

# Confiabilidad y Validez



Pearson  
TalentLens

## Puntuación de la Confiabilidad

La confiabilidad de una prueba es la medición de la consistencia de las puntuaciones; es decir en qué medida dos personas con la misma habilidad o la misma persona puesta a prueba en diferentes ocasiones recibirá la misma puntuación. La confiabilidad se expresa como un coeficiente que va desde cero a uno. Cuanto más cerca está el coeficiente de confiabilidad a 1.00, más confiable será la prueba y menor será el error de cálculo asociado con la puntuación de pruebas. Cuando las pruebas se utilizan en contextos laborales, los coeficientes de confiabilidad que estén por encima de .89, en general, son considerados excelentes; .80-.89, buenos y .70-.79, adecuados. Los valores por debajo de .70 indican que la prueba podría tener una aplicabilidad más limitada. Por ejemplo, podría ser usado para dar una retroalimentación para el desarrollo, pero no sería apropiado para tomar decisiones de selección o promoción.

Pueden usarse varios métodos para medir la confiabilidad de la prueba. Esto incluye la consistencia interna de los ítems de la prueba (por ejemplo, coeficiente alfa de Cronbach y el método de división por mitades), la confiabilidad test-retest (la estabilidad de las puntuaciones de la prueba en el tiempo) y el análisis de las formas paralelas (la consistencia del puntaje a lo largo de las formas paralelas de una prueba).

## Validez

La validez se refiere a la medida en la que cierta información específica, como la investigación o la teoría, permite la interpretación de la puntuación de la prueba (AERA et al., 1999). "En la medida en que la prueba brinde la información necesaria según la persona a tomar la decisión, su validez será alta." (Cronbach, 1970). La validez del constructo es el grado en que una prueba mide el marco teórico o significado para la cual fue diseñada. La validez de criterio es la relación entre la medición y el resultado de la prueba. Por ejemplo, el desempeño en un empleo o aprobar/reprobar un curso.

No existe una norma para su interpretación

- Por debajo de 0.2 - poco probable que sea eficaz
- 0.2 - 0.35 - eficaz
- 0.35 o más - altamente eficaz

Los valores serán menores a la estimación de la confiabilidad por las siguientes razones:

- El desempeño laboral es difícil de pronosticar
- El desempeño laboral no se mide correctamente
- Generalmente hay restricciones en los rangos



## Watson Glaser III

### Confiabilidad

**Consistencia Interna:** .82 -.86 (Muestras del Reino Unido y EE.UU.) Test-Retest: Actualmente en progreso. Estudios anteriores muestran correlaciones de .73 -.89.

**Forma Paralela:** .80 (Muestras del Reino Unido y EE.UU)

### Validez del constructo

Estudios demostraron que Watson-Glaser tiene correlación con la medición de otras habilidades cognitivas:

- Correlación con las pruebas de rendimiento que oscila entre .39 y .51
- Correlación con otras pruebas de razonamiento que oscilan entre .47 y .70
- Correlación con pruebas de Coeficiente Intelectual que oscilan entre .41 y .60
- Correlaciones moderadas con el Test de Matrices Progresivas de Raven [Raven's APM por sus siglas en inglés] que oscilan entre .37 y .53

### Validez de criterio

La prueba Watson-Glaser y El curso de preparación Profesional del Colegio de Abogados [the Bar Professional Training Course]

- En el 2011, 1501 estudiantes en el curso completaron la prueba Watson-Glaser III.
- Se recopilaron las calificaciones de la evaluación final al final del curso.
- La correlación entre los resultados del curso y los de Watson-Glaser fue de .51.
- Esta investigación se actualizó en el 2015 utilizando el mismo método con 998 estudiantes.
- Esta información confirmó una vez más la fuerte correlación que existe entre la prueba y el desempeño del curso con la cifra .55.



## Ravens

### Confiabilidad

- Test de Raven-III consistencia interna: .73 (Muestra internacional de 466 aspirantes a un curso de educación superior en el Reino Unido en el 2015).
- Test de Raven-II consistencia interna: de .74 a .84, según la muestra
- Consistencia interna del Test de Raven: de .83 a .87, según la muestra
- Confiabilidad del test-retest del Test de Raven: .91 (n=243, intervalo de prueba de seis a ocho semanas)

### Validez del constructo

Años de estudios previos sobre la versión del Test de Raven de 36 ítems (antes estaba reducido a 23 ítems) respaldan su validez del constructo.

