

# Fiabilidad, validez y sensibilidad.

Evaluación Neurológica. Tema 1.  
Máster U. Neurocontrol Motor

©2024. Autor: Francisco Molina-Rueda

Algunos derechos reservados

Este documento se distribuye bajo la licencia  
“[Atribución-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/deed.es)”  
de Creative Commons, disponible en:

<https://creativecommons.org/licenses/by-sa/4.0/deed.es>

# Índice de contenidos

---

Fiabilidad - Reliability

---

Validez - Validity

---

Sensibilidad -  
Responsiveness

**Consulta:** <https://www.cosmin.nl/>



 Find the right tool ▾

 COSMIN tools ▾

 Courses & Resources

## COSMIN helps you select the most suitable outcome measurement instruments

COSMIN aims to improve the selection of outcome measurement instruments both in research and in clinical practice by developing methodology and practical tools for selecting the most suitable outcome measurement instrument.

### Make your research better

You can use our tools to improve the way you do research and the trustworthiness of your results.

[LET US HELP YOU FIND THE RIGHT TOOL >](#)

## What are you working on?

To find the right tool for your research, tell us what you are working on:

- > I'm deciding what to measure
- > I'm looking for available outcome measurement instruments
- > I want to select the most suitable outcome measurement instrument
- > I'm conducting a study on measurement properties
- > I'm conducting a systematic review of outcome measurement instruments
- > I'm developing a Core Outcome Set

# Fiabilidad

Aborda el grado en el que una puntuación está libre de error de medición

**Evalúa si las puntuaciones obtenidas al evaluar los pacientes son consistentes bajo diversas circunstancias:**

- 1. Consistencia Interna:** interrelación entre los ítems o dominios de una escala
- 2. Fiabilidad:** proporción de la varianza total en las mediciones debida a las diferencias "verdaderas" entre los pacientes
  - **Test – retest.** A lo largo del tiempo
  - **Interevaluador – Interrater.** Por diferentes personas en la misma ocasión
  - **Intraevaluador – Intrarater.** Por la misma persona en distintas ocasiones

# Fiabilidad

- Interpretación estadística: **Coeficiente de Correlación Intraclassa (CCI)**

## Estimación del valor “r”:

- > 0,90 – EXCELENTE
- Entre 0,76 y 0,90 – BUENO
- Entre 0,5 y 0,75 – MODERADO
- < 0,5 - POBRE



- Interpretación estadística: **Cambio Mínimo Detectable (CMD)**: magnitud de la variación del valor de cada escala por debajo de la que ese cambio puede ser interpretado como inherente a la variabilidad del propio método de valoración, sin que haya existido un verdadero cambio en la situación clínica del paciente

# Fiabilidad



PM R xx (2019) 1-7



www.pmrjournal.org

Original Research

## Reliability and Minimal Detectable Change in the Gait Assessment and Intervention Tool in Patients With Multiple Sclerosis

María D. Gor-García-Fogeda, PhD , Silvia Tomé-Redondo, PT,  
Cristina Simón-Hidalgo, PT, Janis J. Daly, PhD, Francisco Molina-Rueda, PhD,  
Roberto Cano-de-la-Cuerda, PhD 

CCI para la fiabilidad intraobservador = 0,91; Intervalo de Confianza del 95% (IC95%) = 0,85-0,95

CCI para la fiabilidad interobservador = 0,93; IC 95% = 0,88 – 0,96

CMD = 1,19 puntos (de la escala) para la fiabilidad intraobservador

# Validez

Evalúa el grado en el que un instrumento mide el constructo que pretende puntuar

1. **Contenido:** grado en el que el contenido del instrumento refleja el constructo que pretende evaluar - **Validez de Apariencia** (ídem validez de contenido)
2. **Constructo:** grado en que las puntuaciones de un instrumento son coherentes con las hipótesis de evaluación (relación interna, relación con otros instrumentos, diferencias entre grupos)
  - **Estructural** (ídem validez de constructo: relación interna de las secciones de la escala)
  - **Prueba de hipótesis** (ídem validez de constructo: relación con otros constructos)
  - **Transcultural.** Contenido traducido o adaptado culturalmente refleja adecuadamente el contenido original
3. **Criterio:** adecuación de las puntuaciones de un instrumento respecto a una prueba de referencia “gold standard”

# Validez

- Interpretación estadística: **Coefficiente de correlación de y de Spearman**

**Estimación del valor “r” / “rho”:** entre -1 y 1

Correlación negativa: las escalas puntúan de forma inversa. *Ej.: Berg Balance Scale (BBS) vs. Wisconsin Gait Scale (WGS). La puntuación máxima de la BBS significa buen equilibrio; puntuación máxima en la WGS significa peor patrón de marcha*

Correlación positiva: las escalas puntúan en el mismo sentido. *Ej.: BBS vs. Tinetti. En ambas escalas la máxima puntuación significa buen funcionamiento*

Valores:  $\geq 0,80$  excelente; 0,50-0,79 moderado; 0-0,49 pobre

# Validez

Gait & Posture 68 (2019) 363–368

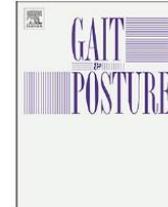


ELSEVIER

Contents lists available at [ScienceDirect](#)

Gait & Posture

journal homepage: [www.elsevier.com/locate/gaitpost](http://www.elsevier.com/locate/gaitpost)



Full length article

Construct validity of the Wisconsin Gait Scale in acute, subacute and chronic stroke



