

## Development of innovative training solutions in the field of functional evaluation aimed at updating of the curricula of health sciences schools



### Modul Biomechanik des Gehens

#### Lerneinheit C: Wie beurteile ich das Gangbild?

C.3: Was sind die Vorteile der Verwendung von instrumentellen Techniken gegenüber Skalen und körperlicher Untersuchung zur Beurteilung des Gangs?



## Index

1. EINLEITUNG UND ZIELSETZUNG	3
2. MERKMALE UND EIGENSCHAFTEN VON GANGBEWERTUNGSINSTRUMENTEN: VERGLEICH ZWISCHEN VERFÜGBAREN TECHNIKEN	4
2.1 Benutzerfreundlichkeit.....	4
2.2 Anforderungen an die Ausrüstung .....	6
2.3 Objektivität der Ergebnisse und statistische Auswertung .....	8
2.4 Gültigkeit .....	11
2.5 Verlässlichkeit .....	14
2.6 Veränderungssensibilität und Reaktionsfähigkeit .....	20
2.7 Boden- und Deckeneffekt.....	25
3. SCHLÜSSELIDEEN	28
4. REFERENZEN	29

## 1. Einführung und Zielsetzung

---

Sowohl die Bewertung des Gangs durch 1) Beobachtung, 2) klinische Skalen, Tests und Fragebögen, als auch 3) Bewertung mit instrumentellen Techniken sind nützlich, um das Gangbild einer Person zu kennen, aber es gibt grundlegende Unterschiede, die bei der Wahl der einen oder anderen Art von Bewertungstechnik oder einer Kombination aus beiden berücksichtigt werden müssen.

In dieser Lerneinheit werden die methodischen Merkmale und statistischen Eigenschaften der in Lerneinheit C.1 des Moduls Biomechanik des Gangs besprochenen Techniken zur Beurteilung des menschlichen Gangs besprochen.

Auf dieser Grundlage schlagen wir die folgenden Ziele vor:

1. Überprüfung der Vor- und Nachteile von Bewertungsmethoden für das menschliche Gangbild.
2. Die statistischen Eigenschaften der verfügbaren Gangbewertungsmethoden zu kennen.
3. Aufbau von technischem Wissen, das es medizinischen Fachkräften ermöglicht, die am besten geeignete Gangbeurteilungstechnik für ihren klinischen oder Forschungskontext auszuwählen.

## 2. Merkmale und Eigenschaften von Gangbeurteilungsinstrumenten: Vergleich zwischen verfügbaren Techniken

### 2.1 Benutzerfreundlichkeit

Benutzerfreundlichkeit bezieht sich auf die Leichtigkeit, mit der Menschen ein bestimmtes Werkzeug verwenden können, um ein bestimmtes Ziel zu erreichen. Angesichts der Berufspraxis eines jeden Gesundheitsprofils und in jedem Bereich werden Fachleute vor der Auswahl einer Bewertungstechnik (oder Behandlung) Fragen stellen wie: *Ist sie einfach zu bedienen? Braucht es viel Zeit? Ist es in meinem Arbeitsbereich durchführbar?* In diesem Sinne haben die verfügbaren Techniken zur Beurteilung des menschlichen Gangs (die in den vorherigen Einheiten besprochen wurden) unterschiedliche Antworten auf die oben gestellten Fragen.

Hinsichtlich der Benutzerfreundlichkeit erfordern instrumentelle Techniken mehrere Schritte, die in ein strenges Bewertungsprotokoll eingebettet sind, das befolgt werden muss, um sicherzustellen, dass der Vergleich zwischen Messungen aus den eigenen Daten des Patienten (z. B. wenn mehrere Bewertungen über einen Zeitraum hinweg durchgeführt werden) oder mit den Daten anderer Probanden gültig ist. Dies macht jede instrumentelle Technik komplexer in der Anwendung als eine Skala, ein Fragebogen oder ein klinischer Test (Tabelle 1). Vom Messprotokoll mit einer instrumentellen Technik ist der wahrscheinlich am komplexesten zu handhabende Schritt die Instrumentierung des Probanden, falls erforderlich, und die Nachbearbeitung der Daten nach Durchführung der Messung. Daher muss bei den instrumentellen Techniken zur Gangbeurteilung ein sehr relevanter Faktor berücksichtigt werden: die **Zeit** (Tabelle 1). Mit dem Fortschritt der verfügbaren Technologien können wir jedoch innerhalb einer Gruppe von instrumentellen Techniken zur Messung der gleichen Parameter verschiedene Szenarien finden.

#### Beispiel

##### **Biomechanische instrumentelle Bewertung der raum-zeitlichen Parameter des Gangs**

Mit einem Photogrammetriesystem und einem instrumentierten Gehweg ist es möglich, die raum-zeitlichen Ergebnisse zu messen, die den menschlichen Gang charakterisieren. Diese Parameter sind in erster Linie die Ganggeschwindigkeit, die Schrittlänge und die Schrittlänge, die Kadenz, die doppelte Stützzeit sowie die Stütz- und Schwungphasenzeit des Gangzyklus. Obwohl wir mit beiden Techniken das gleiche Ergebnis erreichen können, variiert die Zeit, die für jede der Techniken benötigt wird, erheblich.

Mit dem instrumentierten Gang (mit Drucksensoren) ist es nicht notwendig, den Probanden zu instrumentieren, da die geschätzte Person nur mit oder ohne Schuhwerk durch den Gang gehen sollte. Wenn die Geräteverwaltungssoftware die Parameter aus den aufgezeichneten Plantardrücken berechnet, ist keine weitere Behandlung zur Messung der Daten erforderlich. Somit benötigt der Bewerter keine übermäßige Ausbildung oder Kenntnisse, um das Gerät zu benutzen.

Andererseits ist bei einem Photogrammetriesystem immer die Instrumentierung des Probanden notwendig. Diese Instrumentierung besteht aus einem biomechanischen Modell mit Landmarken, die im Körper platziert werden, um die Segmente zu repräsentieren, die an der zu bewertenden Bewegung beteiligt sind. Je nach Komplexität des Modells wird die Instrumentierungszeit länger oder kürzer sein. In jedem Fall muss die Haut, an der der Marker angebracht ist, immer gereinigt und rasiert werden. Andererseits gibt die Verwaltungssoftware der Photogrammetriesysteme in der Regel als Ergebnis die Koordinaten jedes Markers in jeder Bewegungsachse und zu jedem Aufnahmezeitpunkt an (zum Beispiel 50 Positionen pro Sekunde, wenn die Aufnahmefrequenz 50 Hz beträgt). Sofern eine Software nicht ein Protokoll mit einem geschlossenen und festen Modell anbietet und die verschiedenen Parameter, die aus der Messung gewonnen werden können, vorher berechnet, muss immer eine Berechnung der Ergebnisse vorgenommen werden, nachdem die Koordinaten jeder Landmarke aus dem biomechanischen Modell extrahiert wurden. All dies setzt voraus, dass der Auswerter ein tieferes Wissen hat und die korrekte Platzierung der Marker übt.

	<i>Instrumented walkway</i>	<i>Photogrammetry system</i>
		
Subject instrumentation:	NO	YES
Data processing after measurement:	NO	YES
Training evaluator :	NO	YES
Approximate assessment time:	5 minutes	1 hour

Auf der anderen Seite erfordert die Beobachtung der Ganganalyse oder die Bewertung durch klinische Skalen einen minimalen Zeitaufwand und das notwendige Training für die Bewertung des Gangs ist mit der Übung der Anweisungen der Skala selbst oder des Tests verbunden. Aus all diesen Gründen hat die Verwendung jeder Messtechnik in verschiedenen Bereichen mehr Nutzen. Die Verwendung der Gangbeurteilung mit Skalen oder klinischen Tests oder der Ganganalyse mit dem Auge ist im Rahmen der täglichen klinischen Konsultation, bei der Patienten mit verschiedenen Pathologien, die zu Gangstörungen führen, vorgestellt werden, einfacher zu verwenden. Auf der anderen Seite werden instrumentelle Techniken normalerweise in Forschungsprojekten eingesetzt. Dennoch werden im Rahmen der Forschung Skalen und validierte klinische Tests in verschiedenen Populationen und Sprachen häufig verwendet.

Tabelle 1 - Benutzerfreundlichkeit von Gangbewertungsinstrumenten

Kennlinie	Beobachtung Ganganalyse	Fragebogen, Skalen und klinischer Test	Instrumentelle Techniken
-----------	-------------------------	--	--------------------------

Zeitaufwand	+	+	+ / ++/ +++ (abhängig von dem verwendeten System)
Auswerter-Schulung	+	+	++ / +++ (abhängig von dem verwendeten System)
Kontext der Verwendung	Klinisch	Klinik und Forschung	Forschung
Benutzerfreundlichkeit	+	++	+++

## 2.2 Anforderungen an die Ausrüstung

Während für die Gangbeurteilung mit einer Skala oder mit einem Fragebogen nicht mehr als das entsprechende Formular und ein Stift benötigt werden, können klinische Tests eine bestimmte Menge an Materialien erfordern, um den entsprechenden Aufbau zusammenzustellen. Einige Tests benötigen einen Stuhl, ein Bodenmarkierungsband und eine Stoppuhr, während andere eine größere Menge an Materialien benötigen (siehe das Beispiel unten). In jedem Fall sind diese Materialien in der Regel kostengünstig (Tabelle 2) oder für jede medizinische Fachkraft zugänglich.

