

# Desarrollo de soluciones de formación innovadoras en el campo de la evaluación funcional destinadas a actualizar los planes de estudio de las escuelas de ciencias de la salud



Módulo de Biomecánica en la Marcha Humana  
Unidad Didáctica C: ¿Cómo evaluar la marcha?

C.3: ¿Cuáles son las ventajas del uso de técnicas instrumentales frente a las escalas y el examen físico en la evaluación de la marcha?



## Índice

1. INTRODUCCIÓN Y OBJETIVOS	2
2. CARACTERÍSTICAS Y PROPIEDADES DE LAS HERRAMIENTAS DE EVALUACIÓN DE LA MARCHA: COMPARACIÓN ENTRE TÉCNICAS DISPONIBLES:	3
2.1 Usabilidad.....	3
2.1 Requisitos del equipo.....	5
2.3 Objetividad de los resultados y análisis estadístico.....	8
2.4 Validez.....	11
2.5 Fiabilidad.....	14
2.6 Sensibilidad al cambio y Capacidad de respuesta.....	22
2.7 Efecto suelo y techo.....	26
3. IDEAS CLAVES	28
4. REFERENCIAS	29

## 1. Introducción y Objetivos

---

Tanto la evaluación de la marcha a través de 1) observación, 2) escalas clínicas, pruebas y cuestionario, y 3) la evaluación mediante técnicas instrumentales son útiles para conocer el patrón de marcha de una persona, pero existen diferencias fundamentales que hay que tener en cuenta a la hora de elegir uno u otro tipo de técnica de valoración o una combinación de ambos.

En esta unidad revisaremos las características metodológicas y propiedades estadísticas de las técnicas disponibles para la evaluación de la marcha humana revisadas en la Unidad C.1 del Módulo Biomecánica de la Marcha.

En base a esto, proponemos los siguientes objetivos:

1. Revisar las ventajas y desventajas de las metodologías de valoración de la marcha humana.
2. Conocer las propiedades estadísticas de las metodologías de evaluación de la marcha disponibles.
3. Establecer los conocimientos técnicos que permitan a los profesionales sanitarios elegir la técnica de valoración de la marcha más adecuada a su contexto clínico o de investigación.

## 2. Características y propiedades de las herramientas de evaluación de la marcha: comparación entre técnicas disponibles:

### 2.1 Usabilidad

La usabilidad se refiere a la facilidad con la que las personas pueden usar una herramienta en particular para lograr un objetivo específico. Ante el ejercicio de la profesión de cualquier perfil sanitario y en cualquier ámbito, antes de elegir una técnica de evaluación (o tratamiento), los profesionales se plantearán preguntas como: ¿es fácil de utilizar? ¿requiere mucho tiempo? ¿es factible usarlo en mi área de trabajo? En este sentido, las técnicas de evaluación de la marcha humana disponibles (revisadas en unidades anteriores) tienen diferentes respuestas a las preguntas planteadas anteriormente.

En cuanto a la facilidad de uso, las técnicas instrumentales requieren varios pasos enmarcados en un estricto protocolo de evaluación que se debe seguir para asegurar que la comparación entre las mediciones de los propios datos del paciente (por ejemplo, cuando se realizan varias evaluaciones durante un período de tiempo) o con los datos de otros sujetos, son válidos. Esto hace que cualquier técnica instrumental sea más compleja de utilizar que una escala, un cuestionario o una prueba clínica (tabla 1). Del protocolo de medida con técnica instrumental, probablemente el paso más complejo de gestionar es la instrumentación del sujeto si es necesario y el tratamiento posterior de los datos tras realizar la medida. Por tanto, teniendo en cuenta lo anterior, las técnicas de evaluación instrumental de la marcha requieren de un factor muy relevante a tener en cuenta: el tiempo (Tabla 1). Sin embargo, con el avance de las tecnologías disponibles, dentro de un grupo de técnicas instrumentales para medir los mismos parámetros, podemos encontrar diferentes escenarios.

#### Ejemplo

#### **Evaluación biomecánica instrumental de los parámetros espacio-temporales de la marcha**

Usando un sistema de fotogrametría y una pasarela instrumentada, es posible medir los resultados espacio-temporales que caracterizan la marcha humana. Estos parámetros son principalmente la velocidad de la marcha, la longitud de la zancada y la longitud del paso, la cadencia, el tiempo de doble apoyo y el tiempo de la fase de apoyo y balanceo del ciclo de la marcha. Aunque con ambas técnicas podemos llegar al mismo resultado, el tiempo que tarda cada una de las técnicas varía significativamente.

Con la pasarela instrumentada (con sensores de presión) no es necesario instrumentar al sujeto ya que la persona valorada solo debe caminar por el pasillo con o sin calzado. Si el software de gestión del equipo calcula los parámetros a partir de las presiones plantares registradas, no será necesario ningún tratamiento adicional para medir los datos. Así, el

evaluador no necesita una formación o conocimientos excesivos para utilizar la herramienta.

Por otro lado, con un sistema de fotogrametría, la instrumentación del sujeto siempre es necesaria. Esta instrumentación consiste en un modelo biomecánico de hitos/items que se colocan en el cuerpo para representar los segmentos involucrados en el movimiento a evaluar. Dependiendo de la complejidad del modelo, el tiempo de instrumentación será mayor o menor. En cualquier caso, la piel donde se coloca el marcador siempre debe limpiarse y afeitarse. Por otro lado, el software de gestión de los sistemas de fotogrametría suele dar como resultado las coordenadas de cada marcador en cada eje de movimiento y en cada tiempo de captura de imagen (por ejemplo, 50 posiciones por segundo si la frecuencia de grabación es 50 Hz). Salvo que un software ofrezca un protocolo con modelo cerrado y fijo y calcule previamente los diferentes parámetros que se pueden obtener de la medición, siempre habrá que realizar un cálculo de resultados después de extraer las coordenadas de cada hito del modelo biomecánico. Todo esto requiere que el evaluador tenga un conocimiento más profundo y practique la correcta colocación de los marcadores.

**Pasarela instrumentada**



**Sistema de fotogrametría**



Instrumentación del sujeto:	NO	SI
Procesamiento de datos tras medir	NO	SI
Entrenamiento del evaluador	NO	SI
Tiempo aproximado	5 minutos	1 hora

Por otro lado, el análisis de la marcha de observación o la valoración por escalas clínicas requiere un tiempo mínimo y el entrenamiento necesario para la evaluación de la marcha está relacionado con la práctica de las instrucciones de la propia escala o prueba. Es por todo lo anterior, que el uso de cada técnica de medición tiene más usabilidad en diferentes áreas. El uso de la evaluación de la marcha con escalas o pruebas clínicas o el análisis de la marcha a simple vista es más fácil de usar en el contexto de la consulta clínica diaria donde los pacientes presentan diferentes patologías que conducen a alteraciones de la marcha. Por otro lado, las técnicas instrumentales se suelen utilizar en proyectos de investigación. Aun así, en el contexto de la investigación, se utilizan ampliamente escalas y pruebas clínicas validadas en diferentes poblaciones e idiomas.

Tabla 1 - Usabilidad de las herramientas de evaluación de la marcha

Característica	Análisis de la marcha por observación	Cuestionarios, escalas y pruebas clínicas	Técnicas instrumentales
Costo de tiempo	+	+	+ / ++/ +++ (dependiendo del sistema utilizado)
Formación de evaluadores	+	+	++ / +++ (dependiendo del sistema utilizado)
Contexto de uso	Clínica	Clínica y de investigación	Investigación
Usabilidad	+	++	+++

## 2.1 Requisitos del equipo

Mientras que la evaluación de la marcha con una escala o con un cuestionario no requiere más que el formulario correspondiente y un bolígrafo, las pruebas clínicas pueden requerir una cierta cantidad de materiales para preparar el montaje correspondiente. Algunas pruebas pueden requerir una silla, cinta para marcar el suelo y un cronómetro, mientras que otras pueden requerir una mayor cantidad de materiales (vea el ejemplo a continuación). En cualquier caso, estos materiales suelen ser económicos (tabla 2) o accesibles para cualquier profesional sanitario.

### Ejemplo

#### ¿Cuántos materiales necesito para hacer la prueba 6-minutes walking test?

La 6-minutes walking test (6MWT) también se ha utilizado como una medida única del estado funcional de los pacientes, así como como predictor de morbilidad y mortalidad. La 6MWT es una prueba práctica y sencilla que requiere un pasillo de 30 metros, pero sin equipo de ejercicio ni capacitación avanzada para los técnicos. Esta prueba mide la distancia que un paciente puede caminar rápidamente sobre una superficie plana y dura en un período de 6 minutos (la 6MWD). Evalúa las respuestas globales e integradas de todos los sistemas involucrados durante el ejercicio, incluidos los sistemas pulmonar y cardiovascular, la circulación sistémica, la circulación periférica, la sangre, las unidades neuromusculares y el metabolismo muscular. La PM6M auto-guiada evalúa el nivel submáximo de capacidad funcional.

El equipo requerido es:

1. Temporizador de cuenta atrás (o cronómetro).
2. Contador de vueltas mecánico.
3. Dos pequeños conos para marcar los puntos de cambio.
4. Una silla que se pueda mover fácilmente a lo largo del recorrido a pie.
5. Hojas de trabajo en un portapapeles (formulario de registro).
6. Cinta adhesiva o pegatinas de colores para marcar el lugar de detención del paciente a los 6 minutos.
7. Escala Borg.
8. Pulsioxímetro.
9. Esfignomanómetro y estetoscopio.
10. Teléfono.
11. Una fuente de oxígeno.
12. Desfibrilador electrónico automático.



