. Capítulo 3, partes II, IV, V, VII. Recuperado de http://newpsi.bvs-psi.org.br/ebooks2010/en/Acervo_files/InstrumentosEvaluacionPsicologica.pdf.
- International Test Commission. (2005). *International Guidelines on Test Adaptation*. Recuperado de www.intestcom.org.
- International Test Commission. (2014). *ITC Statement on the Use of Tests and Other Assessment Instruments for Research Purposes*. V1.2. Recuperado de www.intestcom.org.
- International Test Commission (2015). *International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal*. Recuperado de www.intestcom.org.
- International Test Commission. (2001). International Guidelines for Test Use. *International Journal of Testing*, 1(2), pp. 93-114.
- Martínez, A. M. R., y Hernández, L. M. V. (2014). *Psicometría*. Madrid, España: Difusora Larousse - Alianza Editorial.
- Martínez, R. (1995). *Psicometría: Teoría de los test psicológicos y educativos*. Madrid, España: Síntesis.
- Thomas, P. H. (2015). *Pruebas psicológicas: una introducción práctica* (2a. ed.). Ciudad de México, México: Editorial El Manual Moderno.
- Tornimbeni, S., et ál. (2004). *Introducción a los test psicológicos*. Buenos Aires, Argentina: Brujas.

Esta obra se terminó de editar en el mes de Septiembre 2018
Tipografía BrownStd Light, 12 puntos
Bogotá D.C,-Colombia.



AREANDINA

Fundación Universitaria del Área Andina

MIEMBRO DE LA RED

ILUMNO