### Beispiel

#### Wie viele Materialien benötige ich für den 6-Minuten-Gehtest?

Der 6-Minuten-Gehtest (6MWT) wurde auch als einmalige Messung des Funktionsstatus von Patienten sowie als Prädiktor für Morbidität und Mortalität verwendet. Der 6MWT ist ein praktischer, einfacher Test, der einen 100-Fuß-Flur, aber keine Übungsgeräte oder fortgeschrittene Ausbildung für Techniker erfordert. Dieser Test misst die Distanz, die ein Patient in einem Zeitraum von 6 Minuten auf einer flachen, harten Oberfläche schnell gehen kann (die 6MWD). Er bewertet die globalen und integrierten Reaktionen aller Systeme, die während der Belastung beteiligt sind, einschließlich des Lungen- und Herz-Kreislauf-Systems, des systemischen Kreislaufs, des peripheren Kreislaufs, des Blutes, der neuromuskulären Einheiten und des Muskelstoffwechsels. Der selbstgesteuerte 6MWT bewertet das submaximale Niveau der funktionellen Kapazität.

Die erforderliche Ausrüstung ist:

1. Countdown-Timer (oder Stoppuhr)
2. Mechanischer Rundenzähler
3. Zwei kleine Kegel zur Markierung der Wendepunkte
4. Ein Stuhl, der leicht entlangbewegt werden kann
5. Arbeitsblätter auf einem Klemmbrett (Anmeldeformular)
6. Klebeband oder farbige Aufkleber zur Markierung des Aufenthaltsortes des Patienten nach Minuten
7. Borg-Skala



8. Pulsoximeter
9. Sphygmomanometer und Stethoskop
10. Telefon
11. Eine Quelle für Sauerstoff
12. Automatisierter elektronischer Defibrillator

*6-Minuten-Gehtest einrichten*

Wenn es jedoch um die Beurteilung mit instrumentellen Techniken geht, ist das benötigte Material nicht nur mengenmäßig größer, sondern auch hoch spezialisiert. Im Allgemeinen besteht eine Messausrüstung zur Erfassung des Gangbildes oder eines bestimmten Merkmals des Gangs aus folgenden Teilen:

- **Sensoren oder/und Messgeräte** beziehen sich auf die Geräte, aus denen die Messgeräte selbst bestehen. Zum Beispiel:
  - Eine dynamometrische Plattform besteht aus einer starren, flachen und statischen Oberfläche, die auf dem Boden installiert ist, und den Sensoren oder Aufnehmern, die am unteren Ende der Plattform angebracht und so verteilt sind, dass sie die auf die obere starre Oberfläche ausgeübten Kräfte erfassen können.
  - Die Elektromyographie-Ausrüstung besteht aus Elektroden (Sensoren), die die elektrische Aktivität im Muskel erfassen, entweder durch Einführen oder durch die Haut, die den Muskel bedeckt. Das System umfasst auch den Verstärker, mit dem die elektrischen Signale aus dem Muskel an einen Monitor übertragen werden können, in der Regel drahtlos, sowie die Verkabelung zwischen den Sensoren und dem Verstärker.
- **Software und Computer:** Es handelt sich um die Software, die die Verwaltung der Registriergeräte und Sensoren ermöglicht, die Speicherung der registrierten Daten und den Computer, auf dem die Software verwendet wird.
- **Zubehör:** Bezieht sich auf alle Arten von Zubehör, die zur Durchführung einer Gangmessung erforderlich sind. Zum Beispiel:
  - Bei einem Elektromyographiesystem ist das notwendige Zubehör dasjenige, das dazu dient, die Haut, auf der die Elektrode platziert wird, vorzubereiten und somit die Impedanz zwischen der Elektrode und der Haut zu reduzieren: Rasierer, Alkohol, Baumwolle und feines Schleifpapier, um abgestorbene Haut zu entfernen. Außerdem benötigen Sie ein leitfähiges Gel, das ebenfalls dazu beiträgt, die Haut-Elektroden-Impedanz zu reduzieren.

**Beispiel**

**Wie viele Materialien benötige ich für eine Gangauswertung mit einem Photogrammetriesystem**



Abbildung 1 - Photogrammetriesystem und seine Komponenten

Neben den Kameras, der Software und dem Computer ist es im Photogrammetriesystem unerlässlich, eine Reihe von Landmarken zu haben, um das biomechanische Modell darzustellen. Damit diese gut auf der Haut haften, werden außerdem ein System von Klebstoffen, Elemente zum Reinigen der Haut und zum Verhindern des Herunterfallens während des Tests (Rasierer, Alkohol, Watte) sowie die für die Durchführung des Tests erforderliche Kleidung benötigt.

Wie man sieht, sind die Materialien und das Zubehör der instrumentellen Techniken zur Gangbeurteilung in der Regel nicht in medizinischen Diensten zu finden. Die Kosten für die Geräte und die Software können je nach Bewertungsinstrument und den Möglichkeiten der Software enorm variieren (Tabelle 2). Einige Software bietet beispielsweise normative Daten, mit denen die an einem Patienten bestimmten Geschlechts und Alters vorgenommene Beurteilung verglichen werden kann, was den Preis erhöht. Geräte ohne jegliche Instrumentierung, wie z. B. ein Gang mit Drucksensoren und die dazugehörige Software, können je nach Länge des Ganges etwa 40 Tausend Euro kosten, während Photogrammetriegeräte mit Kameras und Software leicht 100 Tausend Euro überschreiten können.

Tabelle 2 - Anforderungen an Hilfsmittel zur Gangbeurteilung

Kennlinie	Beobachtung Ganganalyse	Fragebogen, Skalen und klinischer Test	Instrumentelle Techniken
Ausrüstung	+	+	+++
Zubehör	-	+	++
Wirtschaftliche Kosten	+	+	+++

### 2.3 Objektivität der Ergebnisse und statistische Auswertung

Die objektiven Daten einer Gangbeurteilung beziehen sich auf Daten, die ohne Interpretation des Beurteilers gewonnen werden (d. h. direkte Beurteilung einer oder mehrerer Dimensionen des Gangbildes), während bei subjektiven Daten das Ergebnis der Interpretation, Wahrnehmung oder Meinung des Beurteilers unterliegt. Bei den für die Gangbeurteilung verfügbaren Instrumenten finden wir beide Arten von Ergebnissen. Während wir mit instrumentellen Bewertungsinstrumenten objektive Messdaten erhalten können, liefert die Ganganalyse durch Beobachtung, Fragebögen oder klinische Skalen

subjektive Informationen. Aus diesem Grund ist es für den Mediziner sinnvoll, vor der Anwendung zu wissen, dass die Bewertungsskalen sehr nützliche Instrumente zur qualitativen Bestimmung des allgemeinen Gesundheitszustands sind; allerdings unterliegen sie in gewisser Weise der Subjektivität sowohl des Patienten (selbstberichtete Bewertung) als auch des Bewerter (Bewertung durch Beobachtung). Darüber hinaus wurden sie manchmal in sehr spezifischen Kontexten (Populationen, Regionen, Länder usw.) entwickelt, mit sehr wichtigen demographischen und kulturellen Verzerrungen, die eine Extrapolation auf andere Gemeinschaften schwierig oder unmöglich machen. Auf der anderen Seite liegt die große Stärke von Selbsteinschätzungen darin, dass sie wiedergeben, was die Person wirklich erlebt und wahrgenommen hat. Außerdem ist die Methode billig, schnell und praktikabel.

### Beispiel

#### Subjektive und objektive Messung der Schrittlänge

Viele Skalen zur Beurteilung des Gangs haben einige Items, die sich auf die Schrittlänge beziehen. Eine davon ist der Tinetti-Mobilitätstest (TMT), bei dem im Abschnitt Gangart gefragt wird, ob der Patient in der Lage ist, beim Schritt einen Fuß über den anderen zu setzen. Wenn wir die Schrittlänge bei einer Person ohne Pathologie messen, kann die Antwort deutlich sichtbar sein und der Beobachter wird zweifellos darauf hinweisen, dass der rechte Fuß den linken überragt (oder umgekehrt). Wird jedoch ein Patient mit Parkinson bewertet, bei dem eines der Merkmale der Krankheit das Gehen mit kleinen Schritten ist, kann es sein, dass die Reaktion auf das TMT-Item nicht so deutlich ist und zwei Forscher unterschiedliche Dinge beobachten. Andererseits ist ein Messgerät wie der instrumentierte Gehweg in der Lage, die Schrittlänge zu messen, indem der plantare Druck erfasst wird, wenn der Patient den Fuß auf den Boden setzt. Bei dieser Art von Instrument wird die aus der Schrittlänge gewonnene Distanz nicht durch die Beobachtung des Auswerters beeinflusst.

Step length and height	
The left foot does not pass right stance foot	0
The left foot passes right stance foot with step	1

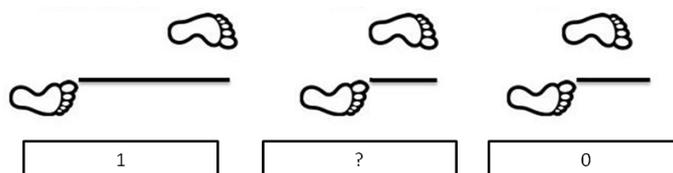
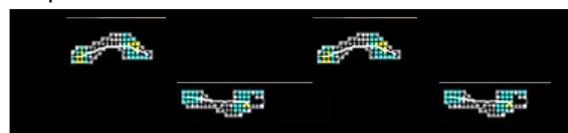
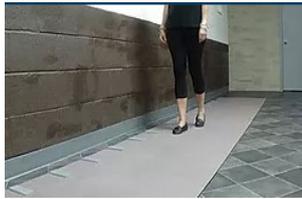


Abbildung 2 - Schrittlänge und -höhe ítem aus dem Tinetti-Mobilitätstest, Abschnitt Gangart. Es werden drei Beispiele für die Schrittlänge gezeigt, bei denen der Beurteiler entsprechend den Kriterien der Skala 1 oder 0 markieren muss. Im Beispiel der Mitte kann es zu Zweifeln kommen, welche Nummerierung zu vergeben ist.





