
Validación Psicométrica Central Test



El arte de la evaluación

Propiedades Psicométricas de los Tests

La evaluación periódica de la validez, así como el ajuste de los tests, son elementos esenciales para la utilidad a largo plazo y la efectividad de las pruebas psicométricas. Central Test pone atención particularmente a estos elementos, revisando regularmente las propiedades psicométricas de todos sus útiles. Estas revisiones implican una actualización continua de contenido, análisis estadísticos y estudios de validez. Antes de ser publicados en línea, nuestros tests son evaluados según cuatro criterios que se presentan a continuación:

Fiabilidad o Fidelidad de un test

La fidelidad se refiere a la manera en que un test mide una dimensión determinada. Si una persona realiza de nuevo el test ¿tendría el mismo resultado o un resultado muy diferente? Un test que proporciona puntuaciones similares cuando una persona lo repite varias veces es considerado como un test fiable. Sin embargo, hay que tomar en cuenta que la medida en las ciencias del comportamiento siempre es influenciada por variables externas que son:

- **El estado psicológico o físico de la persona que realiza el test.** Por ejemplo, los diferentes niveles de fatiga, ansiedad, o motivación pueden influenciar el desempeño del individuo y por lo tanto sus resultados.
- **Los factores ambientales.** Las diferencias en el ambiente y el lugar en donde se realiza el test, como la luz, el ruido, pueden influenciar el desempeño de un individuo durante la prueba.

Estos factores son considerados como fuente de “errores de medida aleatorio”. Si estos errores de medida aleatorios no existieran, un individuo tendría exactamente la misma puntuación en el test, su “puntuación verdadera”, en cada repetición y bajo cualquier circunstancia. Un índice de la fiabilidad del test es la parte de la puntuación del test que no es afectada por el error de medida aleatoria.

Existen diversos tipos de estimación de la fidelidad, cada uno influenciado por diferentes fuentes de errores de medida. Los tests desarrollados por Central Test ponen atención en dos tipos de fidelidad, particularmente importantes:

- **la consistencia interna:** indica si los ítems de un test miden la misma dimensión. Un coeficiente de consistencia interna elevado indica que los ítems del test son similares en su contenido (es decir homogéneos). Es importante tomar en cuenta que la extensión del test puede afectar la consistencia interna. Por ejemplo, un test muy largo puede aumentar el coeficiente de fidelidad de forma errónea. La consistencia interna es comúnmente medida por un Alpha de Cronbach que varía entre 0 (bajo) y 1 (elevado).
- **la fidelidad test-retest:** indica el grado de repetición de la puntuación de un test con el paso del tiempo. Además, esta estimación refleja la estabilidad de las características o del constructo evaluados por el test. Algunos constructos son más estables que otros. Por ejemplo, en un mismo lapso de tiempo, la capacidad de lectura de un individuo es más estable que su nivel de ansiedad. Para los constructos que suponen variar con el tiempo, un coeficiente de fidelidad test-retest aceptable puede ser más bajo que el de los constructos estables en el tiempo.

Como creadores de tests, en Central Test tomamos la responsabilidad de hacer un reporte sobre las estimaciones de la fidelidad pertinentes para un test en particular. El nivel de fidelidad aceptable depende del tipo de test y del tipo de coeficiente utilizado. Las medidas de fidelidad son proporcionadas en los documentos técnicos de los tests.

Validez del Test

La validez es el criterio más importante para la elección de un test. Hace referencia a las *características* que mide el test y a la *exactitud* con las que evalúa. La validez da sentido a las puntuaciones del test. Las pruebas de validez indican que existe un vínculo entre el desempeño en el test y el desempeño en el puesto de trabajo. Así, nos informa el grado en el que es posible obtener predicciones o conclusiones específicas de la persona, en función de la puntuación obtenida en el test. En otras palabras, la validez informa sobre la utilidad del test. Por ejemplo, si un test es un predictor válido del desempeño en un puesto de trabajo determinado, nosotros podemos concluir que si una persona presenta una puntuación elevada en el test, es propensa a ser más competente en ese puesto que una persona que ha obtenido una puntuación baja, aún teniendo características similares.