*Preparación de la 6-minutes walking test*

Sin embargo, cuando se trata de evaluar con técnicas instrumentales, el material requerido no es solo más cantidad, es altamente especializado. En general, el equipo de medición para registrar el patrón de la marcha o una característica específica de la marcha consta de las siguientes partes:

- **Sensores y/o equipos de medición:** referidos a los dispositivos que componen el propio equipo de medición en sí. Por ejemplo:
  - Una plataforma dinamométrica está formada por una superficie rígida, plana y estática instalada en el suelo y los sensores o transductores que se colocan en la parte inferior de la plataforma, distribuidos de manera que sean capaces de registrar las fuerzas ejercidas sobre el superficie superior rígida.
  - El equipo de electromiografía está compuesto por electrodos (sensores) que recolectan la actividad eléctrica dentro del músculo, ya sea insertándolo o a través de la piel que cubre el músculo. El sistema también incluye el amplificador que permite que las señales eléctricas del músculo se transfieran a un monitor, generalmente de forma inalámbrica, y un cableado entre los sensores y el amplificador..
- **Software y ordenador/computadora:** se refiere al software que permite la gestión de los equipos y sensores de registro, el almacenamiento de los datos registrados y la computadora donde se utiliza el software.
- **Suministros:** se refiere a todo tipo de accesorios necesarios para realizar una medición de la marcha. Por ejemplo:
  - En un sistema de electromiografía, los accesorios necesarios son los que servirán para preparar la piel donde se colocará el electrodo y así reducir la impedancia entre el electrodo y la piel: afeitadora, alcohol, algodón y lija fina para eliminar la piel muerta. Además, necesitará un gel conductor que también ayudará a reducir la impedancia de los electrodos de la piel..



**Ejemplo**

**¿Cuántos materiales necesito para hacer una evaluación de la marcha con un sistema de fotogrametría?**

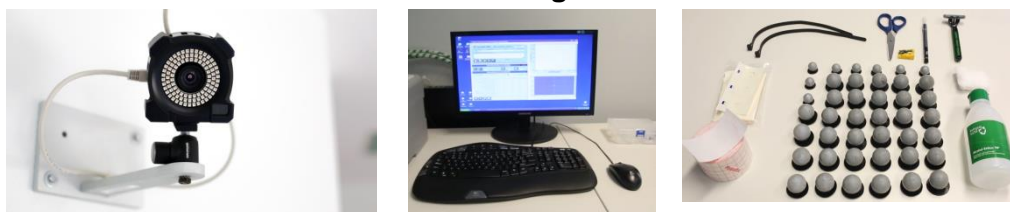


Figura 1 - Sistema de fotogrametría y sus componentes

Además de las cámaras, el software y la computadora, en el sistema de fotogrametría será fundamental contar con un conjunto de hitos para representar el modelo biomecánico. Además, para que estos queden bien adheridos a la piel, será necesario un sistema de adhesivos, elementos para limpiar la piel, y evitar que se caigan durante la prueba (afeitadora, alcohol, algodón) y ropa necesaria para realizar la prueba.

Como puede verse, los materiales y accesorios de las técnicas instrumentales para la evaluación de la marcha no suelen encontrarse en los servicios médicos. El costo del equipo y del software puede variar enormemente según el instrumento de evaluación y las posibilidades del software (Tabla 2). Por ejemplo, algún software puede ofrecer datos normativos con los que comparar la valoración realizada a un paciente de determinado sexo y edad, lo que incrementa su precio. Los equipos sin instrumentación como una pasarela con sensores de presión y su respectivo software pueden costar alrededor de 40 mil euros dependiendo de la longitud del pasillo, mientras que los equipos de fotogrametría con cámaras y software pueden superar fácilmente los 100 mil euros.

Tabla 2 - Requisitos de las herramientas de evaluación de la marcha

Característica	Análisis de la marcha por observación	Cuestionarios, escalas y pruebas clínicas	Técnicas instrumentales
Equipo	+	+	+++
Materiales	-	+	++
Costo económico	+	+	+++



### 2.3 Objetividad de los resultados y análisis estadístico

Los datos objetivos de una evaluación de la marcha se refieren a los datos obtenidos sin la interpretación del evaluador (es decir, la evaluación directa de una o más dimensiones del patrón de la marcha), mientras que en los datos subjetivos, el resultado está sujeto a la interpretación, percepción u opinión del evaluador. En las herramientas disponibles para la evaluación de la marcha encontramos ambos tipos de resultados. Mientras que con las herramientas de evaluación instrumental podemos obtener datos de medición objetiva, el análisis de la marcha a través de la observación, cuestionarios o escalas clínicas proporcionará información subjetiva. Es por ello que conviene al profesional de la salud conocer antes de su aplicación que las escalas de evaluación son instrumentos muy útiles para determinar cualitativamente estados generales de salud; sin embargo, están algo sujetos a la subjetividad tanto del paciente (evaluación auto-reportada) como del evaluador (evaluación de escala de observación). Además, en ocasiones se desarrollan en contextos muy específicos (poblaciones, regiones, países, etc.), con sesgos demográficos y culturales muy importantes, que dificultan o imposibilitan la extrapolación a otras comunidades. Por otro lado, el principal beneficio de las evaluaciones auto-reportadas es que dan cuenta de lo que la persona realmente experimenta y percibe. Además, el método es económico, rápido y factible.

#### Ejemplo

#### Medida subjetiva y objetiva de la longitud del paso

Muchas escalas de evaluación de la marcha tienen algunos elementos relacionados con la longitud del paso. Uno de ellos es la Tinetti Mobility Test (TMT), en el que en la sección de la marcha, se pregunta si el paciente es capaz de sobrepasar un pie al otro al dar el paso. Cuando medimos el paso de longitud en una persona sin patología, la respuesta puede ser claramente visible y el observador sin duda señalará que el pie derecho sobrepasa al izquierdo (o viceversa). Sin embargo, si se evalúa a un paciente con Parkinson, en el que una de las características de la enfermedad es caminar con pequeños pasos, puede darse el caso de que la respuesta al ítem TMT no sea tan clara y dos investigadores observen cosas distintas. Por otro lado, un instrumento de medición como la pasarela instrumentada es capaz de medir la longitud del paso detectando la presión plantar cuando el paciente coloca el pie en el suelo. Con este tipo de instrumento, la distancia obtenida de la longitud del paso no está influenciada por la observación del evaluador.

Longitud y altura del paso	
El pie izquierdo no sobrepasa el derecho en apoyo	0
El pie izquierdo sobrepasa el derecho en apoyo	1

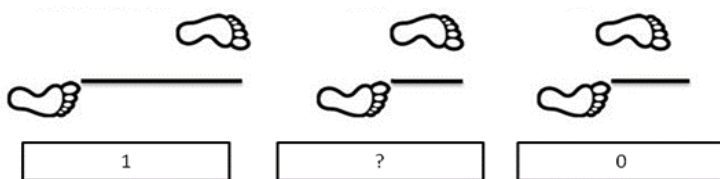


Figura 2 - Ítem de longitud y altura del paso de la prueba de movilidad de Tinetti. Se muestran tres ejemplos de longitud de paso en los que el evaluador debe marcar 1 o 0 según los criterios establecidos por la escala. En el ejemplo del centro, puede dar lugar a dudas sobre qué numeración asignar.

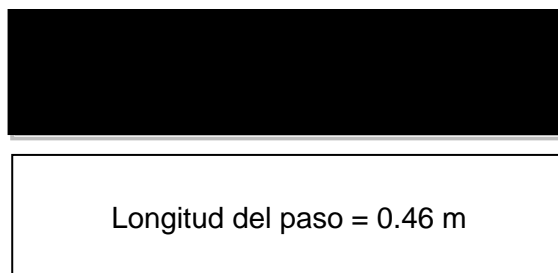


Figura 3 – Análisis de longitud del paso mediante una pasarela instrumentada (GAITrite).

La principal ventaja de utilizar resultados objetivos medidos con técnicas instrumentales es que estas permiten, entre otras cosas, que los datos sean comparables con otros datos del mismo paciente o con resultados de otros sujetos. Sin embargo, el investigador o el evaluador profesional no debe olvidar que la comparación de datos objetivos entre sujetos debe estar normalizada por la característica más influyente sobre estos datos. Por ejemplo, la longitud del paso también está influenciada por la altura del sujeto y por lo tanto la longitud de las extremidades inferiores, así pues, los datos objetivos de la longitud del paso deben dividirse por la altura o la longitud de las piernas. Lo mismo ocurrirá con variables como las fuerzas de reacción y el peso de los sujetos.

Aunque las escalas o cuestionarios y las herramientas instrumentales no brindan el mismo tipo de información (subjetiva versus objetiva), las medidas subjetivas pueden estar altamente correlacionadas con medidas objetivas. Este es otro valor añadido a las escalas de valoración utilizadas en el ámbito clínico. Una escala o cuestionario que esté altamente correlacionado con los resultados de la evaluación utilizando una técnica instrumental, será una herramienta válida para medir poblaciones en muestras grandes.

### Ejemplo

Journal List > Age Ageing > PMC4476851



[Age Ageing](#). 2015 Jul; 44(4): 691–694. PMID: PMC4476851  
Published online 2015 May 27. doi: [10.1093/ageing/afv062](https://doi.org/10.1093/ageing/afv062) PMID: [26018999](https://pubmed.ncbi.nlm.nih.gov/26018999/)

**A comparison of subjective and objective measures of physical activity from the Newcastle 85+ study**

[Paul Innerd](#),<sup>1</sup> [Michael Catt](#),<sup>1</sup> [Joanna Collerton](#),<sup>2</sup> [Karen Davies](#),<sup>2</sup> [Michael Trenell](#),<sup>3</sup> [Thomas B. L. Kirkwood](#),<sup>4</sup> and [Carol Jagger](#)<sup>4</sup>

Age A

En la investigación epidemiológica, la Actividad Física se evalúa comúnmente mediante cuestionarios de actividad física debido a su practicidad y bajo costo. Sería extremadamente caro y complejo evaluar muestras grandes de participantes con alguna

técnica instrumental. Es por ello que existen estudios que buscan determinar la correlación de los cuestionarios de actividad física autoinformados con medidas objetivas como la actividad física registrada con acelerómetros utilizados en una pequeña muestra de la población objetiva durante una semana. Así, el objetivo de este tipo de estudios es evaluar la validez de la prueba subjetiva como herramienta de evaluación de la actividad física para grandes muestras de participantes.