Stufenlänge = 0,46 m

Abbildung 3 - Bewertung der Schrittlänge mit einem instrumentierten Gehweg (GAITrite)

Der Hauptvorteil der Verwendung von objektiven Ergebnissen, die mit instrumentellen Techniken gemessen werden, ist, dass diese unter anderem ermöglichen, dass die Daten mit anderen Daten desselben Patienten oder mit Ergebnissen von anderen Probanden vergleichbar sind. Der Forscher oder der professionelle Gutachter sollte jedoch nicht vergessen, dass der Vergleich von objektiven Daten zwischen Probanden durch das einflussreichste Merkmal auf diese Daten normalisiert werden sollte. Zum Beispiel wird die Schrittlänge auch von der Körpergröße des Probanden und damit von der Länge der unteren Gliedmaßen beeinflusst, daher sollten die objektiven Daten der Schrittlänge durch die Körpergröße oder die Länge der Beine geteilt werden. Das Gleiche geschieht mit Variablen wie den Reaktionskräften und dem Gewicht der Testpersonen.

Obwohl die Skalen oder Fragebögen und die instrumentellen Werkzeuge nicht die gleiche Art von Informationen liefern (subjektiv versus objektiv), können subjektive Messungen hoch mit objektiven Messungen korreliert sein. Dies ist ein weiterer Mehrwert der Bewertungsskalen, die in der klinischen Umgebung verwendet werden. Eine Skala oder ein Fragebogen, der hoch mit den Ergebnissen der Bewertung mit einer instrumentellen Technik korreliert ist, wird ein valides Werkzeug zur Messung von Populationen in großen Stichproben sein.

### Beispiel

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>

In der epidemiologischen Forschung wird die körperliche Aktivität üblicherweise mit Fragebögen zur körperlichen Aktivität erfasst, da diese praktisch und kostengünstig sind. Es wäre extrem teuer und komplex, große Stichproben von Teilnehmern mit einer instrumentellen Technik zu erfassen. Aus diesem Grund gibt es Studien, die die Korrelation von selbstberichteten Fragebögen zur körperlichen Aktivität mit objektiven Messungen, wie

z. B. der mit Beschleunigungsmessern aufgezeichneten körperlichen Aktivität, in einer kleinen Stichprobe der Zielbevölkerung für eine Woche untersuchen. Das Ziel dieser Art von Studien ist es also, die Validität des subjektiven Tests als Instrument zur Bewertung der körperlichen Aktivität für große Stichproben von Teilnehmern zu bewerten.

Sind die subjektiven Daten nun geeignet, statistisch analysiert zu werden? Die Antwort ist ja, sowohl die subjektiven Informationen, die durch Skalen oder Fragebögen gewonnen werden, als auch die objektiven Informationen, die durch eine instrumentelle Technik gewonnen werden, sind in der Lage, statistisch analysiert zu werden. Es ist klar, dass z. B. die Gehgeschwindigkeit als eine quantitative Variable betrachtet wird, d. h. es ist eine Variable, die numerische Größen als Modalitäten hat, mit denen wir arithmetische Operationen durchführen können (Tabelle 3).

Tabelle 3 - Objektivität der Ergebnisse und statistische Analyse

Kennlinie	Beobachtung Ganganalyse	Fragebogen, Skalen und klinischer Test	Instrumentelle Techniken
Objektivität	Subjektive Informationen	Subjektive Informationen	Objektive Informationen
Statistische Auswertung	Nein (a priori)	Ja	Ja
Typ der Variablen in der statistischen Analyse	-	Qualitativ, semi-quantitativ und quantitativ (abhängig von der zu analysierenden Antwort)	Quantitativ

Auf der anderen Seite können wir von einem Fragebogen oder einer Skala den Endwert analysieren. Wenn dieses Endergebnis eine Zahl ist, wird es statistisch als quantitative Variable analysiert, weil es das ist, was wir in das statistische Programm eingeben (eine Zahl), aber immer noch eine subjektive Information ist. Wenn wir jedoch jede einzelne Frage der Skala oder des Fragebogens der Analyse unterziehen, können wir verschiedene statistische Variablen erhalten (Tabelle 3), je nachdem, welche Art von Antworten wir finden können:

1. Die Dynamic Parkinson Gait Scale hat Fragen, bei denen der Bewerter eine Punktzahl zwischen 0 und 5 für die Leistung des Patienten beim Gehen vergeben muss. Zum Beispiel kann der Bewerter bei Item 1, das sich darauf bezieht, wie der Patient sieben Meter vorwärts geht, dieses Merkmal von normal (0 Punkte) bis "unfähig, einen Schritt vorwärts zu initiieren oder scheitert" (5 Punkte) bewerten. Obwohl es sich immer noch um kategoriale Modalitäten innerhalb der Frage handelt, haben diese Antwortmöglichkeiten eine Reihenfolge: eins ist schlechter als null, zwei

ist schlechter als eins, drei ist schlechter als zwei, und so weiter. Dies ist eine semi-quantitative Variable.

Dynamic Parkinson Gait Scale (DYPAGS)	
	Score
1. Walking 7 m forwards	
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

- Der Tinetti-Mobilitätstest hat Fragen, bei denen der Beobachter ein Merkmal des Gangs beim Patienten überprüfen muss und 0 Punkte vergibt, wenn das, was er beobachtet, eine negative oder veränderte Reaktion ist, oder 1 Punkt, wenn das bewertete Gangmerkmal als positiv oder normal dargestellt wird. Mit anderen Worten: Diese Art von Frage oder Skalen-Item ist eine Ja/Nein-Antwort. Bei einer statistischen Analyse würden wir es als eine qualitative, kategoriale Variable mit zwei Modalitäten betrachten: 0 = Merkmal eines veränderten Gangs und 1 = Merkmal eines normalen Gangs.

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

Wenn der Beobachter schließlich eine Analyse des Gangs des Patienten mit dem bloßen Auge durchführt und eine Reihe von Ideen oder Merkmalen über das von ihm beobachtete Gangbild erhält, sind diese Informationen subjektiv, aber da sie nicht standardisiert sind, können sie nicht mit konventionellen statistischen Methoden analysiert werden (Tabelle 3).

## 2.4 Gültigkeit

Um als geeignet zu gelten, müssen die Instrumente genaue, valide und interpretierbare Daten zur Beurteilung der Gesundheit der Bevölkerung liefern. Außerdem sollen die Maßnahmen wissenschaftlich belastbare Ergebnisse liefern. Die Leistungsfähigkeit der Ergebnisse dieser Maßnahmen ergibt sich aus der Reliabilität und Validität der Instrumente.

Die Validität bezieht sich auf die Tatsache, dass ein Instrument genau das misst, was es zu messen vorschlägt, oder anders gesagt, die Validität bezieht sich auf die Genauigkeit der Messung (Abbildung 4). Die Validität ist keine Instrumenteneigenschaft und muss in Bezug auf einen bestimmten Sachverhalt bestimmt werden, sobald sie sich auf eine definierte Population bezieht. Die Messeigenschaften - Validität und Reliabilität - sind nicht völlig unabhängig voneinander. Ein Instrument, das nicht reliabel ist, kann nicht valide sein;

umgekehrt kann ein reliables Instrument unter Umständen auch ungültig sein. Eine hohe Reliabilität garantiert also nicht die Validität eines Instruments.

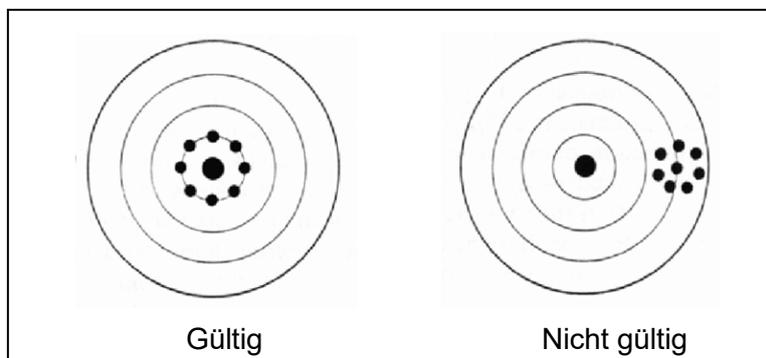


Abbildung 4 - Bei der linken Scheibe trafen die Schüsse die Stelle, die sie sollten, und waren konsistent, genau in der Scheibenmitte. In der rechten Scheibe waren die Schüsse zuverlässig und trafen denselben Punkt; allerdings hat keiner die Mitte der Scheibe getroffen, was nicht als gültig angesehen wird (Bild aus De Souza A. et al. 2017)

Die Validität von Beurteilungsinstrumenten erfordert mehrere Evidenzquellen, um den Beweis zu erbringen, dass das Instrument das misst, was es messen soll. Die Bestimmung der Validität kann als das Konstruieren eines evidenzbasierten Arguments darüber betrachtet werden, wie gut ein Instrument das misst, was es tun soll. Es können Belege zusammengestellt werden, die eine bestimmte Verwendung des Bewertungsinstruments unterstützen oder nicht unterstützen. Aus diesem Grund finden wir in der Bibliographie mehrere Studien, die versuchen, die Validität eines Instruments zur Messung des Gangs in einer bestimmten Zielpopulation zu untersuchen (Gail M. Sullivan 2011).

