

**CONFIABILIDAD**

- Para que un test pueda ser utilizado responsablemente es necesario que cumpla con determinados estándares técnicos.

- La medición psicológica parte de ciertos supuestos fundamentales.
- Uno de éstos expresa que el resultado de la medición es un valor observado que no coincide con el valor verdadero y, en consecuencia, siempre se mide con un margen de error

- Este valor verdadero es un valor teórico, un concepto matemático.
- En términos matemáticos, este valor es la esperanza matemática de la puntuación observada y podría pensarse como la media de las puntuaciones observadas obtenida de infinitas administraciones de un instrumento dado a una persona (Muñiz, 2001).

- Es importante considerar que con el término “error” nos referimos a cualquier variación de las puntuaciones de un test que no sea asimilable a las variaciones en la magnitud del rasgo que está siendo evaluado (por ejemplo, los cambios en la autoeficacia de una persona entre una medición y otra).

- Siempre que medimos un fenómeno, sea éste de naturaleza física o social, es inevitable una cierta dosis de error, debido a imprecisiones del instrumento o a la influencia de las posibles fuentes de variación de las puntuaciones de un test.

- Cuanto mayor es la variabilidad entre las medidas del mismo fenómeno en repetidas ocasiones tanto menor es la confiabilidad, y cuanto menor es la variabilidad mayor la confiabilidad.

# **El concepto de confiabilidad en la teoría clásica de los tests**



- La hipótesis fundamental de la teoría clásica de los tests (TCT) es que la puntuación observada de una persona en un test es una función de dos componentes: su puntaje verdadero. (que es inobservable) y el error de medición implícito en toda medición.
- El postulado esencial de la TCT se expresa como:
$$O_i = V_i + E_i$$
- Esto es, la puntuación observada de un individuo es igual a la puntuación verdadera más el error.

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

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

- La confiabilidad significa la consistencia entre los puntajes de un test obtenidos por los mismos individuos en distintas ocasiones o entre diferentes conjuntos de ítems equivalentes (APA, 1999).
- Obviamente, debemos distinguir entre la falta de consistencia debido a un cambio real en el rasgo medido y las fluctuaciones azarosas de las puntuaciones ocasionadas por cambios en circunstancias personales u otras que afecten la situación de evaluación.

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

**VALIDEZ**

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

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



- Según Martínez Arias (1995), en la validación relacionada con contenido deben realizarse las siguientes operaciones:
- A. Definición del dominio de conocimiento o comportamiento a medir.
- B. Identificación de expertos en ese dominio.
- C. Juicio de los expertos acerca del grado en que el contenido del test es relevante y representativo del dominio.
- D. Procedimiento estadístico para resumir los datos de la fase precedente.

- Cuando se entregan a los jueces los ítems preliminares de un test es conveniente adjuntar una forma estandarizada de calificación.
- Los jueces revisan de manera independiente cada uno de los ítems, utilizando esa guía preestablecida (Herrera Rojas, 1998).

- La validación de un test está vinculada a la muestra particular que se ha empleado en ese proceso. Esta importante limitación intenta ser atenuada por los procedimientos de generalización de la validez.
- El más empleado en la actualidad es el meta-análisis, método que permite integrar y combinar los resultados de diversos estudios empíricos mediante técnicas específicas.
- En lo que se refiere a la validez, la utilización de este método surgió como respuesta a los coeficientes débiles de correlación observados en muchas investigaciones en relación con la evidencia de las relaciones test-criterio (Martínez Arias, 1995).

# **METODO PARA EVALUAR LA VALIDEZ**

- El meta-análisis es útil para este propósito de generalización al agregar estudios bien diseñados pero obstaculizados por una muestra pequeña y al promediar efectos falsos que operan en ambas direcciones (Nunnally y Bernstein, 1995). Es considerado en la actualidad como una metodología potente de investigación que permite integrar los resultados de diversos estudios empíricos y sus descubrimientos particulares.