Es importante comprender la diferencia entre la fidelidad y la validez. La validez nos dice qué tan bueno es un test para evaluar una situación particular; la fidelidad nos dice a qué punto el test es creíble y estable. La validez de un test es establecida en referencia a un objetivo en particular.

Existen muchos tipos de validez que han sido conceptualizados por investigadores y científicos. Éstos pueden ser agrupados en tres categorías:

- **La validez de criterio:** es evaluada examinando la correlación u otra relación estadística entre el desempeño del test y otro criterio externo de desempeño. En otras palabras, los individuos que presentan una puntuación elevada en el test tienden a tener un mejor desempeño en ese criterio externo que los individuos que tienen una puntuación baja en el test. Si el criterio es obtenido al mismo tiempo que se realiza el test, hablamos de validez concurrente. Si el criterio es obtenido más tarde, hablamos de validez predictiva.

- **La validez de contenido:** se evalúa examinando si el contenido del test es representativo del constructo que queremos medir. Por ejemplo, para un test de evaluación de potencial comercial, los ítems del test deben ser pertinentes y evaluar directamente las exigencias y cualidades para cierto puesto comercial. El contenido del test debe igualmente corresponder al nivel de comprensión de una población determinada.
- **La validez del constructo:** requiere la demostración que el test mida la característica que pretende medir. Por ejemplo, un test para los intereses en el trabajo debe evaluar los intereses en el trabajo y no el nivel de motivación. La validez del constructo es generalmente evaluado de dos formas distintas:
 - Presentando el test a un panel de expertos y pedir su opinión acerca la correspondencia entre los ítems del test (el contenido) y el constructo evaluado (**validez aparente**).
 - Administrando el test junto a otros tests desarrollados con constructos teóricamente similares y examinando la correlación entre los dos (**validez convergente**); o bien, administrando el test al mismo tiempo que otros test teóricamente opuestos, examinando su correlación (**validez divergente**).

Los tres tipos de validez -de criterio, de contenido y de constructo- son utilizados como base de la validación. Estos tres métodos se pueden presentar al mismo tiempo; uno o más métodos podrían ser apropiados dependiendo de la situación. En Central Test, los estudios de validación de nuestros tests son llevados a cabo de manera interna y con la ayuda de expertos. Los resúmenes de estos estudios se encuentran en los documentos técnicos de cada test y son utilizados para un mejorarlos constantemente.

Deseabilidad social

La deseabilidad social es un sesgo en la evaluación que corresponde a la inclinación de una persona a responder de tal manera que sea vista como favorable por otros. Ya que somos sociales por naturaleza, todo individuo se inclina a buscar un cierto grado de aceptación social. La deseabilidad social es uno de los principales sesgos de los test de personalidad; pero es aún más prominente en el contexto del reclutamiento. Un candidato es normalmente tentado a responder de manera que pueda impresionar positivamente al reclutador.

Para nuestros tests, hemos desarrollado una estrategia en dos puntos principales para sobrellevar el problema. En primer lugar, todos los ítems son examinados con el fin de asegurar que no son sensibles a la deseabilidad social. En segundo, hemos incluido una escala de deseabilidad social dentro de los tests, con el fin de identificar el posible engaño.

Consideraciones socio-legales

Un buen test psicométrico no debe inclinarse (positivamente o negativamente) hacia un grupo socio-cultural en particular. No debe mostrar ningún tipo de discriminación en base a la religión, género, origen étnico o cultural de la persona que pasa el test.

Desde el inicio del desarrollo de nuestros tests, todos los ítems son examinados para que no favorezcan a una población particular. En los estudios de validación, las relaciones entre las variables demográficas y las puntuaciones son examinadas con el fin de evitar una posible inclinación. Esto nos permite construir tests sensibles y evitar preferencia por cierta población.