### Wie kann man die Validität eines Werkzeugs messen?

Die Technik oder das Werkzeug, dessen Validität untersucht werden soll, muss mit einem "Goldstandard"-Werkzeug verglichen werden, dessen Validität zur Messung eines bestimmten Ergebnisses bereits ausgiebig getestet wurde (konkurrierende Validität). Bei dieser Art von Studie geht es um die Beantwortung der Frage: Misst Werkzeug A genauso genau wie Werkzeug B den menschlichen Gang? Dies wird üblicherweise mit dem Pearson oder Spearman Korrelationskoeffizienten ( $r$ ) analysiert. Er gilt als exzellent, wenn der Korrelationskoeffizient zwischen dem Messgerät und dem Referenzstandard über 0,6 liegt, adäquat zwischen 0,59 und 0,31 und schlecht unter 0,3.

#### Beispiel 1: Gültigkeit der Skala

Diane M. Wrisley und Neeraj A. Kumar (2010) untersuchten die Validität des Functional Gait Assessment (FGA) zur Klassifizierung des Sturzrisikos und zur Vorhersage ungeklärter Stürze bei in der Gemeinschaft lebenden älteren Erwachsenen. Zu diesem Zweck füllten fünfunddreißig ältere Erwachsene im Alter von 60 bis 90 Jahren zusätzlich zu dem zu validierenden Tool (FGA) die Activities specific Balance Confidence Scale (ABC),

die Berg Balance Scale (BBS), den Dynamic Gait Index (DGI) und den Timed "Up & Go" Test (TUG) aus. Spearman-Korrelationskoeffizienten wurden verwendet, um die gleichzeitige Validität zwischen ABC, BBS, TUG, DGI und FGA zu bestimmen. Der FGA korrelierte mit dem ABC ( $r = .053$ ,  $P < .001$ ), BBS ( $r = .84$ ,  $P < .001$ ) und TUG ( $r = .84$ ,  $P < .001$ ). Die Autoren stellten fest, dass der FGA mit einem Cutoff-Score von 22/30 effektiv bei der Klassifizierung des Sturzrisikos bei älteren Erwachsenen und der Vorhersage unerklärter Stürze bei in der Gemeinschaft lebenden älteren Erwachsenen ist.

### Beispiel 2: Gültigkeit der dreidimensionalen Ganganalyse auf dem Laufband

Pinto R. et al. (2020) ermittelten die Validität von dreidimensionalen (3D) Gangbiomechaniken, die von laufbandbasierten Systemen abgeleitet wurden. Hierfür untersuchten die Forscher die gleichzeitige Validität, indem sie die Assoziationen zwischen laufbandbasierten und oberirdischen (Goldstandard) Messungen bei Teilnehmern mit Kniearthrose schätzten. Die Gehgeschwindigkeit auf dem Laufband wurde an die selbst gewählte Geschwindigkeit auf dem Boden angepasst. Markerset, Kniewinkel- und Momentberechnungen waren für beide Systeme konsistent. Variablen, die aus den Kniewinkel- und Moment-Gangwellenformen während des Standes berechnet wurden, wurden neben anderen statistischen Parametern mit Pearson-Korrelationen ( $r$ ) ausgewertet. Die Ergebnisse zeigten, dass die Pearson-Korrelationen zwischen dem Laufband- und dem Bodensystem von 0,56 bis 0,97 reichten. Obwohl hoch assoziiert, gab es erhebliche Unterschiede in den Momenten, was unterstreicht, dass sie nicht austauschbar verwendet werden können. Dies deutet darauf hin, dass die Kniewinkel und Momente in der Frontal- und Sagittalebene bei Patienten mit Kniearthrose, die mit einem Laufband-basierten System bewertet wurden, gültig sind.

Nachdem wir nun die Validität definiert haben und wie sie funktioniert, lohnt es sich zu fragen: *Welche Arten von Werkzeugen haben die größte Validität, um den Gang oder ein bestimmtes Merkmal des Gangs zu messen? Skalen / klinische Tests oder die instrumentellen Techniken?* Die Antwort ist klar, instrumentelle Messtechniken sind für die Messung des Gangs valider als Skalen oder klinische Tests, weil sie präzisere Instrumente sind, um eine bestimmte Variable des Gangs zu messen. Es ist erwähnenswert, dass es sogar innerhalb der instrumentellen Techniken zur Ganganalyse einige gibt, die präziser sind als andere. Diese Präzisionsunterschiede hängen mit dem Messfehler der Geräte zusammen, der in der Methodik selbst impliziert ist. Abbildung 5 zeigt einen Vergleich der verfügbaren instrumentellen Techniken zur Messung der raum-zeitlichen Parameter des Gangs wie Geschwindigkeit, Schrittlänge, Kadenz, Dauer des Gangzyklus und Gangphasen.

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

Abbildung 5 - Vergleich der gängigen Technologien zur Messung raum-zeitlicher Gangparameter (Moissenet F. und Armand S. 2016). Für jede instrumentelle Technik werden der Grad der Präzision und die Kosten der Technik genannt.

## 2.5 Verlässlichkeit

Auf der anderen Seite ist die Reliabilität die Fähigkeit, ein konsistentes Ergebnis in Zeit und Raum oder von verschiedenen Beobachtern zu reproduzieren (Abbildung 6) und stellt Aspekte der Kohärenz, Stabilität, Äquivalenz und Homogenität dar. Sie ist eines der Qualitätskriterien eines Instruments (de Souza et al. 2017). Die Reliabilität befasst sich auch mit der Wiederholbarkeit. Zum Beispiel wird eine Skala oder ein Test als zuverlässig bezeichnet, wenn wiederholte Messungen unter konstanten Bedingungen das gleiche Ergebnis liefern (Taherdoost H. 2016).

Es ist wichtig zu betonen, dass die Reliabilität keine feste Eigenschaft eines Fragebogens ist. Im Gegenteil, die Reliabilität hängt von der Funktion des Instruments, von der Population, in der es verwendet wird, von den Umständen und vom Kontext ab; das heißt, dass das gleiche Instrument unter verschiedenen Bedingungen nicht als zuverlässig angesehen werden kann. Reliabilitätsschätzungen werden von verschiedenen Aspekten der Bewertungsumgebung (Rater, Stichprobencharakteristika, Art des Instruments, Administrationsmethode) und von der verwendeten statistischen Methode beeinflusst. Daher können die Ergebnisse einer Untersuchung mit Messinstrumenten nur interpretiert werden, wenn die Beurteilungsbedingungen und der statistische Ansatz klar dargestellt werden. Abschließend bezieht sich die Reliabilität darauf, ob ein Bewertungsinstrument jedes Mal, wenn es in der gleichen Umgebung mit der gleichen Art von Probanden verwendet wird, die gleichen Ergebnisse liefert (Sullivan G.M. 2011).

Die Wahl der statistischen Tests, die zur Beurteilung der Zuverlässigkeit verwendet werden, kann variieren, je nachdem, was gemessen werden soll. Es gibt drei wichtige Reliabilitätskriterien, die für Forscher von großem Interesse sind: 1) Stabilität, 2) interne Konsistenz und 2) Äquivalenz.

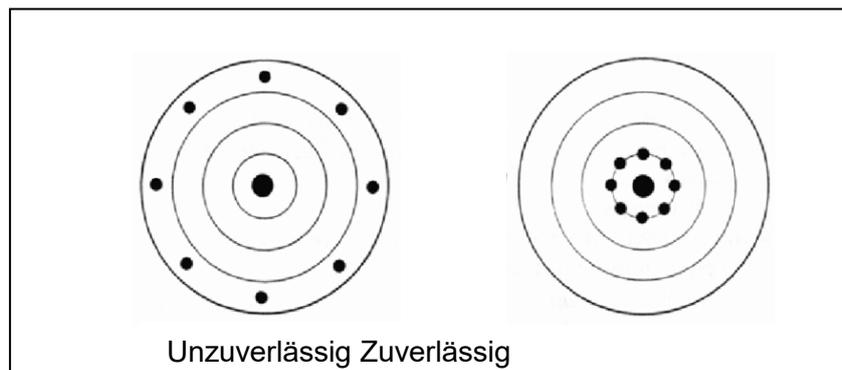


Abbildung 6 - Bei der linken Scheibe trafen die Schüsse nicht zuverlässig, da die getroffenen Punkte nicht an einer bestimmten Stelle lagen, sondern über die gesamte Scheibe verteilt waren. In der rechten Scheibe waren die Schüsse zuverlässig und trafen denselben Punkt (Bild aus De Souza A. et al. 2017).

### Wie wird die Reliabilität eines Instruments gemessen?

Wie bereits erwähnt, hängt das statistische Verfahren zur Messung der Zuverlässigkeit davon ab, was gemessen werden soll.

1) **Stabilität:** Die Stabilität misst, wie ähnlich die Ergebnisse sind, wenn sie zu zwei verschiedenen Zeitpunkten gemessen werden, d. h. sie schätzt die Konsistenz der Messwiederholung. Die Bewertung der Stabilität kann mit der Test-Retest-Methode durchgeführt werden, d. h. das Verfahren besteht darin, dieselbe Messung zu zwei verschiedenen Zeitpunkten anzuwenden. Der Intraklassen-Korrelationskoeffizient (ICC) ist einer der am häufigsten verwendeten Tests zur Abschätzung der Stabilität kontinuierlicher Variablen, da er die Messfehler berücksichtigt. Für die Interpretation der Ergebnisse werden Mindestwerte von 0,70 als zufriedenstellend angesehen (de Souza et al. 2017).

2) **Interne Konsistenz:** Die interne Konsistenz (oder Homogenität) zeigt, ob alle Teilbereiche eines Instruments das gleiche Merkmal messen. Dies ist eine wichtige Messeigenschaft für Instrumente, die ein einzelnes Konstrukt bewerten, das eine Vielzahl von Items verwendet. Eine niedrige interne Konsistenz kann darauf hinweisen, dass die Items unterschiedliche Konstrukte messen (de Souza et al. 2017). Bei der internen Konsistenzanalyse wird die Differenz zwischen den Antworten der Items eines Konstrukts berechnet. Cronbach Alpha ist ein Test, der häufig verwendet wird, um die Korrelationswerte zwischen den Antworten eines Assessment-Instruments zu berechnen.