Ahora bien, ¿los datos subjetivos son susceptibles de ser analizados estadísticamente? La respuesta es sí, tanto la información subjetiva obtenida a través de escalas o cuestionarios como la información objetiva obtenida de una técnica instrumental son susceptibles de ser analizadas estadísticamente. Claramente, la velocidad al caminar, por ejemplo, se considera una variable cuantitativa, es decir, es una variable que tiene cantidades numéricas como modalidades con las que podemos hacer operaciones aritméticas (Tabla 3).

Tabla 3 - Objetividad de los resultados y análisis estadístico

Característica	Análisis de la marcha por observación	Cuestionario, escalas y prueba clínica	Técnicas instrumentales
Objetividad	Información subjetiva	Información subjetiva	Información objetiva
análisis estadístico	No (a priori)	Si	Si
Tipo de variable en análisis estadístico	-	Cualitativa, semi-cuantitativa y cuantitativa. (dependiendo de la respuesta a analizar)	Cuantitativa

Por otro lado, a partir de un cuestionario o una escala, podemos analizar su puntuación final. Si esta puntuación final es un número, se analizará estadísticamente como una variable cuantitativa, porque es la que introduciremos en el programa estadístico (un número), pero sigue siendo información subjetiva. Sin embargo, si lo que sometemos a análisis es cada pregunta de la escala o cuestionario, podemos obtener diferentes variables estadísticas (Tabla 3), dependiendo del tipo de respuestas que podamos encontrar::

1. La Escala Dinámica de la Marcha de Parkinson tiene preguntas en las que el evaluador debe asignar una puntuación entre 0 y 5 en el desempeño del paciente al caminar. Por ejemplo, en el ítem 1 referente a cómo el paciente camina siete metros hacia adelante, el evaluador puede calificar esta característica de normal (0 puntos) a "incapaz de dar un paso adelante o fallar" (5 puntos). Aunque sigue siendo modalidades categóricas dentro de la pregunta, estas posibles respuestas tienen un orden: uno es peor que cero, dos es peor que uno, tres es peor que dos, y así sucesivamente. Esta es una variable semicuantitativa.

Dynamic Parkinson Gait Scale (DYPAGS)	
	Score
<b>1. Walking 7 m forwards</b>	
Normal	0
Subtle start hesitation (<1 s) or slow gait or increased double-support time	1
Start hesitation >1 s or destination hesitation or impaired feet clearance	2
Block or accelerated short steps	3
Unable to perform the entire distance or near fall	4
Unable to initiate a step forward or fall	5

2. El Test de Movilidad de Tinetti tiene preguntas donde el observador debe verificar una característica de la marcha en el paciente y asignar 0 punto si lo que observa es una respuesta negativa o alterada o 1 punto si la característica de la marcha evaluada se presenta como positiva o normal. En otras palabras, este tipo de pregunta o ítem de escala es una respuesta sí / no. Analizada estadísticamente, la consideraríamos como una variable cualitativa, categórica con dos modalidades: 0 = característica de la marcha alterada y 1 = característica de la marcha normal.

Walking time	Heels apart	= 0
	Heels almost touching while walking	= 1

Finalmente, cuando el observador realiza un análisis de la marcha del paciente a simple vista y obtiene un conjunto de ideas o características sobre el patrón de la marcha que observa, esta información es subjetiva pero al no estar estandarizada no podría ser analizar utilizando métodos estadísticos convencionales (Tabla 3).

## 2.4 Validez

Antes de ser considerados adecuados, los instrumentos deben ofrecer datos precisos, válidos e interpretables para la evaluación de la salud de la población. Además, se supone que las medidas proporcionarán resultados científicamente sólidos. El desempeño de los resultados de estas medidas proviene de la fiabilidad y validez de los instrumentos.

La validez se refiere al hecho de que una herramienta mide exactamente lo que se propone medir o, en otras palabras, la validez se refiere a la precisión de la medición (Figura 4). La validez no es una característica del instrumento y debe determinarse sobre un tema específico, una vez que se refiere a una población definida. Las propiedades de medición, validez y fiabilidad, no son totalmente independientes. Un instrumento que no es fiable no puede ser válido; sin embargo, un instrumento fiable, a veces, puede ser inválido. Por tanto, una alta fiabilidad no asegura la validez de un instrumento.

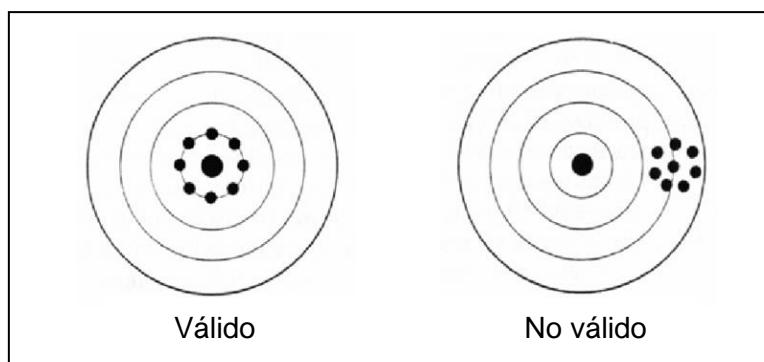


Figura 4 - En el blanco de la izquierda, los disparos dieron en el lugar donde se suponía que debían y fueron consistentes, justo en el centro del blanco. En la diana de la derecha, los disparos fueron fiables, acertando en el mismo punto; sin embargo, ninguno ha alcanzado el centro del objetivo, por lo que no se considera válido (Imagen de De Souza A. et al.2017).

La validez de los instrumentos de evaluación requiere varias fuentes de evidencia para crear el caso en el cual el instrumento mida lo que se supone que debe medir. La determinación de la validez puede verse como la construcción de un argumento basado en evidencia sobre qué tan bien mide una herramienta lo que se supone que debe hacer. La evidencia se puede reunir para respaldar, o no respaldar, un uso específico de la herramienta de evaluación. Es por ello que en la bibliografía encontramos varios estudios que buscan estudiar la validez de una herramienta para medir la marcha en una determinada población objetivo (Gail M. Sullivan 2011).

### ¿Cuál es el procedimiento para medir la validez de una herramienta?

La técnica o herramienta cuya validez se pretende estudiar debe compararse con una herramienta "gold standard" cuya validez ya ha sido ampliamente probada para medir un resultado específico (validez concurrente). En este tipo de estudio, el objetivo es responder a la pregunta: ¿la herramienta A mide con tanta precisión como la herramienta B la marcha humana? Esto generalmente se analiza usando el coeficiente de correlación de Pearson o Spearman ( $r$ ). Se considera excelente cuando el coeficiente de correlación entre el instrumento de medición y el estándar de referencia es superior a 0,6, adecuado entre 0,59 y 0,31 y pobre por debajo de 0,3.

#### Ejemplo 1: validez de escala

Diane M. Wrisley y Neeraj A. Kumar (2010) determinaron la validez de la Functional Gait Assessment (FGA) para clasificar el riesgo de caídas y predecir caídas inexplicables en adultos mayores que viven en la comunidad. Para ello, treinta y cinco adultos mayores de 60 a 90 años completaron la Activities specific Balance Confidence Scale (ABC), la Berg Balance Scale (BBS), el Dynamic Gait Index (DGI), el Timed "Up & Go" Test (TUG), además de la herramienta a validar (FGA). Se utilizaron coeficientes de correlación de



Spearman para determinar la validez concurrente entre ABC, BBS, TUG, DGI y FGA. El FGA se correlacionó con ABC ( $r = .053$ ,  $P < .001$ ), BBS ( $r = .84$ ,  $P < .001$ ) y TUG ( $r = .84$ ,  $P < .001$ ). Los autores determinaron que la FGA con una puntuación de corte de 22/30 es eficaz para clasificar el riesgo de caídas en adultos mayores y predecir caídas inexplicables en adultos mayores que viven en la comunidad.

### **Ejemplo 2: validez del análisis tridimensional de la marcha en cinta rodante**

Pinto R. y col. (2020) determinaron la validez de la biomecánica de la marcha tridimensional (3D) derivada de sistemas basados en cintas de correr mecánicas. Para ello, los investigadores examinaron la validez concurrente mediante la estimación de las asociaciones entre las medidas en cinta de correr mecánica y sobre el suelo (gold standard) en participantes con osteoartritis de rodilla. La velocidad de la cinta de correr mecánica se ajustó para auto-regularse a la velocidad sobre suelo. Una vez colocados los marcadores, los cálculos de la angulación y momento de la rodilla fueron consistentes en ambos sistemas. Las variables calculadas a partir de la angulación y las formas del momento de las ondas de la marcha en la rodilla durante la fase de apoyo fueron evaluadas usando la correlación de Pearson ( $r$ ) entre otros parámetros estadísticos. Los resultados mostraron que las correlaciones de Pearson entre la cinta de correr mecánica y los sistemas sobre suelo oscilaron entre 0,56 y 0,97. A pesar de estar altamente asociado, hubo diferencias sustanciales en los momentos, de esta forma se enfatizó que no se pueden usar indistintamente. Esto sugiere que los ángulos y momentos de la rodilla en los planos frontal y sagital, en pacientes con osteoartritis de rodilla, evaluados con un sistema de cinta de correr mecánica, son válidos.