- El meta-análisis transforma los resultados estadísticos de investigaciones empíricas independientes a una métrica común, provee una estimación simple de la fortaleza de la relación entre determinadas variables y permite comprobar estadísticamente si una serie de estudios, conjuntamente considerados, apoyan o refutan las hipótesis de investigación (Multon, Brown y Lent, 1991).

- Este método comienza con la reunión, clasificación y codificación de las investigaciones existentes sobre un tema.
- Dicho proceso implica la consideración, clasificación y codificación de las características sustantivas y metodológicas de las investigaciones particulares (tales como tipo y duración de una intervención o tratamiento experimental, tipo de muestra e instrumentos empleados).
- El propósito de la codificación de los estudios particulares es verificar si los resultados difieren en función de las características de los mismos.

- Debido a que los resultados de las investigaciones podrían ser difíciles de comparar directamente, se los debe transformar en una medida común.
- Las dos medidas más utilizadas para cuantificar e integrar los resultados de las investigaciones independientes son los niveles de significación y las medidas de tamaño del efecto.
- El nivel de significación informa si los resultados obtenidos han ocurrido probablemente por azar, mientras que el tamaño del efecto indica la intensidad de la relación o del efecto de interés (Gómez Benito, 1987).



# **VALIDEZ DE CONSTRUCTO**

- El constructo viene a ser un concepto hipotético que forma parte de las teorías que intentan explicar la conducta humana: inteligencia, creatividad, dependencia de campo, etc.
- La validez de constructo es la obtención de evidencias que apoyan que las conductas observadas en un test son (algunos) indicadores del constructo.

- El proceso de validación de constructo implica a partir del establecimiento de deducciones de la teoría:
- a) Formular hipótesis y relaciones entre elementos del constructo, de éste con otros constructos de la teoría y con otros constructos externos.
- b) Seleccionar ítemes o tests (indicadores) que representen manifestaciones concretas del constructo.
- c) Recogida de datos.
- d) Establecer consistencia entre datos e hipótesis, y examinar el grado en que los datos podrían explicarse mediante hipótesis alternativas

- e) Hay diversos procedimientos para establecer la validez de constructo. Si elaboramos una “miniteoría” esta tendrá tres pasos:
- (1) en base a la teoría sostenida en ese momento respecto del test, el psicólogo deduce ciertas hipótesis sobre la conducta esperada de las personas que obtienen puntajes diferentes en el test,
- (2) se reúne datos que confirman o no esas hipótesis,
- (3) en base a los datos acumulados, se toma la decisión relativa a si la teoría explica adecuadamente los datos. Si no es así se tiene que revisar la teoría y repetir el proceso hasta lograr una explicación más adecuada.

- El proceso de validación, en ese sentido, es de continua reformulación y refinamiento. Al determinar la validez de construcción, el propósito es identificar todos los factores que influyen en la ejecución del test y determinar el grado que influyen cada uno de ellos.

**Validez de criterio.**

- Validez Predictiva (evidencia externa)
- ¿Predicen las puntuaciones del test un rendimiento o conducta futura? (Junto con la validez concurrente se le denomina también validez empírica del test).
- Un uso común de los tests es predecir la conducta futura; utilizamos el test para ayudarnos a tomar alguna decisión práctica (selección, clasificación, etc.).

- Por ejemplo, el test será un componente aceptable de un proceso de selección de personal, si sus calificaciones o puntuaciones predicen la ejecución de algún componente importante del trabajo (criterio externo); en otras palabras, para que el test se pueda utilizar como parte de un proceso de selección es preciso demostrar la validez de la prueba relacionándola con los criterios pertinentes.



- Un criterio es cualquier desempeño que los sujetos tienen en la vida real, por ejemplo, las medidas de rendimiento académico, medidas de rendimiento laboral, clasificaciones psiquiátricas, etcétera

**Validez de contenido.**

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

- La validez de contenido consiste en determinar lo adecuado del muestreo de reactivos del universo de reactivos posibles; en este sentido, es una “medida” de lo adecuado del muestreo. Ponemos “medida” entre comillas debido a que este tipo de validez consiste en una serie de estimaciones u opiniones, que no proporcionan un índice cuantitativo de validez (para su obtención no se utiliza procedimientos estadísticos).