Eine hohe Reliabilitätsschätzung sollte möglichst nahe bei 1 liegen (Sullivan G.M. 2011).

3) *Äquivalenz*: Äquivalenz ist der Grad der Übereinstimmung von zwei oder mehr Beobachtern bezüglich eines Instrumentenwertes. Die gebräuchlichste Methode zur Beurteilung der Äquivalenz ist die Inter-Beobachter-Reliabilität, die die unabhängige Teilnahme von zwei oder mehreren Ratern beinhaltet. Der Kappa-Koeffizient ist ein Maß zur Beurteilung der Inter-Beobachter, das auf kategoriale Variablen angewandt wird und einen Maximalwert von 1 hat. Je höher der Kappa-Wert ist, desto höher ist die Übereinstimmung zwischen den Ratern (Sullivan G.M. 2011).

### Beispiel 1: Stabilität Zuverlässigkeit

#### Instrumentaltechnik

##### *Inter-Rater-Test-Retest (unterschiedliche Beurteiler)*

Meng L. *et al.* (2020) untersuchten, ob eine geringere Abhängigkeit von der manuellen Identifizierung anatomischer Landmarken die Inter-Assessor-Reliabilität der Gelenkinematik im Vergleich zu drei kinematischen Gangmodellen verbessern könnte. Dieses Studienziel wurde erhoben, weil eine Hauptfehlerquelle für die Zuverlässigkeit der Ganganalyse aus der Palpation anatomischer Landmarken resultiert. Die Hypothese der Studie war, dass das funktionelle Strathclyde-Cluster-Modell (SFCM) eine größere Inter-Assessor-Reliabilität als die anatomischen Modelle erzielen würde: Plug-in gait (PIG) und das Human body model gait (HBM2). Um dies zu demonstrieren, absolvierten 10 Teilnehmer drei Versuche, die von verschiedenen Assessoren an nicht aufeinanderfolgenden Tagen durchgeführt wurden. In jeder Sitzung legte der Untersucher den Teilnehmern das kombinierte Marker-Set an. Dann wurde ein statischer Versuch für 5 s aufgezeichnet, wobei der Teilnehmer in einer natürlichen Haltung stand. Danach ging jeder Teilnehmer auf dem Laufband mit seiner angenehmen Geschwindigkeit. Nach einer zweiminütigen Eingewöhnungsphase wurden 30 Sekunden Daten aufgezeichnet und die markierten Trajektorien wurden von 12 Kameras mit einer Abtastrate von 100 Hz aufgenommen. Der Intraklassenkorrelationskoeffizient über die Gangzyklen wurde zum Vergleich der Interassesspr-Reliabilität verwendet und der Wert der ICC ( $\alpha$ ) wurde in vier Stufen der Reliabilität interpretiert:  $\alpha < 0,5$  = schlecht,  $0,5 < \alpha < 0,75$  = mäßig,  $0,75 < \alpha < 0,9$  = gut und  $\alpha > 0,9$  = ausgezeichnet. Die Ergebnisse zeigen, dass alle Modelle eine "gute bis exzellente" Inter-Assessor-Reliabilität für alle Flexions-/Extensionswinkel und Hüftab-/Adduktionswinkel zeigten, aber eine "schlechte bis mäßige" Inter-Assessor-Reliabilität für andere nicht-sagittale Winkel. Das SFCM erreichte jedoch eine höhere Zuverlässigkeit bei geringerer Variation im Vergleich zu den anatomischen Modellen. Die Ergebnisse zeigen, dass die SFCM für weniger erfahrene Untersucher vorteilhafter sein könnte.



Abbildung 7 - Marker-Layout des Strathclyde-Funktionscluster-Modells (links) und ein kombiniertes Marker-Setup für das SFCM-, PiG- und HBM2-Modell (rechts).

### *Intra-Rater-Test-Retest (unterschiedliche Beurteilungszeitpunkte)*

Geerse D. et al. (2020) ermittelten die Test-Retest-Reliabilität der HoloLens von Microsoft, einem Mixed-Reality-Headset, das neben Hologrammen auch reichhaltige Positionsdaten des Kopfes liefert, die zur Quantifizierung verschiedener Aufgaben wie dem Gehen verwendet werden können. Um zu untersuchen, wo die Grenzen der Übereinstimmung von abgeleiteten raum-zeitlichen Gangparametern bei Wiederholungen liegen, wurden 23 gesunde junge Erwachsene, die mit langsamer, bequemer und schneller Geschwindigkeit gehen, sowie 24 Menschen mit Parkinson-Krankheit, die mit selbst gewählter Geschwindigkeit gehen, untersucht. Die HoloLens liefert Positions- und Orientierungsdaten des Kopfes in der Umgebung in drei Richtungen, und die Positionsdaten wurden verwendet, um raum-zeitliche Gangparameter abzuleiten. Die Test-Retest-Reliabilität wurde mit dem Intraclass-Korrelationskoeffizienten für absolute Übereinstimmung bewertet. ICCs über 0,60 und 0,75 bedeuteten gute bzw. ausgezeichnete Übereinstimmung. Die ICCs zeigten eine exzellente Test-Retest-Reliabilität für beide Systeme bei gesunden Erwachsenen und Menschen mit Parkinson-Krankheit.

			Trial 1	Trial 2	ICC <sub>(A,1)</sub>
			Mean ± SD	Mean ± SD	
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

Abbildung 8 - Test-Retest-Reliabilität für raum-zeitliche Gangparameter der instruierten langsamen Gehgeschwindigkeit (SWS), komfortablen Gehgeschwindigkeit (CWS) und schnellen Gehgeschwindigkeit (FWS) bei gesunden jungen Erwachsenen (HYA) und CWS bei Menschen mit Morbus Parkinson (PD) abgeleitet aus HoloLens-Daten (Geerse D. et al. 2020).

Erwartungsgemäß sind die ICC-Werte in den Reliabilitätsanalysen einer instrumentellen Technik hoch, da die Messverfahren in der Regel von hoher Präzision sind. Faktoren wie das Befolgen eines strengen Auswerteprotokolls tragen zur Aufrechterhaltung einer hohen Test-Retest-Reliabilität bei. Wenn der Titrator eine instrumentelle Messtechnik verwendet, ohne das gleiche Messprotokoll in einer standardisierten Weise für die verschiedenen Beurteilungszeitpunkte zu befolgen, ist es wahrscheinlich, dass die ICC-Werte nicht die in der Beispielstudie beobachteten sind. Es ist auch zu beobachten, dass die Zuverlässigkeit der Messtechnik in der Kadenz der Teilnehmer mit Morbus Parkinson abnimmt. Dies liegt daran, dass diese Patientengruppe die Trittfrequenz in Abhängigkeit von der Schrittlänge verändert, die durch die Krankheit verändert ist. Dieses Beispiel ist nützlich, um darauf hinzuweisen, dass die Zuverlässigkeit der Messung, auch wenn es sich um eine instrumentelle Technik handelt, in verschiedenen Populationen variiert.

### Klinischer Test

Reliabilitätsanalysen werden vom klinischen Personal bei häufig verwendeten klinischen Tests noch mehr geschätzt. Hee-jae Kim et al. (2016) untersuchten die Reliabilität (Intra-Rater Test-Retest) der bei verschiedenen Entfernungen und Geschwindigkeiten gemessenen Ganggeschwindigkeit bei älteren Koreanern. Vierundfünfzig weibliche Teilnehmerinnen im Alter von ≥70 Jahren wurden aus einer lokalen Seniorengemeinschaft rekrutiert. Die Ganggeschwindigkeit wurde über 4, 6 und 10 Meter sowie im normalen und

schnellen Gehmodus gemessen. Die Teilnehmer wurden angewiesen, aus dem Stand in einem für sie normalen und angenehmen Tempo zu gehen oder so schnell zu gehen, wie sie konnten, bis sie das Ende des markierten Weges erreichten. Ein geschulter Tester ging hinter dem Teilnehmer und stoppte die Zeitmessung, wenn der Fuß des Teilnehmers am Ende der Gehstrecke den Boden berührte. Die Gehgeschwindigkeit wurde also aus der zurückgelegten Strecke und der Zeit, die die Teilnehmer dafür benötigten, berechnet. Die Ergebnisse sind in Abbildung 9 dargestellt.

	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*

Abbildung 9 - Reliabilität des Gehens (ICC) für verschiedene Distanzen und Geschwindigkeiten.

Höhere ICC-Werte wurden bei der längsten Gehstrecke von 10 Metern im Vergleich zu 4 und 6 Metern beobachtet. Darüber hinaus waren die ICC-Werte des Gehversuchs bei maximaler Geschwindigkeit höher als bei normalem Tempo. Die Autoren erwähnten, dass, obwohl der Gehstest bei maximalem Tempo über eine längere Distanz eine bessere Zuverlässigkeit bei älteren Personen aufweist, die Testdistanz und das Tempo nach dem Zweck der Messungen und dem klinischen Gesundheitszustand der Teilnehmer und nicht nach dem Kriterium einer hohen Zuverlässigkeit betrachtet werden müssen.