Ahora que hemos definido la validez y cómo funciona, cabe preguntarse: ¿qué tipo de herramientas tienen más validez para medir la marcha o una característica específica de la marcha? ¿Escala / pruebas clínicas o técnicas instrumentales? La respuesta es clara, las técnicas de medición instrumental son más válidas para medir la marcha que las escalas o las pruebas clínicas porque son instrumentos más precisos para medir una determinada variable de la marcha. Cabe mencionar que incluso dentro de las técnicas instrumentales para el análisis de la marcha, hay algunas más precisas que otras. Estas diferencias de precisión están relacionadas con el error de medición del equipo implícito en la propia metodología. La figura 5 muestra una comparación de las técnicas instrumentales disponibles para medir los parámetros espacio-temporales de la marcha como la velocidad, la longitud del paso, la cadencia, la duración del ciclo de la marcha y las fases de la marcha.

Devices	Precision	Cost
Chronometer	+	+
Pedometer	+	+
GPS	++	++
Radar speed	+++	++++
Cross line detector	+++	++
Inertial measurement unit	++	+++
Footswitch	+++	++
Instrumented walkway	+++	++++
Optoelectronic cameras	++++	+++++

Figura 5 - Comparación de las tecnologías comunes utilizadas para medir los parámetros de la marcha espacio-temporal (Moissenet F. y Armand S. 2016). Para cada técnica instrumental se menciona el grado de precisión y el costo de la técnica.

## 2.5 Fiabilidad

Por otro lado, la fiabilidad es la capacidad de reproducir un resultado consistente en el tiempo y el espacio, o por diferentes observadores (Figura 6), presentando aspectos de coherencia, estabilidad, equivalencia y homogeneidad. Es uno de los criterios de calidad de un instrumento (de Souza et al.2017). La fiabilidad también está relacionada con la repetibilidad. Por ejemplo, una escala o prueba es fiable si las mediciones realizadas con el instrumento dan el mismo resultado cuando se evalúa bajo condiciones constantes (Taherdoost H. 2016).

Es importante destacar que la fiabilidad de un cuestionario no es una propiedad fija. Por el contrario, la fiabilidad depende de la función del instrumento, de la población en la que se utiliza, de las circunstancias, del contexto; es decir, el mismo instrumento puede no considerarse fiable en diferentes condiciones. Las estimaciones de fiabilidad se ven afectadas por varios aspectos del entorno de evaluación (evaluadores, características de la muestra, tipo de instrumento, método de administración) y por el método estadístico utilizado. Por lo tanto, los resultados de una investigación que utiliza instrumentos de medición solamente se pueden interpretar cuando las condiciones de evaluación y el enfoque estadístico se presentan claramente. En conclusión, la fiabilidad se refiere a si un instrumento de evaluación da los mismos resultados cada vez que se utiliza en el mismo entorno con el mismo tipo de sujetos (Sullivan G.M.2011).

La elección de las pruebas estadísticas para evaluar la fiabilidad pueden variar, dependiendo de lo que se pretende medir. Existen tres importantes criterios de fiabilidad de



gran interés para las y los investigadores: 1) Estabilidad, 2) Consistencia interna y 3) Equivalencia.

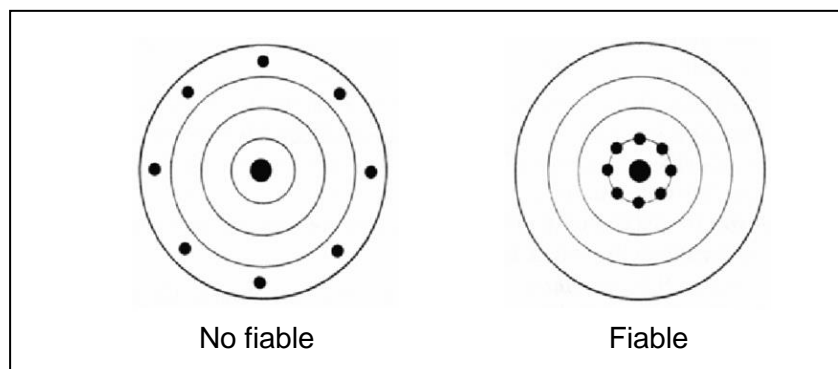


Figura 6 - En la diana de la izquierda, los disparos no son fiables porque los puntos impactados no están ubicados en un lugar específico, sino que se reparten por toda la diana. En la diana de la derecha, los disparos fueron fiables, acertando en el mismo punto cada vez (Imagen de De Souza A. et al. 2017).

### ¿Cuál es el procedimiento para medir la fiabilidad de un instrumento?

Como se mencionó anteriormente, el procedimiento estadístico para medir la fiabilidad depende de lo que se pretenda medir.

1) *Estabilidad*: la estabilidad mide qué tan similares son los resultados cuando se miden en dos momentos diferentes, es decir, estima la consistencia de la repetición de la medición. La evaluación de la estabilidad se puede realizar mediante el método test-retest, es decir, el procedimiento consiste en aplicar la misma medición en dos momentos diferentes. El coeficiente de correlación intraclass (Intraclass correlation coefficient, ICC) es una de las pruebas más utilizadas para estimar la estabilidad de las variables continuas, ya que tiene en cuenta los errores de medición. Para la interpretación de los resultados se consideran satisfactorios valores mínimos de 0,70 (de Souza et al. 2017).

2) *Consistencia interna*: la consistencia interna (u homogeneidad) muestra si todas las subpartes de un instrumento miden la misma característica. Esta es una propiedad de medida importante para los instrumentos que evalúan un constructo que utiliza una variedad de ítems. Una consistencia interna baja puede indicar que los ítems miden diferentes constructos (de Souza et al. 2017). En el análisis de consistencia interna, se calcula la diferencia entre las respuestas de los ítems del constructo valorado. Alfa de Cronbach es la prueba que se utiliza con frecuencia para calcular los valores de correlación entre las respuestas de una herramienta de evaluación. Una estimación de alta fiabilidad debe ser lo más cercana posible a 1 (Sullivan G.M. 2011).

3) *Equivalencia*: la equivalencia es el grado de concordancia de dos o más observadores con respecto a la puntuación de un instrumento. La forma más común de evaluar la equivalencia es a través de la fiabilidad interobservador, que implica la participación independiente de dos o más evaluadores. El coeficiente Kappa es una medida utilizada para evaluar interobservadores, aplicada a variables de categoría y tiene un valor máximo de 1. Cuanto mayor sea el valor Kappa, mayor será la concordancia entre los evaluadores (Sullivan G.M. 2011).

### Example 1: Fiabilidad de estabilidad

#### Técnica Instrumental

##### *Inter-evaluador test-retest (diferentes evaluadores)*

Meng L. *et al.* (2020) investigaron si una menor dependencia de la indentificación manual de puntos de referencia anatómicos podría mejorar la fiabilidad entre evaluadores de la cinemática articular, a través de la comparación de tres modelos biomecánicos para medir la marcha humana. Este objetivo del estudio se planteó porque una fuente importante de error en la fiabilidad del análisis de la marcha surge de la palpación de puntos de referencia anatómicos por parte de la persona que instrumenta al sujeto de estudio. La hipótesis del estudio fue que el modelo de clúster funcional de Strathclyde (Strathclyde functional cluster model, SFCM) obtendría una mayor confiabilidad interevaluador que los modelos anatómicos: Plug-in gait (PIG) y el modelo de marcha para el cuerpo humano (Human body model gait ,HBM2). Para demostrar esto, 10 participantes completaron tres ensayos realizados por diferentes asesores en días no consecutivos. En cada sesión, el evaluador aplicó el conjunto de marcadores de todos los modelos biomecánica a los participantes. Primero, se registró una prueba estática durante 5 s con el participante de pie en una postura natural. Después de esto, cada participante caminó en la cinta rodante a su velocidad confortable. Se capturaron 30 s de marcha después de un período de familiarización de dos minutos; 12 cámaras capturaron las trayectorias de los marcadores a una frecuencia de muestreo de 100 Hz.

Se utilizó el coeficiente de correlación intraclase a través de los ciclos de la marcha para comparar la fiabilidad entre evaluaciones. El valor de ICC ( $\alpha$ ) se interpretó en cuatro niveles de confiabilidad:  $\alpha < 0.5$  = pobre,  $0.5 < \alpha < 0.75$  = moderado,  $0.75 < \alpha < 0,9$  = bueno y  $\alpha > 0,9$  = excelente. Los resultados demuestran que todos los modelos mostraron una fiabilidad entre evaluadores de "buena a excelente" para todos los ángulos de flexión / extensión y ángulo de abducción / aducción de la cadera. Sin embargo, se demostró una fiabilidad entre evaluadores de "mala a moderada" para otros ángulos no sagitales. El modelo de clúster funcional de Strathclyde obtuvo una mayor fiabilidad con menos variación en comparación con los modelos anatómicos. Los resultados demuestran que el modelo de clúster funcional de Strathclyde puede ser más beneficioso para los evaluadores con menos experiencia.