C. Estrada-Barranco<sup>a,b</sup>, R. Cano-de-la-Cuerda<sup>c,\*</sup>, F. Molina-Rueda<sup>c</sup>

**Objetivo del trabajo: estudiar la validez de constructo de la escala WGS en pacientes con ictus crónico**

**Constructos: marcha, equilibrio y funcionalidad**

**Resultados: WGS vs. BBS →  $r = -0,908$ ; IC95% (-0,94)-(-0,86)**

# Sensibilidad

Capacidad de un instrumento para evaluar cambios a lo largo del tiempo

Evaluar la sensibilidad del instrumento de evaluación antes y después de una intervención. Dos formas de cálculo:

**Métodos basados en la distribución (distribution-based methods):** métodos estadísticos y capacidad de detectar el cambio:

- **Error estándar de la media y cambio mínimo detectable (visto en fiabilidad)**
- **Efectos suelo y techo:** porcentaje de la muestra que obtuvo las puntuaciones mínimas o máximas posibles, respectivamente. Según varios autores, los efectos de techo y suelo del 20% o más se consideran relevantes

# Sensibilidad

**Métodos basados en un criterio (anchor-based methods):** basado en un criterio externo para determinar si los cambios son clínicamente significativos

**Criterio externo:** *global rating of change (GRC), caídas, superar el punto de corte en una escala, etc.*

→ a través del criterio externo se puede obtener el **Cambio Mínimo Importante** de una escala.

GRC: es un instrumento diseñado para cuantificar la mejora o deterioro de los pacientes a lo largo del tiempo.

Pide al paciente que califique su cambio con respecto a una condición particular durante un período de tiempo específico. Ejemplo: *¿Con respecto a su dolor lumbar, ¿cómo se describiría ahora en comparación con su primera sesión?*

Escala entre -7 (mucho peor) – 0 (sin cambios) – 7  
(completa recuperación)

Pequeño cambio (GRC ≤ 3), cambio moderado (3-5),  
cambio importante (GRC ≥ 5)  
GRC > 3: cambio importante

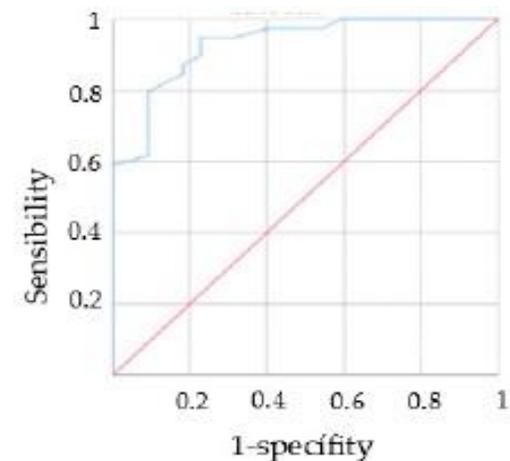
**Lectura sobre Cambio Mínimo Importante.** Godi M, Franchignoni F, Caligari M, Giordano A, Turcato AM, Nardone A. Comparison of reliability, validity, and responsiveness of the mini-BESTest and Berg Balance Scale in patients with balance disorders. Phys Ther. 2013 Feb;93(2):158-67. doi: 10.2522/ptj.20120171. Haz [click](#).

La **curva ROC** representa la sensibilidad (razón de verdaderos positivos) y el complementario de la especificidad (razón de falsos positivos) para cada puntuación de una prueba para detectar o predecir una variable resultado (por ejemplo, mejora o no mejora, caída o no caída). **Rendimiento diagnóstico de un instrumento.**

Sensibilidad: capacidad de un punto de corte de una escala para detectar o clasificar los casos positivos de forma correcta (por ejemplo: personas que se han caído o que perciben mejoría -  $GRC > 3$ ), de entre todos los casos positivos disponibles durante la prueba (verdaderos positivos y falsos negativos).

**Complementario de la especificidad**: proporción de falsos positivos para un punto de corte concreto de la escala entre todos los casos negativos de la prueba (verdaderos negativos y falsos positivos).

Área bajo la curva (AUC): posee un valor comprendido entre 0,5 y 1, donde 1 representa un valor diagnóstico perfecto y 0,5 es una prueba sin capacidad discriminatoria diagnóstica. Si  $AUC = 0,8$  significa que existe un 80% de probabilidad de que el diagnóstico realizado a un enfermo sea correcto



**Puedes consultar esta figura en:** Estrada-Barranco C, Sanz-Esteban I, Giménez-Mestre MJ, Cano-de-la-Cuerda R, Molina-Rueda F. Predictive Validity of the Postural Assessment Scale for Stroke (PASS) to Classify the Functionality in Stroke Patients: A Retrospective Study. *J Clin Med.* 2022 Jun 29;11(13):3771. doi: 10.3390/jcm11133771. PMID: 35807054; PMCID: PMC9267227.

Los autores tratan de comprobar la capacidad predictiva de la escala PASS respecto a la funcionalidad de los pacientes dado un punto de corte que indica buena o mala funcionalidad.

**Consulta:** <https://www.sralab.org/rehabilitation-measures>

[g & COVID-19 Precautions](#) | [TeleHealth Visits](#)



# Fiabilidad, validez y sensibilidad.

Evaluación Neurológica. Tema 1.  
Máster U. Neurocontrol Motor

©2024. Autor: Francisco Molina-Rueda.  
Algunos derechos reservados.

Este documento se distribuye bajo la licencia  
[“Atribución-CompartirIgual 4.0 Internacional”  
de Creative Commons](https://creativecommons.org/licenses/by-sa/4.0/deed.es), disponible en:  
<https://creativecommons.org/licenses/by-sa/4.0/deed.es>