## Beispiel 2: Interne Konsistenz-Reliabilität

### Klinischer Test

Im Gegensatz zu instrumentellen Werkzeugen, mit denen wir eine spezifische Variable wie Ganggeschwindigkeit, Schrittlänge oder Gelenkbewegungswinkel messen, erlauben uns klinische Tests, den Gang von einer globaleren Leistung aus zu beurteilen. Dies ist der Fall bei Das Functional Gait Assessment (FGA), ein Maß für die Gleichgewichtsfähigkeit beim Gehen, wurde entwickelt, um den Deckeneffekt zu beseitigen, der beim Dynamic Gait Index (DGI) beobachtet wurde. Zu den ursprünglichen 8 DGI-Aufgaben wurden drei vermutlich schwierigere Aufgaben hinzugefügt und 1 leichtere Aufgabe entfernt (Beninato M. und Ludlow L., 2016). Die Items des FGA sind: 1) Gang auf ebener Fläche, 2) Veränderung der Ganggeschwindigkeit, 3) Gang mit horizontalen Kopfdrehungen, 4) Gang mit vertikalen Kopfdrehungen, 5) Gang mit Drehung, 6) Schritt über ein Hindernis, 7) Gang mit schmaler Auflagefläche, 8) Gang mit geschlossenen Augen, 9) Rückwärtsgehen und 10) Schritte. Für jedes Item kann der Bewerter eine Punktzahl von 0 bis 3 vergeben, so dass insgesamt eine maximale Punktzahl von 30 Punkten erreicht werden kann.

Wrisley D. et al. (2004) untersuchten die interne Konsistenz der Daten, die mit dem FGA bei Menschen mit vestibulären Störungen erhoben wurden. Sechs Patienten mit vestibulären Störungen füllten den FGA zweimal aus, mit einer einstündigen Pause zwischen den Sitzungen. Die interne Konsistenz, oder die Homogenität, der in der FGA enthaltenen Items wurde mit Hilfe des Cronbach-Alpha bestimmt. Diese Bewertung wurde über beide Testsitzungen und innerhalb der einzelnen Tests durchgeführt. Der Cronbach-Alpha-Wert variiert von 0 bis 1, die Konsistenz ist hoch, wenn der Alpha-Wert größer als 0,8 ist.

Die FGA zeigte interne Konsistenz innerhalb und über beide FGA-Testversuche für jeden Patienten. Die Cronbach-Alpha-Werte betragen .81 und .77 für die einzelnen Versuche 1 bzw. 2. Das Cronbach-Alpha betrug .79 über beide Versuche hinweg. Dies ist eine der Analysen, die zur Validierung eines itembasierten Messinstruments durchgeführt werden. Der Vorteil dieser Art von Test ist, dass wir einen Score für die Gesamtleistung der Patienten in einem Bereich erhalten. So erhält man ein viel funktionaleres Maß als ein einzelnes Ergebnis, das mit einer instrumentellen Technik gemessen wurde.

### Beispiel 3: Äquivalenzsicherheit

#### Klinischer Test

In der gleichen Studie von Wrisley D. et al. (2004) wurde auch die Äquivalenz oder der Konkordanzgrad von zwei oder mehr Beobachtern bezüglich des Functional Gait Assessment (FGA) durch den Kappa-Koeffizienten bewertet. Für diesen Versuch meldeten sich sieben Physiotherapeuten aus verschiedenen Praxen und 3 Physiotherapiestudenten freiwillig zur Teilnahme. Der Aufbau für diese Bewertung war wie in Abbildung 10 dargestellt, wobei alle Bewerter dieselbe Patientenleistung beobachteten.

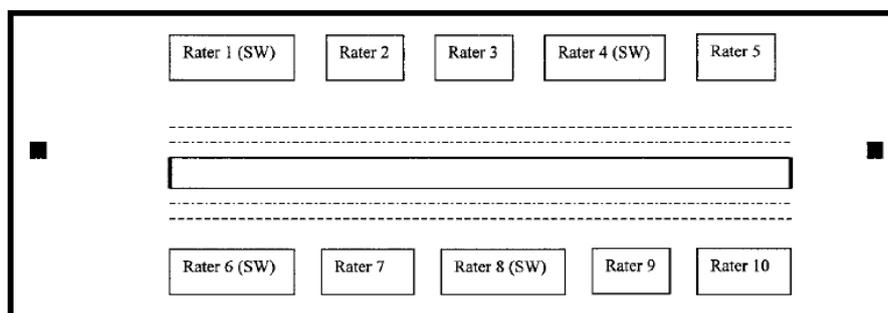


Abbildung 10 - Ausschnitt des Laufweges und der Laufwegmarkierungen sowie die Position der Rater für die Verwaltung des Tests. Die schwarzen Quadrate stellen Kegel dar, die die Start- und Abschlagspunkte anzeigen. Die Bewerter SW benutzen während des Tests eine Stoppuhr.

Die Kappa-Statistik wurde zur Bewertung der Interrater-Übereinstimmung (zwischen den Bewertern) für einzelne FGA-Items und die FGA-Gesamtwerte verwendet. Abbildung 11 enthält die prozentuale Übereinstimmung und die Kappa-Werte für jedes Item und die FGA-Gesamtwerte. Cohen schlug vor, das Kappa-Ergebnis wie folgt zu interpretieren:

Werte  $\leq 0$  bedeuten keine Übereinstimmung, 0,01-0,20 keine bis geringe Übereinstimmung, 0,21-0,40 mittlere Übereinstimmung, 0,41- 0,60 mäßige Übereinstimmung, 0,61-0,80 erhebliche Übereinstimmung und 0,81-1,00 fast perfekte Übereinstimmung (McHugh M. 2012).

	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

Abbildung 11 - Interrater-Reliabilität für einzelne und gesamte Items der funktionellen Gangbeurteilung (FGA).

Die Übereinstimmung wurde unter 10 Klinikern ermittelt, die 6 Patienten beim ersten Messversuch gemessen haben. Da es 45 mögliche Klinikerpaar-Übereinstimmungen pro Proband gibt, ergaben sich bei 6 Probanden 270 mögliche Übereinstimmungen für jedes FGA-Item und die gesamten FGA-Scores. Die Werte in Abbildung 11 stellen die mittlere Übereinstimmung und Kappa über die 45 Klinikerpaare für jedes Item dar. Die mittlere Test-Retest-Übereinstimmung über alle Klinikerpaare hinweg war unserer Meinung nach für alle Items mäßig oder besser, außer für die Items 2 ("Veränderung der Ganggeschwindigkeit") und 5 ("Gangart und Drehung").

Die Ergebnisse über die Interrater-Reliabilität der Studie von Wrisley D. et al. (2004) sind nützlich, um zu verdeutlichen, dass die Skalen und Tests, die durch Beobachtung des Bewerter entwickelt werden, im Ergebnis variieren können, je nachdem, was der Bewerter beobachtet, was als Nachteil der klinischen Skalen und Tests gegenüber den instrumentellen Instrumenten angesehen wird.

## 2.6 Sensibilität für Veränderungen und Reaktionsfähigkeit

Die *Änderungssensitivität* ist definiert als die Fähigkeit eines Instruments, Zustandsänderungen zu messen, unabhängig davon, ob die Änderung für den Entscheidungsträger relevant oder sinnvoll ist. Obwohl sie notwendig ist, wurde die Änderungssensitivität als unzureichend für die Bewertung von Veränderungen und die Feststellung der Wirksamkeit einer Behandlung beschrieben. Ein Test kann für einen Zustand oder eine Diagnose empfindlich sein, aber ob er aussagekräftig oder wichtig ist, kann nicht allein aus dieser Eigenschaft abgeleitet werden.

Die Bedeutung dieser Eigenschaft von Gangmessgeräten steht im Zusammenhang mit der Bewertung der Auswirkungen von Programmen und Behandlungen in der klinischen Wissenschaft. Wenn ein Instrument nicht empfindlich genug ist, wird es nicht in der Lage sein, die Wirkung einer bestimmten Pathologie oder Behandlung auf den menschlichen Gang zu bewerten. Eine solche Sensitivität ist besonders in angewandten Settings relevant, wo Programm- oder Behandlungseffekte oft nicht besonders stark sind und die Messbedingungen recht variabel sein können (Lipsey M. 1983). Wenn ein Instrument nicht

empfindlich genug ist, um eine Veränderung (Effektgröße) bei einer bestimmten Bedingung oder Behandlung zu erkennen, werden häufig große Stichproben benötigt, um eine ausreichende statistische Power zu erreichen, um den Effekt dieser Interventionen zu erkennen.

Andererseits ist die *Responsivität* definiert als die Fähigkeit eines Instruments, eine sinnvolle oder klinisch wichtige Veränderung eines klinischen Zustands zu messen, und wird als wesentliche Eigenschaft von Instrumenten zur Messung von Veränderungen und der Wirksamkeit von Interventionen befürwortet. Ähnlich wie die Reliabilität und Validität wird die Responsivität nicht als verallgemeinerbare Eigenschaft angesehen und sollte für jede Population und jeden Zweck, für den die Messung verwendet wird, bewertet werden.

Eine klinisch bedeutsame oder wichtige Veränderung kann aus der Perspektive des Patienten, seines Bevollmächtigten, der Gesellschaft oder des medizinischen Fachpersonals definiert und somit bewertet werden. Sie impliziert eine Veränderung, die merklich, spürbar anders ist und für den Patienten (oder Arzt) von Wert ist. Diese Veränderung kann es der Person ermöglichen, einige wesentliche Aufgaben auszuführen oder Aufgaben effizienter oder mit weniger Schmerzen oder Schwierigkeiten auszuführen. Diese Veränderungen sollten auch über die Variation hinausgehen, die dem Zufall zugeschrieben werden kann. Jaeschke *et al.* (1989) schlugen vor, dass eine klinisch bedeutsame Veränderung als der minimal wichtige Unterschied definiert werden könnte. Dies könnte definiert werden als der kleinste Unterschied in der Punktzahl in der interessierenden Domäne, den ein Patient als Veränderung wahrnimmt und der ohne Nebenwirkungen und übermäßige Kosten eine Änderung in der Behandlung des Patienten erfordern würde.