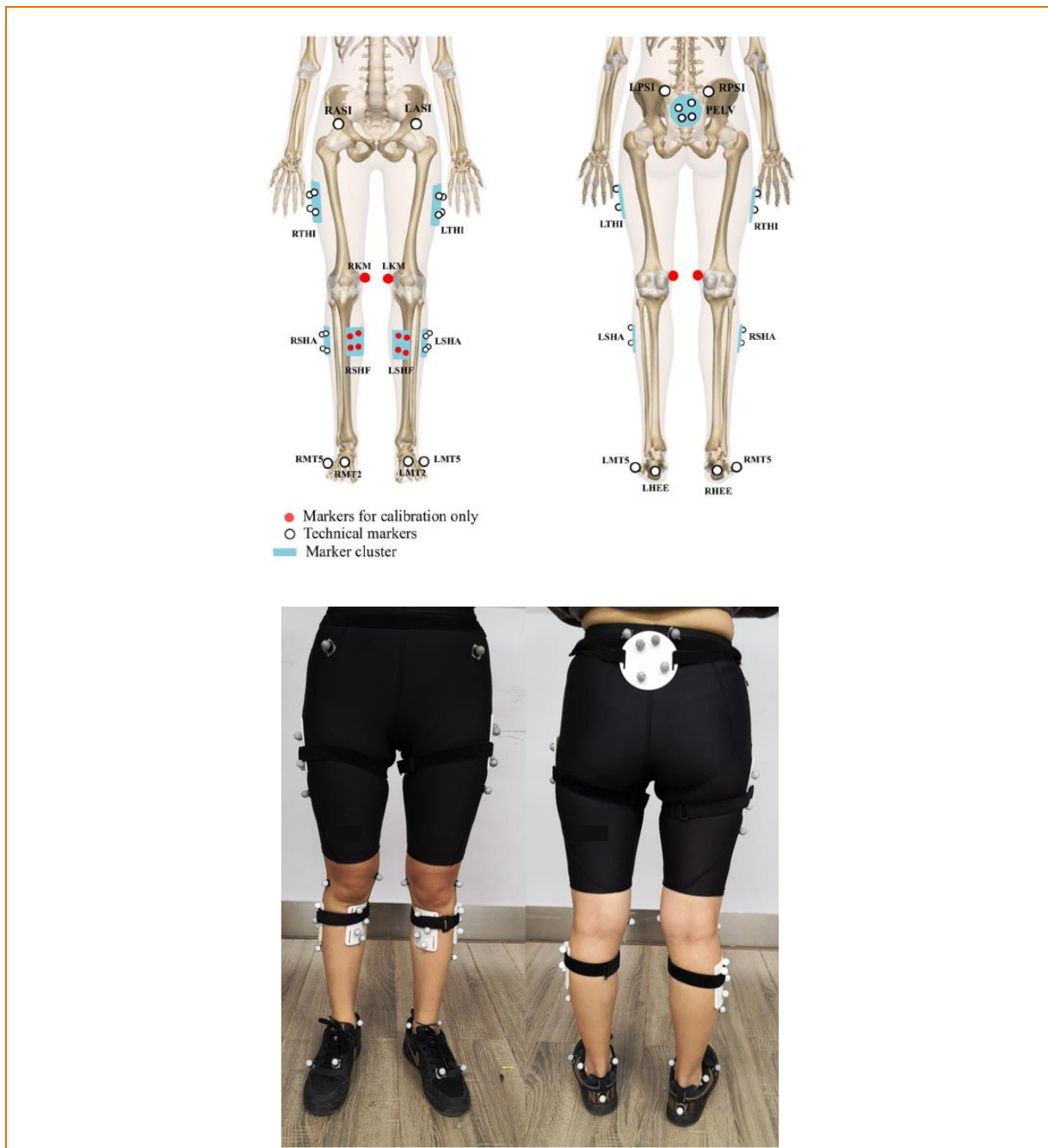


Figura 7 – Disposición de marcadores del modelo de clúster funcional de strathclyde (superior) y la configuración de marcador combinada de los modelos SFCM, PiG y HBM2 (inferior).

*Intra-evaluador test-retest (diferentes tiempos de evaluación)*

Geerse D. et al. (2020) determinaron la fiabilidad test-retest de los HoloLens de Microsoft, un auricular mixto de realidad virtual que proporciona, además de hologramas, datos de posición de la cabeza, los que se pueden usar para cuantificar diferentes tareas como caminar. Para estudiar cuáles son los límites de concordancia de los parámetros de la marcha espacio-temporal derivados de distintas repeticiones, se reclutaron 23 adultos

jóvenes sanos que caminaban a velocidad lenta, cómoda y rápida, así como 24 personas con enfermedad de Parkinson que caminaban a una velocidad auto-seleccionada. HoloLens proporciona datos de posición y orientación de la cabeza en el espacio en tres direcciones; estos datos de posición se utilizaron para derivar los parámetros de la marcha espacio-temporales. La fiabilidad test-retest se evaluó con el coeficiente de correlación intraclase. Los ICC por encima de 0,60 y 0,75 representaron un acuerdo bueno y excelente, respectivamente. Los ICC indicaron una excelente fiabilidad test-retest para ambos sistemas, tanto para adultos jóvenes sanos y personas con enfermedad de Parkinson.

			Trial 1		Trial 2	
			Mean ± SD	Mean ± SD	ICC <sub>(A,1)</sub>	
Walking speed (cm/s)	HYA	SWS	74.2 ± 12.6	76.2 ± 12.2	0.861	
		CWS	116.2 ± 16.2	122.1 ± 16.4	0.870	
	PD	FWS	176.3 ± 21.8	174.7 ± 22.8	0.930	
		CWS	104.5 ± 20.7	106.1 ± 20.8	0.935	
Step length (cm)	HYA	SWS	56.2 ± 6.4	56.8 ± 5.4	0.884	
		CWS	67.9 ± 7.0	69.7 ± 6.6	0.911	
	PD	FWS	86.2 ± 8.0	85.5 ± 8.3	0.928	
		CWS	57.8 ± 11.5	57.8 ± 10.9	0.939	
Cadence (steps/min)	HYA	SWS	79.2 ± 9.8	80.6 ± 9.6	0.903	
		CWS	103.8 ± 8.5	106.5 ± 8.7	0.890	
	PD	FWS	124.1 ± 10.4	123.7 ± 9.8	0.953	
		CWS	109.6 ± 7.6	110.5 ± 7.0	0.778	

Figura 8 – Fiabilidad test-retest reliability para los parámetros espacio-temporales de la marcha extraídos del dispositivo HoloLens, en personas que caminaron a velocidad lenta (slow walking speed, SWS), auto-seleccionada (comfortable walking speed, CWS) y rápida (fast walking speed, FWS) en adultos jóvenes sanos (healthy young adults, HYA) (Geerse D. et al. 2020).

Como era de esperar, en los análisis de fiabilidad de una técnica instrumental, los valores de ICC son elevados, ya que las técnicas de medida suelen ser de alta precisión. Factores como seguir un protocolo de evaluación estricto ayudan a mantener una alta fiabilidad test-retest. Si el valorador utiliza una técnica de medición instrumental sin seguir el mismo protocolo de instrumentación de forma estandarizada para los diferentes tiempos de evaluación, es probable que los valores de ICC no fueran los observados en el estudio de ejemplo. También se puede observar que la fiabilidad de la técnica de medición disminuye en la cadencia de los participantes con enfermedad de Parkinson. Esto ocurre porque este grupo de pacientes cambia la cadencia en función de la longitud de la zancada que realizan, la que se alterada por la enfermedad. Este ejemplo es útil para señalar que la fiabilidad de la medición, aunque se utilice una técnica instrumental, varía en diferentes poblaciones.

### Tes Clínico

Los análisis de fiabilidad son aún más valorados en el ámbito asistencial por el personal clínico. Hee-jae Kim et al. (2016) examinaron la fiabilidad (Intra-evaluador test-retest) de la velocidad de la marcha medida en pasillos con diferentes distancias y ritmos, en personas mayores koreanas. Cincuenta y cuatro participantes mujeres de 70 años o más fueron reclutadas de una comunidad de jubilados local. La velocidad de la marcha se evaluó en un pasillo de 4, 6 y 10 metros, y en el modo de marcha habitual y caminando rápido. Se indicó a las participantes que caminaran desde la posición de inicio hasta la marca que señalizaba el final del pasillo, primero a velocidad confortable y luego todo lo rápido que pudieran. Un evaluador instruido caminó detrás de las participantes y detuvo el cronómetro cuando el pie de cada sujeto alcanzó el final del recorrido. Esto significa que la velocidad de marcha se calculó con la distancia del pasillo recorrida y el tiempo cronometrado por el evaluador (d/t). Los resultados se muestran en la Figura 9.

	Normal pace		
	4 M	6 M	10 M
ICC (95% CI)	.715**	.861**	.902**
	Maximal pace		
	4 M	6 M	10 M
	.837**	.905**	.933*

Figura 9 – Fiabilidad del test de marcha (ICC) para las diferentes distancias y ritmos.

Se observaron valores de ICC más altos en la distancia de caminata más larga (10 metros) en comparación con las pruebas de marcha en los pasillos de 4 y 6 metros. Además, los valores de ICC de la prueba de la marcha a la velocidad máxima fueron más altos que los del paso normal. Los autores mencionaron que, aunque la prueba de marcha a un ritmo máximo en una distancia más larga tiene mejor fiabilidad en personas mayores, la distancia y el ritmo de la prueba deben considerarse en relación con el propósito de las mediciones y las condiciones clínicas de salud de los participantes, más que por el criterio de un alto nivel de fiabilidad.

### Example 2: Fiabilidad de consistencia interna

#### Test Clínico

A diferencia de las herramientas instrumentales, con las que medimos una variable objetiva como la velocidad de la marcha, la longitud de la zancada o el ángulo de movimiento articular, las pruebas clínicas nos permiten evaluar el desempeño global de la marcha desde una perspectiva clínica. Este es el caso de la prueba Functional Gait Assessment (FGA), una medida de la capacidad de equilibrio al caminar que se desarrolló para eliminar el efecto techo observado en otra prueba de marcha, el Índice dinámico de



Marcha (Dynamic Gait Index, DGI). Se agregaron tres tareas presuntamente más difíciles y se eliminó una de las tareas más sencillas originales de las 8 tareas DGI (Beninato M. y Ludlow L., 2016). Los elementos en el FGA son: 1) Marcha en una superficie nivelada, 2) Cambio en la velocidad de la marcha, 3) Marcha con giros horizontales de la cabeza, 4) Marcha con giros verticales de la cabeza, 5) Marcha y giro de pivote, 6) Pasar por encima de un obstáculo, 7) Marcha con base de apoyo estrecha, 8) Marcha con los ojos cerrados, 9) Andar hacia atrás y 10) Pasos. En cada ítem, el evaluador puede dar una puntuación de 0 a 3, obteniendo una puntuación máxima de 30 puntos en total.

