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

Module Biomechanics of Gait

Didactic Unit C: How do I assess gait?

C.3: What are the advantages of the use of instrumental techniques versus scales and physical examination to assess gait?

Index

1. Introduction and Objectives

Both the evaluation of gait through 1) observation, 2) clinical scales, tests and questionnaire, and 3) assessment using instrumental techniques are useful to know the gait pattern of a person, but there are fundamental differences that must be taken into account when choose one or the other type of valuation technique or a combination of both.

In this unit we will review the methodological characteristics and statistical properties of the techniques available for the assessment of human gait reviewed in Unit C.1 of the Module Biomechanics of Gait.

Based on this, we propose the following objectives:

1. To review the advantages and disadvantages of valuation methodologies for human gait.

2. To know the statistical properties of the gait assessment methodologies available.

3. To establish the technical knowledge that allow healthcare professionals to choose the most appropriate gait assessment technique for their clinical or research context.

2. Features and properties of gait assessment tools: comparison between available techniques

2.1 Usability

Usability refers to the ease with which people can use a particular tool in order to achieve a specific goal. Faced with the practice of the profession of any health profile and in any area, before choosing an evaluation technique (or treatment), professionals will ask questions such as: *is it easy to use? Does it take a long time? Is it feasible to use it in my work area?* In this sense, the available human gait assessment techniques (reviewed in previous units) have different answers to the questions posed above.

Regarding ease of use, instrumental techniques require several steps framed in a strict assessment protocol that must be followed to ensure that the comparison between measurements from the patient's own data (for example, when several assessments are carried out over a period of time) or with other subjects' data are valid. This makes any instrumental technique more complex to use than a scale, questionnaire, or clinical test (Table 1). Of the measurement protocol with an instrumental technique, probably the most complex step to manage is the instrumentation of the subject if needed and the post-treatment of data after performing the measurement. Therefore, taking into account the above, the instrumental gait assessment techniques require a very relevant factor to take into account: ***time*** (Table 1). However, with the advancement of available technologies, within a group of instrumental techniques to measure the same parameters, we can find different scenarios.

Example

Biomechanical instrumental assessment of the spatiotemporal parameters of gait

Using a photogrammetry system and an instrumented walkway, it is possible to measure the spatiotemporal outcomes that characterize human gait. These parameters are primarily gait velocity, stride length and step length, cadence, double support time, and support and swing phase time of the gait cycle. Although with both techniques we can reach the same result, the time it takes for each of the techniques varies significantly.

With the instrumented walkway (with pressure sensors) it is not necessary to instrument the subject since the valued person should only walk through the corridor with or without footwear. If the equipment management software calculates the parameters from the recorded plantar pressures, no further treatment will be required to measure the data. Thus, the evaluator does not need excessive training or knowledge to use the tool.

On the other hand, with a photogrammetry system, the instrumentation of the subject is always necessary. This instrumentation consists of a biomechanical model of landmarks that are placed in the body to represent the segments involved in the movement to be evaluated. Depending on the complexity of the model, the instrumentation time will be longer or shorter. In any case, the skin where the marker is attached must always be cleaned and shaved. On the other hand, the management software of the photogrammetry systems usually gives as results the coordinates of each marker in each axis of movement and at each image capture time (for example, 50 positions per second if the recording frequency is 50 Hz). Unless a software offers a protocol with a closed and fixed model and previously calculates the different parameters that can be obtained from the measurement, a calculation of outcomes will always have to be made after extracting the coordinates of each landmark from the biomechanical model. All this requires implies that the evaluator to have a deeper knowledge and to practice the correct placement of the markers.

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

On the other hand, the observation gait analysis or the assessment by clinical scales requires a minimum amount of time and the necessary training for the evaluation of gait is related to the practice of the instructions of the scale itself or test. It is for all of the above, that the use of each measurement technique has more usability in different areas. The use of gait evaluation with scales or clinical tests or eye-naked gait analysis is easier to use in the context of daily clinical consultation where patients present with different pathologies that lead to gait disturbances. On the other hand, instrumental techniques are usually used in research projects. Even so, in the context of research, scales and validated clinical tests in different populations and languages are widely used.