So wird die Sensitivität üblicherweise durch die Schätzung der minimal wichtigen Differenz (MID) oder der minimalen klinisch wichtigen Differenz (MCID) angegeben, wobei ein Veränderungswert für eine Maßnahme gleich oder größer als die MID-Schätzung sein sollte, um als wichtig zu gelten. Die Schätzung der MID von Messwerten verbessert die Interpretation von Veränderungswerten, indem sie Benchmarks festlegt, die helfen, die Aussagekraft von Veränderungen zu bestimmen.

### Praktischer Ansatz

Bei der Bewertung des klinischen Nutzens von Therapien, die subjektive Ergebnisse verbessern sollen, muss das Ausmaß der Verbesserung bestimmt werden, das für die Patienten wichtig ist. Die minimale klinisch bedeutsame Differenz (MCID) ist der kleinste Nutzen, der für die Patienten von Bedeutung ist. Der MCID ist ein patientenzentriertes Konzept, das sowohl das Ausmaß der Verbesserung als auch den Wert, den Patienten der Veränderung beimessen, erfasst. Mit anderen Worten: Der MCID definiert den kleinsten Betrag, um den sich ein Ergebnis ändern muss, damit es für die Patienten von Bedeutung ist (McGlothlin A. *et al* 2014).

Im Jahr 2014 erstellten Bohannon R. *et al.* eine systematische Übersichtsarbeit über den minimalen klinisch bedeutsamen Unterschied für die Veränderung der komfortablen Ganggeschwindigkeit von Erwachsenen mit Pathologie. Die Schlussfolgerung der Studie war, dass Veränderungen der Ganggeschwindigkeit von 0,10 bis 0,20 ms<sup>-1</sup> über mehrere

Patientengruppen hinweg wichtig sein können, wie z. B. bei Menschen mit Schlaganfall, Hüftfraktur, Multipler Sklerose und anderen. Wenn wir also einen Eingriff oder eine Entwicklung bei Menschen mit den in der vorherigen Studie erwähnten Pathologien messen wollen, müssen wir ein Messgerät mit ausreichender Empfindlichkeit zur Aufzeichnung von Änderungen der Ganggeschwindigkeit verwenden, wie sie in der Studie von Bohannon R. *et al.* angegeben wurde, d. h. mit der Fähigkeit, Geschwindigkeitsänderungen von  $0,1 \text{ ms}^{-1}$  zu registrieren.

Zusammenfassend lässt sich sagen, dass medizinisches Fachpersonal bei der Auswahl eines Messinstruments berücksichtigen sollte, dass die Sensitivität des Geräts ausreichend sein muss, um einen minimalen klinisch bedeutsamen Unterschied in der Variable zu messen, die sie in einer bestimmten Population beobachten möchten.

In Bezug auf die Sensitivität haben biomechanische Messverfahren eine größere Sensitivität als klinische Bewertungsskalen, um Veränderungen der Gangmerkmale zu erkennen. Dies liegt daran, dass biomechanische Messgeräte viel genauere Geräte sind. Um dieses Konzept zu verstehen, sehen Sie sich Abbildung 12 an.

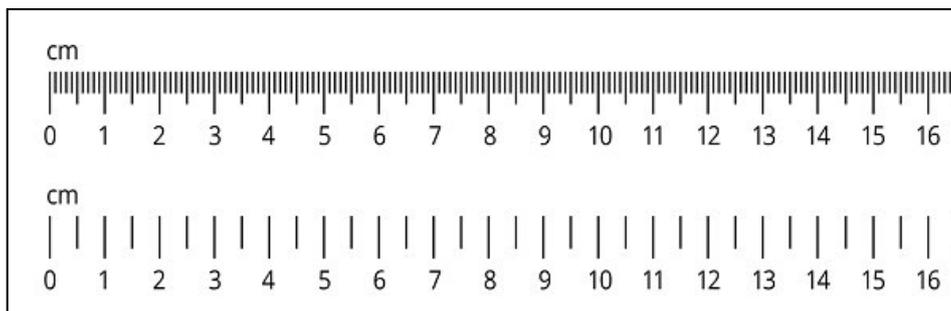


Abbildung 12 - Das Bild zeigt zwei Lineale. Das obere Lineal ist in mm und das untere Lineal in cm eingeteilt, und ihre Genauigkeiten betragen jeweils  $\pm 1 \text{ mm}$  und  $\pm 1 \text{ cm} = \pm 10 \text{ mm}$ . Da Ersteres genauer ist als Letzteres, ist das obere Lineal empfindlicher als das untere Lineal.

Das obere Lineal ist auf 1 mm genau und das untere Lineal in der Abbildung ist auf 1 cm (10 mm) genau. Da das obere Lineal präziser ist als das untere Lineal, ist es wiederum empfindlicher für die Längenmessung. Dies ist auch bei den biomechanischen Bewertungsinstrumenten zur Messung des menschlichen Gangs der Fall. In Abbildung 5 ist die Präzision der verschiedenen biomechanischen Instrumente zur Messung der raumzeitlichen Parameter während des menschlichen Gangs zu sehen. Photogrammetriesysteme und optoelektronische Kameras sind im Vergleich zu anderen Systemen die genauesten Geräte, daher sind sie auch empfindlicher bei der Messung von raumzeitlichen Variablen als Gehhilfen, die mit Drucksensoren, Inertialsensoren oder einer Stoppuhr ausgestattet sind (Moissenet F. *et al.* ). Leider sind die wissenschaftlichen Erkenntnisse über die Empfindlichkeit der Messgeräte begrenzt.

### Beispiel: Sensitivität bei klinischen Tests des Gangs bei Menschen mit Rückenmarksverletzungen

Genauso wie sich biomechanische Assessment-Instrumente in ihrer Sensitivität zur Messung bestimmter Outcomes unterscheiden, haben auch klinische Skalen diese Eigenschaft. Jackson A. et al. (2008) analysierten den Nutzen in der klinischen Praxis von aktuellen Ergebnismessungen, die als Indikatoren für die Verbesserung des Gangs und der Ambulation bei der rückenmarksverletzten Bevölkerung verwendet werden. Im Einzelnen bewerteten sie die folgenden Gang- und Ambulationsmaße für: den Walking Index for Spinal Cord Injury II (WISCI II), den 50-Fuß-Gehtest (50FTWT), den 6-Minuten-Gehtest (6MWT), den 10-Meter-Gehtest (10MWT) und den Functional Independence Measure-Locomotor (FIM-L). Die Ergebnisse sind in Tabelle 4 dargestellt.

Tabelle 4 - Empfindlichkeit gegenüber Veränderungen in der Fortbewegungsleistung von Personen mit Rückenmarksverletzungen

Klinischer Test	Empfindlichkeit informiert
Gehindex für Rückenmarksverletzungen II (WISCI II)	<ul style="list-style-type: none"> <li>- Gute Empfindlichkeit gegenüber Veränderungen bei Patienten mit stärker beeinträchtigtem Gangbild.</li> <li>- Keine Elemente der Geschwindigkeit oder Ausdauer einbeziehen.</li> </ul>
50-Fuß-Gehtest (50FTWT)	<ul style="list-style-type: none"> <li>- Probanden, die keine 50 Fuß gehen können, sind nicht in der Lage, an der Messung teilzunehmen (Bodeneffekte).</li> <li>- Diejenigen Patienten, die die Strecke in einem schnellen Tempo gehen können, sind möglicherweise auch in der Lage, eine größere Strecke mit der gleichen Geschwindigkeit zu gehen (Deckeneffekt).</li> </ul>
6-Minuten-Gehtest (6MWT)	<ul style="list-style-type: none"> <li>- Gute Sensitivität für Veränderungen bei Probanden mit besserer ambulanter Funktion bei akuter Verletzung und 6 Monate nach der Verletzung (während der WISCI II und der Lower Extremity Motor Score dies nicht taten).</li> <li>- Bei Patienten mit Schlaganfall zeigte der Test eine geringere Empfindlichkeit gegenüber Veränderungen als der 12 MWT.</li> <li>- Die Sensitivität für den 6MWT wird durch einen Bodeneffekt bei Patienten beeinträchtigt, die nicht 6 Minuten gehen können.</li> <li>- Aufrechtes Ruhen ist während des Tests erlaubt, aber wenn die Testperson sitzt, dann wird sie disqualifiziert. In ähnlicher Weise wird ein Deckeneffekt bei Patienten gesehen, die über 6 Minuten hinaus im gleichen Tempo weiterlaufen können.</li> </ul>
10-Meter-Gehtest	<ul style="list-style-type: none"> <li>- Mehr Sensitivität bei Patienten mit mehr als 90 %</li> </ul>

(10MWT)	<p>Erholung des motorischen Scores der unteren Extremitäten 6 und 12 Monate nach der Verletzung (ASIA D).</p> <ul style="list-style-type: none"> <li>- Bessere Sensitivität als der WISCI II, insbesondere bei Probanden, die weniger beeinträchtigt sind, aufgrund des Deckeneffekts dieses Tests.</li> <li>- Seine Sensitivität gegenüber Veränderungen der Gangart ist eher vergleichbar mit anderen Ganggeschwindigkeitstests wie dem 6MWT, da ähnliche Merkmale gemessen werden.</li> <li>- Die Reaktionsfähigkeit auf Veränderungen wurde auch in der Schlaganfallpopulation mit einer modifizierten 6-Meter-Distanz als Variation des 10MWT gezeigt.</li> <li>- Bessere Sensitivität als der Lower Extremity Motor Score über einen längeren Zeitraum (6 Monate).</li> <li>- Die Ergebnismessung ist weniger empfindlich für Veränderungen in der Fortbewegung bei chronischen SCI, bei denen es kaum Veränderungen in der Kraft gibt.</li> </ul>
Funktionale Unabhängigkeitsmessung - Lokomotorik (FIM-L)	<ul style="list-style-type: none"> <li>- Geringe Sensitivität, um eine Veränderung der Punktzahl eines Probanden vor und nach dem Training zu zeigen, obwohl die Art der Fortbewegung vom Rollstuhl zum Gehen wechselte.</li> <li>- Ein Deckeneffekt zeigt sich darin, dass der Test eine geringe Sensitivität für Veränderungen bei Probanden mit besseren Gehfähigkeiten aufweist. Außerdem berücksichtigt die Auswertung des FIM-L nicht die Geschwindigkeit, mit der die Testpersonen die geforderten Strecken absolvieren.</li> </ul>

Die Schlussfolgerung der Studie war, dass bei Menschen mit Rückenmarksverletzungen der 10-Meter-Gehtest und der Walking Index for Spinal Cord Injury II der valideste und klinisch nützlichste Test zur Bewertung von Verbesserungen im Gang und in der Ambulation sind, unter anderem aufgrund der guten Sensitivität zur Bewertung von Veränderungen in der Ambulation in dieser Gruppe von Patienten mit hoher Erholung bzw. mit stärker verändertem Gang.