Wrisley D. et al (2004), evaluaron la consistencia interna de los datos obtenidos con la FGA cuando se usa con personas con trastornos vestibulares. Seis pacientes con trastornos vestibulares completaron la FGA dos veces, con una hora de descanso entre sesiones. La consistencia interna, u homogeneidad de los ítems incluidos en el FGA, se determinó mediante el alfa de Cronbach. Esta evaluación se realizó dentro de cada test desarrollado y entre ambas sesiones de evaluación. El valor alfa de Cronbach varía de 0 a 1 y de debe considerar una consistencia alta cuando el valor alfa es superior a 0,8.

En este estudio, se demostró que el test tiene buena consistencia interna y entre los dos ensayos que se hicieron por paciente. Los valores alfa de Cronbach fueron 0,81 y 0,77 para los ensayos individuales 1 y 2, respectivamente. El alfa de Cronbach fue de 0,79 cuando se analizaron ambas pruebas. Este es uno de los análisis realizados para validar un instrumento de medición basado en ítems. La ventaja de este tipo de prueba es que obtenemos una puntuación del desempeño general de los pacientes en un área, obteniendo así una medida mucho más funcional que unos resultados únicos medidos con una técnica instrumental.

### Example 3: Fiabilidad de equivalencia

#### Test Clínico

En el mismo estudio de Wrisley D. et al (2004), también se evaluó la equivalencia o grado de concordancia de dos o más observadores que utilizaron la prueba Functional Gait Assessment (FGA), a través del Coeficiente Kappa. Para este propósito, siete fisioterapeutas de diferentes entornos clínicos, y 3 estudiantes de fisioterapia, se ofrecieron como voluntarios para participar. La configuración para esta evaluación fue la que aparece en la Figura 10, donde todos los evaluadores observaron a la vez el desempeño de cada paciente.

El estadístico de Kappa fue utilizado para evaluar el acuerdo entre valoradores sobre las puntuaciones de cada ítem del test, así como sobre la puntuación final de la prueba FGA. La Figura 11 contiene los porcentajes de concordancia y los valores de Kappa para cada ítem y las puntuaciones totales de FGA. Cohen sugirió que el resultado de Kappa se interprete de la siguiente manera (McHugh M. 2012):

- Los valores  $\leq 0$  indican que no hay concordancia
- Valores entre 0,01–0,20 indican concordancia nula o leve

- Los valores entre 0,21–0,40 indican una concordancia regular
- Los valores entre 0,41– 0,60 indican concordancia moderada
- Valores entre 0,61–0,80 indican concordancia sustancial
- Los valores entre 0,81–1,00 indican concordancia casi perfecta.

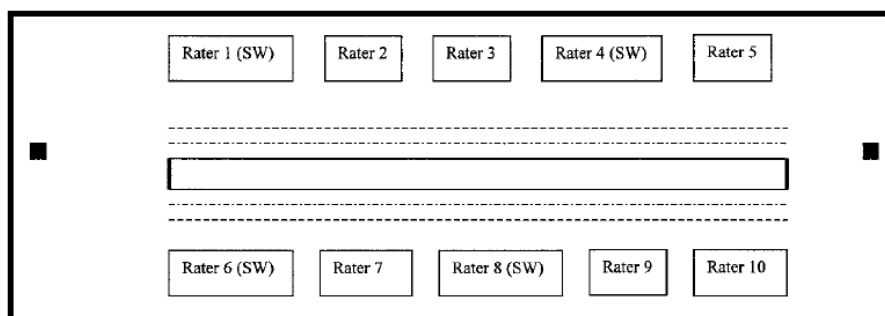


Figura 10 – Diseño del pasillo de marcha y de la posición que cada evaluador ocupó para observar la prueba. Los cuadrados negros representan conos que indican los puntos de inicio y final del recorrido. Los evaluadores con las iniciales SW, utilizaron un cronómetro durante el test.

	FGA Item										Total	
	1	2	3	4	5	6	7	8	9	10		
Interrater reliability												
% agreement	88	60	58	68	60	68	66	66	67	90	58	
Kappa	.78	.37	.40	.53	.34	.41	.45	.46	.54	.76	.50	

Figura 11 – Fiabilidad entre valoradores para los ítems individuales de la prueba Funcional Gait Assessment (FGA) y para su puntuación final.

El grado de acuerdo se determinó entre 10 clínicos que midieron a 6 pacientes durante una prueba de marcha. Como hay 45 posibles pares de acuerdos por sujeto de estudio, los 6 sujetos medidos crearon 270 posibles acuerdos para cada ítem del FGA y las puntuaciones totales. Los valores de la Figura 11 representan la concordancia media entre los pares de acuerdo para ítem y la puntuación final, siendo moderada o buena para todos los ítems, excepto en los ítems 2, “cambio de velocidad”, y 5 “giro con pivot”.

Los resultados sobre fiabilidad inter-evaluador del estudio de Wrisley D. et al (2004) son útiles para ejemplificar que las escalas y pruebas que se desarrollan observando al sujeto evaluado, pueden variar en el resultado según lo que cada observador detecta, lo cual se considera un desventaja de las escalas y pruebas clínicas con respecto a las herramientas instrumentales.



## 2.6 Sensibilidad al cambio y Capacidad de respuesta

La sensibilidad al cambio se define como la capacidad de un instrumento para medir el cambio de estado, independientemente de si el cambio es relevante o significativo para quien toma las decisiones. Aunque necesaria, la sensibilidad al cambio se ha descrito como insuficiente para evaluar el cambio y establecer la efectividad del tratamiento. Una prueba puede ser sensible a un estado o diagnóstico, pero si es significativa o importante no se puede deducir solamente de esta propiedad.

La importancia de esta propiedad de los instrumentos de medición de la marcha está relacionada con la evaluación de los impactos de los programas y tratamientos clínicos. tratamientos en la ciencia clínica. Si un instrumento no es lo suficientemente sensible, no podrá evaluar el efecto de una determinada patología o tratamiento en la marcha humana. Dicha sensibilidad es especialmente relevante en entornos aplicados donde los efectos del programa o del tratamiento a menudo no son particularmente fuertes y las condiciones de medición pueden ser bastante variables (Lipsey M. 1983). Si un instrumento no es lo suficientemente sensible para detectar un cambio (tamaño del efecto) en una determinada condición o tratamiento, con frecuencia se necesitan muestras grandes para lograr un poder estadístico adecuado para detectar el efecto de estas intervenciones.

Por otro lado, la capacidad de respuesta (en inglés, *responsiveness*) se define como la capacidad de un instrumento para medir un cambio significativo o clínicamente importante en un estado clínico y se ha defendido como una propiedad esencial de los instrumentos diseñados para medir el cambio y la eficacia de las intervenciones. De manera similar a la fiabilidad y la validez, la capacidad de respuesta no se considera una propiedad generalizable y debe evaluarse para cada población y propósito para el cual se usa la medida.

Un cambio clínicamente significativo o importante puede definirse y, por lo tanto, evaluarse, desde la perspectiva del paciente, su representante, la sociedad o el profesional de la salud. Implica un cambio que es notablemente diferente y que es valioso para el paciente (o el médico). Este cambio puede permitirle al individuo realizar alguna tarea esencial o realizar tareas de manera más eficiente o con menos dolor o dificultad. Estos cambios también deberían superar la variación que se puede atribuir al azar. Jaeschke et al. (1989) sugirió que un cambio clínicamente significativo podría definirse como la diferencia mínima importante; lo que a su vez puede establecerse como la diferencia más pequeña en la puntuación en el dominio de interés que un paciente percibe como un cambio y que, obligaría, en ausencia de efectos secundarios y costos excesivos, a modificar el manejo del paciente.

Así es como, la capacidad de respuesta se informa comúnmente a través de la estimación de la diferencia mínimamente importante (Minimally Important Difference, MID) o la diferencia mínima clínicamente importante (Minimal Clinically Important Difference, MCID) por lo que una puntuación de cambio en una medida debe igualar o superar la estimación de MID para que se considere importante. La estimación de MID de las medidas, mejora la interpretación de la puntuación de cambio, estableciendo puntos de referencia para ayudar a determinar la relevancia del cambio.

### Ejemplo práctico

Al evaluar la utilidad clínica de las terapias destinadas a mejorar las variables o resultados subjetivos, se debe determinar la cantidad de mejora que es importante para los pacientes. La diferencia mínima clínicamente importante (MCID) es el beneficio más pequeño de valor para los pacientes. El MCID es un concepto centrado en el paciente, que captura tanto la magnitud de la mejora como el valor que los pacientes le dan al cambio. En otras palabras, el MCID define la cantidad más pequeña que debe cambiar un resultado para que sea significativo para los pacientes (McGlothlin A. *et al* 2014).

En 2014, Bohannon R. *et al.* desarrollaron una revisión sistemática sobre la diferencia mínima clínicamente importante para el cambio en la velocidad de la marcha cómoda de los adultos con patología. La conclusión del estudio fue que los cambios en la velocidad de la marcha de 0,10 a 0,20 ms<sup>-1</sup> pueden ser importantes en varios grupos de pacientes, como personas con accidente cerebrovascular, fractura de cadera, esclerosis múltiple, entre otros. Por tanto, si queremos medir alguna intervención o evolución en personas con las patologías mencionadas en el estudio anterior, debemos utilizar un equipo de medición con suficiente sensibilidad para registrar cambios en la velocidad de la marcha como los indicados en el estudio de Bohannon R. *et al.*, es decir, capaz de registrar cambios de velocidad de 0,1 ms<sup>-1</sup>.