- En una muestra a 149 aspirantes a la educación superior, los puntajes del Test de Raven analizados fueron correlativos en .56 con aquellos obtenidos en matemática en el American College Test (Koenig, Frey, & Detterman, 2007).
- En un estudio realizado a 104 estudiantes universitarios, Frey y Detterman (2004) informaron que los puntajes del Test de Raven tenían una correlación de .48 con los del examen SAT (The Scholastic Assessment Test). Las pruebas de la validez del constructo de la versión actual del Test de Raven están respaldadas por diversos resultados previos.
- En un subconjunto de 41 encuestados de la muestra estandarizada, los puntajes del Test de Raven analizados tuvieron una correlación de .54 con los obtenidos en la prueba Watson-Glaser (Watson-Glaser, 2006)
- Asimismo, en una muestra N=276 del Test de Raven la correlación fue de  $r = .51$  con la puntuación total de la Valoración del Razonamiento Numérico Avanzado [Advanced Numerical Reasoning Appraisal -ANRA]; una evaluación de la capacidad cognitiva que mide el razonamiento cuantitativo.

### Validez de criterio

Los estudios que utilizaron el Test de Raven en los últimos 70 años muestran evidencia de su validez de relación de criterio:

- Chan (1996) – Un estudio de validación de centros de evaluación demostró que las puntuaciones en el Test de Raven tuvieron correlación con la calificación de los participantes en “iniciativa/ creatividad” ( $r = .28$ ).

- Gonzalez, Thomas and Vanyukov (2005) - informaron una relación positiva entre las puntuaciones del Test de Raven y el desempeño en la toma de decisiones.
- Fay y Frese (2001) - descubrieron que las puntuaciones del Test de Raven estaban "asociadas consistentemente y de manera positiva con un aumento de la iniciativa personal a lo largo del tiempo".



## Prueba de interpretación de datos numéricos NDI

### Confiabilidad

Consistencia Interna; .76-.78 (Muestras del Reino Unido y EE.UU.)

### Validez del Constructo

Se comparó a la NDI con la Prueba de Razonamiento Numérico de Gestión Empresarial & Graduados [Reasoning for Business Managerial & Graduate—Numerical] (RfB MG Numerical; Cubiks, 2009). La correlación entre ambas pruebas fue de: .71

### Validez de Criterio

NDI y el Desempeño Profesional

- Muestras de 104 participantes en EE.UU.
- Evaluó su desempeño profesional general y su desempeño en los trabajos que requieren "matemática".

**Muestra total:** todos los participantes

Aquellos que usan matemática: informaron que a veces las utilizan en sus empleos, en contraste con pocas veces o nunca.

Grupo	N	Desempeño Profesional General	Desempeño en relación a matemática
Muestra Total	104	.19	.52**
Los que utilizan matemática	71	.32**	.50**



## SOSIE

### Internal Consistency (UK)

Los valores de confiabilidad oscilan entre .75 (Responsabilidad y Cautela) y .83 (Dominio) para las escalas de Personalidad, y entre .61 (Convicción) y .87 (Variedad) para las escalas de valores.

### Consistencia Interna (Muestra Internacional)

Los indicadores alfa oscilaron entre .62 (Convicción) y .85 (Variedad). Dos escalas obtuvieron un alfa <.70 (Materialismo y Convicción). Los indicadores alfa de las Escalas de Personalidad oscilaron entre .75 (Vigor) y .81 (Dominio y Sociabilidad). Los indicadores alfa de las Escalas de Valores Interpersonales oscilaron entre .70 (Apoyo) y .81 (Conformidad), y los indicadores alfa de las Escalas de Valores Personales oscilaron entre .62 (Convicción) y .85 (Variedad).

### Test Retest (Reino Unido)

Los coeficientes para las Escalas de Personalidad oscilaron entre .66 (Responsabilidad y Sociabilidad) y .86 (Cautela) y para las Escalas de Valores, entre .53 (Reconocimiento y Orden) y .82 (Conformidad).

### Test Retest (Muestra Internacional)

Los coeficientes para las Escalas de Personalidad oscilaron entre .78 (Responsabilidad y Sociabilidad) y .84 (Dominio y Cautela) y entre .55 (Convicción) y .79 (Variedad) para las Escalas de Valores.