Im vorherigen Beispiel haben wir gesehen, dass die Empfindlichkeit der Tests zur Gangbeurteilung durch zwei Phänomene beeinflusst wird, die als Deckeneffekt und Bodeneffekt bezeichnet werden und mit der Durchführbarkeit des Tests selbst durch die Zielpersonen zusammenhängen.

## 2.7 Boden- und Deckeneffekt

Wenn der Funktionsbereich, der von einer Messung abgedeckt wird, kleiner ist als der von den Patienten erlebte Bereich, kann es der Messung an Reaktionsfähigkeit fehlen. Das Potenzial für Boden- und Deckeneffekte wird oft durch die Untersuchung von

Antwortmustern beurteilt. Wenn es Spitzen bei der höchsten oder niedrigsten Antwortoption gibt, wird dies oft als Hinweis auf Decken- bzw. Bodeneffekte interpretiert. Bei der Verwendung von Messwerten zur Bewertung der Wirksamkeit von Interventionen ist der prospektive Nachweis der Leistungsfähigkeit eines Messwerts jedoch wichtiger als die Frage, ob es Spitzen gibt oder nicht (Feeny DH. *et al.* 2013).

Wenn sich die Werte der Teilnehmer am oberen Ende (oder bestmöglichen Wert) der Messung/des Instruments (Deckeneffekt) oder am unteren Ende (Bodeneffekt) häufen, können die von den Patienten erlebten Veränderungen der Gangleistung verzerrt sein. Dies ist darauf zurückzuführen, dass die Patienten "schlechter dran" sein könnten, als die Messung erfassen kann, oder "besser dran", als das Instrument messen kann.

Wenn wir die instrumentellen Bewertungstechniken mit den klinischen Bewertungsskalen vergleichen, werden wir feststellen, dass die instrumentellen Techniken einen Bodeneffekt haben, während die klinischen Skalen zur Gangbeurteilung einen Deckeneffekt haben. Der Grund für diesen Unterschied ist die Leichtigkeit oder Schwierigkeit, mit der jede von ihnen vom Patienten durchgeführt werden kann. Die Bewertungsskalen werden anhand von strukturierten Fragebögen definiert, wobei kategorische oder diskrete Klassifizierungen vorgenommen werden, die die Sensitivität in Bezug auf andere klinische Bewertungsinstrumente oder -techniken verringern. Mit anderen Worten, kleine Veränderungen, die in der funktionellen Kapazität des Patienten als Folge der Intervention des Fachmanns auftreten, sind sehr schwer zu identifizieren. Diese Tatsache impliziert, abhängig von der Konstruktion der Skala, einen Deckeneffekt. Menschen mit schweren Gangstörungen werden jedoch nicht in der Lage sein, die komplexen Bewertungsprotokolle durchzuführen, die von vielen der biomechanischen Instrumentaltechniken verlangt werden, so dass die empfindlichste Bewertung für diese Art von Patienten die klinischen Bewertungsskalen sein werden.

Auf der anderen Seite haben die instrumentellen Techniken der biomechanischen Beurteilung einen Bodeneffekt, weil die Beurteilungsprotokolle eine größere Anforderung an den Patienten stellen, um durchgeführt zu werden. Das heißt, in den meisten Fällen ist die Beurteilung des Gangs von Menschen mit schweren Beeinträchtigungen beim Gehen mit instrumentellen Techniken nicht möglich, so dass der Bodeneffekt den Eintritt von schwer verletzten Personen begrenzen kann, bis sie mit den erforderlichen Instrumenten gehen können.

### Beispiel

Obwohl die Studien, die die statistischen Eigenschaften von Bewertungsinstrumenten analysieren, begrenzt sind, haben Middleton A. und Fritz S. (2013), berichteten Evidenz über die genaue Bewertung des Gangs bei älteren Erwachsenen, da dies ein wichtiger Aspekt der klinischen Praxis für Kliniker ist, die mit dieser Population arbeiten. Ausgewählte Maße sind wie folgt: Gang: Ganggeschwindigkeit, Gangsymmetrie, Gangausdauer, Anpassungsfähigkeit des Gangs, Dual-Task-Leistung während des Gangs und selbstberichtete Sicherheit während des Gangs. Abbildung 13 zeigt eine Zusammenfassung der in dem Artikel enthaltenen Ergebnismessungen als Schnellreferenz für Kliniker.

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

Abbildung 13 - Zusammenfassende Tabelle der Gang-Outcomes-Maßnahmen. Minimal nachweisbare Veränderung (MDC), nicht anwendbar (n/a).

Die minimal nachweisbare Veränderung (MDC) bietet Klinikern einen Referenzwert, um festzustellen, ob eine echte Veränderung stattgefunden hat. Der MDC95 quantifiziert den kleinsten Betrag an Veränderung, der erforderlich ist, um Messfehler und Variabilität mit einem Vertrauensniveau von 95 % zu übertreffen. Wenn in der Literatur berichtet wird, können MDC-Scores für Ergebnismessungen wertvolle Werkzeuge für Kliniker sein. In Abbildung 13 ist die kleinste nachweisbare Veränderung für die in der klinischen Praxis am häufigsten verwendeten Gang-Outcomes, gemessen mit den in der ersten Spalte genannten Instrumenten, zu sehen.

Diese Studie ist nützlich, um zu erklären, dass Patienten, die in jeder der genannten Variablen unter der minimalen nachweisbaren Veränderung liegen, es vorziehen, dass sie mit anderen Instrumenten gemessen werden, die es ermöglichen, Veränderungen unterhalb des referenzierten Wertes als minimale nachweisbare Veränderung zu erkennen.

### 3. Wichtige Ideen

---

- Das medizinische Personal muss die methodischen Merkmale und statistischen Eigenschaften zum Zeitpunkt der Auswahl eines Instruments zur Gangbeurteilung kennen. Dies ist notwendig, um methodische Fehler und Verzerrungen in den Messergebnissen zu vermeiden.
- Hinsichtlich der Anwendbarkeit haben klinische Skalen und Tests den Vorteil, dass sie in kurzer Zeit entwickelt werden können, keine spezielle Ausbildung des Bewerter erfordern und in jedem Kontext, wie z.B. in der klinischen Praxis, eingesetzt werden können.
- Die Ausrüstung, die für die Verwendung von klinischen Tests und Skalen erforderlich ist, ist viel geringer und zugänglicher als die Ausrüstung, die für die Durchführung einer Gangbeurteilung mit biomechanischen Beurteilungsinstrumenten erforderlich ist.
- Die wichtigste Eigenschaft der instrumentellen biomechanischen Bewertungstechniken ist, dass sie objektive Daten liefern, die ohne Interpretation des Bewerter gewonnen werden (d. h. direkte Bewertung einer oder mehrerer Dimensionen des Gangbildes), so dass ihre Verwendung hauptsächlich im Forschungsbereich liegt. Im Gegensatz dazu werden die durch Skalen und klinische Tests gewonnenen Informationen durch die Interpretation und Wahrnehmung des Bewerter beeinflusst.
- Die hohe Präzision der instrumentellen Messtechniken verleiht ihnen die Eigenschaft, ein Gangmerkmal valider zu messen als die Skalen oder klinischen Tests.
- Die Zuverlässigkeit ist bei biomechanischen Geräten in der Regel besser, da die Wiederholbarkeit der Messung nicht vom Beobachter, sondern von anderen Faktoren abhängt, wie z. B. der Durchführung der Messung mit einem standardisierten Protokoll.
- Je genauer ein Messgerät ist, desto empfindlicher reagiert es auf Veränderungen. Die Empfindlichkeit der Geräte muss ausreichen, um den minimalen klinisch bedeutsamen Unterschied in den Ergebnissen zu messen, die die Fachleute in einer bestimmten Population zu beobachten beabsichtigen.
- Die klinischen Skalen und Tests haben eine größere Tendenz zu einem Deckeneffekt, d. h. die Werte der Teilnehmer häufen sich in Richtung des oberen Endes (oder der bestmöglichen Punktzahl) der Messung/des Instruments. Auf der anderen Seite haben die instrumentellen Verfahren einen größeren Bodeneffekt, bei dem sich die Werte der Teilnehmer zum unteren Ende hin anhäufen. Dies liegt daran, dass die Patienten "besser dran" sein könnten, als die Messung erfassen kann, oder "schlechter dran", als das Instrument messen kann.

## 4. Referenzen

---

- [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.



Die Unterstützung der Europäischen Kommission für die Erstellung dieser Veröffentlichung stellt keine Billigung des Inhalts dar, welcher nur die Ansichten der Verfasser wiedergibt, und die Kommission kann nicht für eine etwaige Verwendung der darin enthaltenen Informationen haftbar gemacht werden.