En resumen, los profesionales de la salud al elegir un instrumento de medición, deben tener en cuenta que la sensibilidad del equipo debe ser suficiente para medir una diferencia mínima clínicamente importante en la variable que desean observar en una población determinada.

En relación con la sensibilidad, las técnicas de evaluación biomecánica tienen una mayor sensibilidad que las escalas de evaluación clínica para detectar cambios en las características de la marcha. Esto se debe a que los instrumentos de medición biomecánicos son equipos mucho más precisos. Para comprender este concepto, mire la Figura 12.

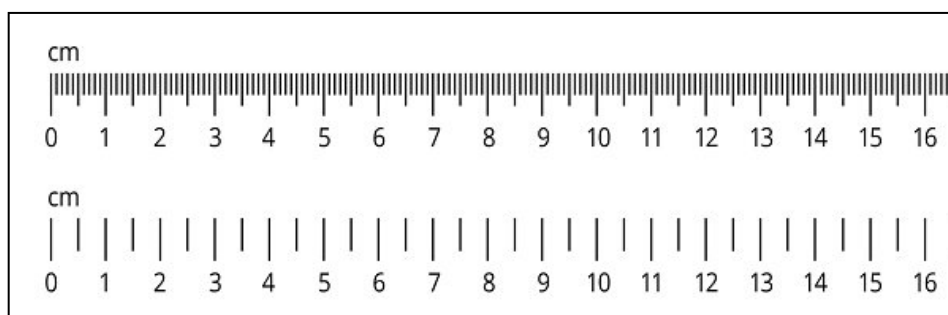


Figura 12 – La imagen muestra dos reglas graduadas. La regla superior está graduada en mm y la regla inferior está graduada en cm, y sus precisiones son respectivamente  $\pm 1$  mm y  $\pm 1$  cm =  $\pm 10$  mm. Dado que el primero es más preciso que el segundo, la regla superior es más sensible que la regla inferior.

La regla gradada superior, de la Figura 12, tiene una precisión de 1 mm y la regla graduada inferior en la figura tiene una precisión de 1 cm (10 mm). Dado que la regla superior es más precisa que la regla inferior, a su vez es más sensible a la medida de longitud. Este es también el mismo caso entre los instrumentos de evaluación biomecánica para medir la marcha humana. En la Figura 5, podemos ver la precisión de los diferentes instrumentos biomecánicos para medir parámetros espacio-temporales durante la marcha humana. Los sistemas de fotogrametría y las cámaras optoelectrónicas, son los equipos más precisos en comparación con otros sistemas, por lo tanto, también son más sensibles a la medición de variables espacio-temporales que las pasarelas equipadas con sensores de presión, sensores inerciales o un cronómetro (Moissenet F. et al.).

### Ejemplo: sensibilidad entre las pruebas clínicas de marcha para personas con lesión medular

Así como los instrumentos de evaluación biomecánica difieren en su sensibilidad para medir ciertos resultados, las escalas clínicas también tienen esta característica. Jackson A. et al. (2008) analizaron la utilidad en la práctica clínica de las medidas de resultado actuales utilizadas como indicadores de mejora en la marcha y la deambulacion en la población con lesión de la médula espinal. Específicamente, evaluaron las siguientes medidas de marcha y deambulacion para: Walking Index for Spinal Cord Injury II (WISCI II), 50-Foot Walk Test (50FTWT), 6-Minute Walk Test (6MWT), 10-Meter Walk Test (10MWT), y Functional Independence Measure- Locomotor (FIM-L). Los resultados se muestran en la Tabla 4.

Tabla 4 – Sensibilidad al cambio en el desempeño de la marcha de personas con lesión medular

Test clínico	Sensibilidad informada
Walking Index for Spinal Cord Injury II (WISCI II)	<ul style="list-style-type: none"> <li>- Buena sensibilidad al cambio en pacientes con mayor alteración de la marcha.</li> <li>- No incorpora elementos de velocidad o resistencia.</li> </ul>
50-Foot Walk Test (50FTWT)	<ul style="list-style-type: none"> <li>- Las personas que no pueden caminar 50 pies (15,24 metros) no puede participar en la medición (efecto suelo).</li> <li>- Aquellos pacientes que pueden caminar la distancia a un ritmo rápido también pueden caminar una distancia mayor a esa misma velocidad (efecto techo).</li> </ul>
6-Minute Walk Test (6MWT)	<ul style="list-style-type: none"> <li>- Buena sensibilidad al cambio en participantes con una mejor capacidad de marcha en la lesión aguda y 6 meses post lesión (with better ambulatory function in acute injury and 6 months post injury (mientras que el WISCI II y el Lower Extremity Motor Score no).</li> <li>- En pacientes con accidente cerebrovascular, la prueba mostró menos sensibilidad al cambio que el 12 MWT.</li> <li>- La sensibilidad de la 6MWT es afectada por un efecto suelo</li> </ul>

	<p>entre los pacientes que no pueden caminar el test de 6 minutos marcha.</p> <ul style="list-style-type: none"> <li>- Se permite el reposo en bípedo durante el test, pero si el sujeto se sienta, queda descalificado. De manera similar, se observa un efecto techo en pacientes que pueden continua caminando más de 6 minutos al mismo ritmo.</li> </ul>
10-Meter Walk Test (10MWT)	<ul style="list-style-type: none"> <li>- Más sensibilidad en pacientes con una recuperación superior al 90% en la puntuación del test Lower Extremity Motor score a los 6 y 12 meses después de la lesión (ASIA D).</li> <li>- Mejor sensibilidad que WISCI II, especialmente en sujetos que tienen menos deterioro debido al efecto techo de esa prueba.</li> <li>- Su sensibilidad al cambio en la deambulación es más comparable a otras pruebas de velocidad de la marcha como la 6MWT porque se miden rasgos similares.</li> <li>- La capacidad de respuesta al cambio también se ha demostrado en la población con accidentes cerebrovasculares utilizando una distancia modificada de 6 metros, como una variación de la 10MWT.</li> <li>- Mejor sensibilidad que la puntuación Lower Extremity Motor Score durante un período más largo (6 meses).</li> <li>- La medida de los resultados es menos sensible al cambio en la locomoción en la lesión medular crónica, donde hay pocos cambios en la fuerza.</li> </ul>
Functional Independence Measure- Locomotor (FIM-L)	<ul style="list-style-type: none"> <li>- Poca sensibilidad para mostrar un cambio en la puntuación de un sujeto antes y después del entrenamiento a pesar de que su modo de deambulación cambia de silla de ruedas a caminar.</li> <li>- Se revela un efecto techo en el hecho de que la prueba muestra poca sensibilidad al cambio en sujetos con mejores habilidades para caminar. Además, la puntuación de la FIM-L no tiene en cuenta la velocidad con la que los sujetos de prueba completan las distancias requeridas.</li> </ul>

La conclusión del estudio fue que en personas con lesión de la médula espinal, la prueba 10-Meter Walk Test y Walking Index for Spinal Cord Injury II, son las pruebas más válidas y clínicamente útiles para evaluar mejoras en la marcha debido, entre otras cosas, a su buena sensibilidad para valorar cambios en la deambulación en este grupo de pacientes con alta recuperación y con marcha más alterada, respectivamente.

En el ejemplo anterior, hemos visto que la sensibilidad de las pruebas de evaluación de la marcha se ve afectada por dos fenómenos denominados efecto techo y efecto suelo, que están relacionados con la viabilidad de la propia prueba para ser realizada por las personas objetivo.

## 2.7 Efecto suelo y techo

Si el rango de función cubierto por una medida es menor que el rango experimentado por los pacientes, la medida puede carecer de capacidad de respuesta. El potencial de efectos suelo y techo, a menudo se evalúa examinando patrones de respuesta. Si hay picos en la opción de respuesta más alta o más baja, esto a menudo se interpreta como evidencia de efectos techo o suelo, respectivamente. Sin embargo, cuando se utilizan medidas para evaluar la efectividad de las intervenciones, la evidencia prospectiva del desempeño de una medida es más importante que si hay picos o no (Feeny DH. *et al.* 2013). Si las puntuaciones de los participantes se agrupan hacia el extremo superior (o la mejor puntuación posible) de la medida / instrumento (efecto techo) o hacia el extremo inferior (efecto suelo), los cambios experimentados por los pacientes en el rendimiento de la marcha pueden estar sesgados. Esto se debe a que los pacientes podrían estar "en peor situación" de lo que la medida podría captar o "mejor" de lo que puede medir el instrumento.

Si comparamos las técnicas de evaluación instrumental con las escalas de evaluación clínica, observaremos que las técnicas instrumentales tienen un efecto suelo, mientras que las escalas de evaluación clínica de la marcha tienen un efecto techo. El motivo de esta diferencia es la facilidad o dificultad con la que cada una de las pruebas puede ser realizada por los pacientes. Las escalas de valoración se definen siguiendo cuestionarios estructurados, realizando clasificaciones categóricas o discretas que reducen la sensibilidad con respecto a otros instrumentos o tecnologías de valoración. Es decir, los pequeños cambios que se producen en la capacidad funcional del paciente como consecuencia de la intervención del profesional, son muy difíciles de identificar. Este hecho implica, según el diseño de la escala, un efecto techo. Sin embargo, las personas con alteración grave de la marcha no podrán realizar los complejos protocolos de valoración que requieren muchas de las técnicas instrumentales biomecánicas, por lo que la valoración más sensible para este tipo de pacientes serán las escalas de valoración clínica.

Por otro lado, las técnicas instrumentales de valoración biomecánica tienen un efecto suelo porque los protocolos de valoración requieren una mayor exigencia por parte del paciente para su realización. Es decir, en la mayoría de las veces, la evaluación de la marcha de personas con discapacidad severa en la deambulación no es posible con técnicas instrumentales, por lo que el efecto suelo puede limitar la entrada de personas severamente lesionadas hasta que puedan caminar con la instrumentación requerida.

### Ejemplo

Aunque los estudios que analizan las propiedades estadísticas de los instrumentos de valoración son limitados, Middleton A. y Fritz S. (2013) informaron evidencia sobre la evaluación precisa de la marcha en adultos mayores, ya que es un aspecto importante de la práctica clínica para los médicos que trabajan con esta población. Las medidas seleccionadas son las siguientes: velocidad de la marcha, simetría de la marcha, resistencia de la marcha, adaptabilidad de la marcha, desempeño de la tarea dual durante la marcha y confianza autoinformada durante la marcha. La Figura 13 proporciona un resumen de las medidas de resultado incluidas en el artículo como referencia rápida para los médicos.

Measure	Assessing	Predictive abilities	Scale <sup>a</sup>	MDC <sub>95</sub>
4-Meter Walk Test [20•, 21, 24, 25, 33••, 35]	Gait speed	Functional status, discharge location, rehabilitation potential, fall risk, mortality	n/a	0.14 m/s
6-Minute Walk Test [35, 38]	Gait endurance	Not established	n/a	58.2 m
Dynamic Gait Index [39]	Ability to adapt to changing task demands during gait	Fall risk	0-24	2.9 points
Walking While Talking Test [48]	Gait performance under divided attention conditions	Fall risk, frailty, disability, mortality	n/a	Not established
Modified Gait Efficacy Scale [49•]	Patients' perceptions of gait abilities	Not established	10-100	14.7 points

Figura 13 – Table resumen de las variables evaluadas en la marcha. MDC, mínimo cambio detectable (*Minimal detectable change*). n/a, no aplicable.

El cambio mínimo detectable (MDC) proporciona a los médicos un valor de referencia para determinar si se ha producido un cambio real. El MDC95 cuantifica la menor cantidad de cambio necesaria para superar el error de medición y la variabilidad con un nivel de confianza del 95%. Cuando se informa en la literatura, las puntuaciones MDC para las medidas resultado pueden ser herramientas valiosas para los médicos. En la Figura 13 es posible observar el cambio mínimo detectable para los resultados de la marcha más utilizados en la práctica clínica, medidos por los instrumentos mencionados en la primera columna.

Este estudio es útil para explicar que los pacientes que rinden por debajo del cambio mínimo detectable en cada una de las variables mencionadas, es preferible que se midan con otros instrumentos que permitan detectar cambios por debajo del valor referenciado como cambio mínimo detectable.



### 3. Ideas claves

---

- El personal médico debe conocer las características metodológicas y propiedades estadísticas a la hora de elegir una herramienta de evaluación de la marcha. Esto es necesario para evitar errores metodológicos y sesgos en los resultados medidos.
- En cuanto a usabilidad, las escalas y pruebas clínicas tienen la ventaja de que son posibles de desarrollar en poco tiempo, no requieren una formación especializada por parte del evaluador y pueden utilizarse en cualquier contexto como en la práctica clínica.
- El equipo necesario para utilizar las pruebas clínicas y las escalas es mucho menor y accesible que el equipo necesario para realizar una evaluación de la marcha con instrumentos de evaluación biomecánica.
- La cualidad más importante de las técnicas de evaluación biomecánica instrumental es que proporcionan datos objetivos obtenidos sin interpretación del evaluador (es decir, evaluación directa de una o más dimensiones del patrón de la marcha), por lo que su uso es principalmente en el área de investigación. Por el contrario, la información obtenida a través de escalas y pruebas clínicas está influenciada por la interpretación y percepción del evaluador.
- La alta precisión de las técnicas de medición instrumental les confiere la cualidad de ser más válidas para medir una característica de la marcha que las escalas o pruebas clínicas.
- La fiabilidad suele ser mejor en instrumentos biomecánicos porque la repetibilidad de la medición no depende del observador sino de otros factores, como realizar la medición con un protocolo estandarizado.
- Cuanto más preciso sea un instrumento de medición, más sensible será al cambio. La sensibilidad del equipo debe ser suficiente para medir una diferencia mínima clínicamente importante en los resultados que el profesional intenta observar en una población determinada.
- Las escalas y pruebas clínicas tienen una mayor tendencia a tener un efecto techo, es decir, las puntuaciones de los participantes se agrupan hacia el extremo superior (o la mejor puntuación posible) de la medida / instrumento. Por otro lado, las técnicas instrumentales tienen un efecto piso mayor, donde las puntuaciones de los participantes se agrupan hacia el extremo inferior. Esto se debe a que los pacientes podrían estar "en mejor situación" de lo que la medida podría captar o "peor" de lo que el instrumento puede medir.



## 4. Referencias

---

- [1] Gutierrez-Clavería M, Beroíza T, Cartagena C, Caviedes I, Céspedes J, Gutiérrez-Navas M, Oyarzún M, Palacios S, Schönfeldt P. Guidelines for the six-minute walk test. *Rev Chil Enf Respir* 2009; 25: 15-24.
- [2] Innerd P, Catt M, Collerton J, Davies K, Trenell M, Kirkwood T, Jagger C. A comparison of subjective and objective measures of physical activity from the Newcastle 85+ study. *Age Ageing*. 2015 Jul;44(4):691-4.
- [3] Crémers J, Phan R, Delvaux V, Garrauxa G. Construction and validation of the Dynamic Parkinson Gait Scale (DYPAGS). *Parkinsonism & Related Disorders*. Volume 18, Issue 6, July 2012, Pages 759-764.
- [4] Tinetti M.E. Performance-Oriented Assessment of Mobility Problems in Elderly Patients. *J Am Geriatr Soc*. 1986 Feb;34(2):119-26.
- [5] Wrisley D, Kumar N. Functional Gait Assessment: Concurrent, Discriminative, and Predictive Validity in Community-Dwelling Older Adults. *Phys Ther*. 2010 May;90(5):761-73.
- [6] Pinto R, Birmingham T, Leitch K, Atkinson H, Jones I, Giffin J.R. Reliability and validity of knee angles and moments in patients with osteoarthritis using a treadmill-based gait analysis system. *Gait & Posture* 80 (2020) 155-161.
- [7] Taherdoost H. Validity and Reliability of the Research Instrument: How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management (IJARM)*. Vol. 5, No. 3, 2016, Page: 28-36.
- [8] De Souza A, Costa N, de Brito E. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol. Serv. Saude, Brasília*, 26(3), Jul-Sep 2017.
- [9] Sullivan G. A primer on the Validity of Assessment Instruments. *J Grad Med Educ*. 2011.
- [10] Meng L, Millar L, Childs C, Buis A. A strathclyde cluster model for gait kinematic measurement using functional methods: a study of inter-assessor reliability analysis with comparison to anatomical models. *Computer methods in biomechanics and biomedical engineering*. *Comput Methods Biomech Biomed Engin*. 2020 Jun 16;1-10.
- [11] Geerse D, Coolen B, Roerdink M. Quantifying Spatiotemporal Gait Parameters with HoloLens in Healthy Adults and People with Parkinson's Disease: Test-Retest Reliability, Concurrent Validity, and Face Validity. *Sensors (Basel)*. 2020 Jun 5;20(11):3216.
- [12] Hee-jae Kim, Ilhyoek Park, Hyo joo Lee, On Lee. The reliability and validity of gait speed with different walking pace and distances against general health, physical function, and chronic disease in aged adults. *J Exerc Nutrition Biochem*. 2016;20(3):046-050.

- [13] Wrisley D, Marchetti G, Kuharsky D, Whitney S. Reliability, Internal Consistency, and Validity of Data Obtained With the Functional Gait Assessment. *Phys Ther.* 2004 Oct;84(10):906-18.
- [14] McHugh M. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012 Oct; 22(3): 276-282.
- [15] Lipsey, M. W. (1983). A scheme for assessing measurement sensitivity in program evaluation and other applied research. *Psychological Bulletin*, 94(1), 152–165.
- [16] Jaeschke R, Singer J, Guyatt GH. Ascertaining the minimal clinically important difference. *Cont Clin Trials.* 1989;10:407–415.
- [17] McGlothlin A. and Lewis R. Minimal Clinically Important Difference Defining What Really Matters to Patients. *JAMA* October 1, 2014 Volume 312, Number 13.
- [18] Bohannon R and Glenney S. Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review. *Journal of Evaluation in Clinical Practice* 20 (2014) 295–300.
- [19] Moissenet F, Armand S. Chapter 17: Qualitative and quantitative methods of assessing gait disorders. *Orthopedic Management of Children with Cerebral Palsy.* 2015 Ed. Nova Science Publishers, Inc. ISBN: 978-1-63483-318-9-
- [20] Jackson A, Carnel C, Ditunno J, Schmidt M. Boninger M, Schmeler M, Williams S, Donovan W. Outcome Measures for Gait and Ambulation in the Spinal Cord Injury Population. *J Spinal Cord Med.* 2008;31:487–499.
- [21] Feeny DH, Eckstrom E, Whitlock EP, Perdue LA. Agency for Healthcare Research and Quality, US. A Primer for Systematic Reviewers on the Measurement of Functional Status and Health-Related Quality of Life in Older Adults. September 2013.
- [22] Middleton A, Fritz S. Assessment of Gait, Balance, and Mobility in Older Adults: Considerations for Clinicians. *Curr Transl Geriatr and Exp Gerontol Rep* (2013) 2:205–214.



El apoyo de la Comisión Europea para la producción de esta publicación no constituye una aprobación del contenido, el cual refleja únicamente las opiniones de los autores, y la Comisión no se hace responsable del uso que pueda hacerse de la información contenida en la misma.