Table 1 - Usability of gait assessment tools

Characteristic	**Observation gait analysis**	**Questionnaire, Scales and clinical Test**	**Instrumental techniques**
Time cost	+	+	+ / ++/ +++ (depending on the system used)
Evaluator training	+	+	++ / +++ (depending on the system used)
Context of use	Clinical	Clinical and research	Research
Usability	+	++	+++

2.2 Equipment requirements

While gait evaluation with a scale or with a questionnaire requires nothing more than the corresponding form and a pen, clinical tests may require a certain amount of materials to assemble the corresponding set up. Some tests may require a chair, ground marking tape, and a stopwatch, while others may require a greater amount of materials (see the example below). In any case, these materials are usually inexpensive (Table 2) or accessible to any healthcare professional.

Example

How many materials do I need to do the 6-minutes walking test?

The 6-minutes walking test (6MWT) has also been used as a one-time measure of functional status of patients, as well as a predictor of morbidity and mortality. The 6MWT is a practical simple test that requires a 100-ft hallway but no exercise equipment or advanced training for technicians. This test measures the distance that a patient can quickly walk on a flat, hard surface in a period of 6 minutes (the 6MWD). It evaluates the global and integrated responses of all the systems involved during exercise, including the pulmonary and cardiovascular systems, systemic circulation, peripheral circulation, blood, neuromuscular units, and muscle metabolism. The self-paced 6MWT assesses the submaximal level of functional capacity.

The required equipment is:

1. Countdown timer (or stopwatch)
2. Mechanical lap counter
3. Two small cones to mark the turnaround points
4. A chair that can be easily moved along the walking course
5. Worksheets on a clipboard (registration form)
6. Adhesive tape or colored stickers to mark the patient's place of detention at 6 minutes
7. Borg scale
8. Pulse oximeter
9. Sphygmomanometer and stethoscope
10. Telephone
11. A source of oxygen
12. Automated electronic defibrillator

6-minutes walking test set up

However, when it comes to appraisal with instrumental techniques, the material required is not only more quantity, it is highly specialized. In general, measurement equipment to register the gait pattern or a specific characteristic of the gait consists of the following parts:

- ***Sensors or/and measurement equipment*** refers to the devices that compose the measurement equipment itself. For example:

- A dynamometric platform is made up of a rigid, flat and static surface installed on the ground and the sensors or transducers that are placed at the bottom of the platform, distributed in such a way that they are capable of recording the forces exerted on the upper rigid surface.
- Electromyography equipment is made up of electrodes (sensors) that collect electrical activity within the muscle, either by inserting it or through the skin that covers the muscle. The system also includes the amplifier that allows the electrical signals from the muscle to be transferred to a monitor, usually wirelessly, and the wiring between the sensors and the amplifier.

- ***Software and computer***: it refers to the software that allows the management of the registration equipment and sensors, the storage of the registered data, and the computer where the software is used.
- ***Supplies***: refers to all kinds of accessories required to carry out a gait measurement. For example:
  - In an electromyography system, the necessary accessories are those that will serve to prepare the skin where the electrode will be placed and thus reduce the impedance between the electrode and the skin: shaver, alcohol, cotton, and fine sandpaper to remove dead skin. In addition, you will need a conductive gel that will also help reduce the skin-electrode impedance.

Example

How many materials do I need to do a gait evaluation with a photogrammetry system

Figure 1 - Photogrammetry system and its components

Besides the cameras, the software, and the computer, in the photogrammetry system, it will be essential to have a set of landmarks to represent the biomechanical model. In addition, for these to be well adhered to the skin, a system of adhesives will be needed, elements to clean the skin, and prevent them from falling during the test (shaver, alcohol, cotton) and clothing necessary to perform the test.

As can be seen, the materials and accessories of instrumental techniques for gait evaluation are not usually found in medical services. The cost of the equipment and the software can vary enormously depending on the assessment instrument and the possibilities of the software (Table 2). For example, some software may offer normative data with which to compare the assessment made on a patient of certain sex and age, which increases its price. Equipment without any instrumentation such as an walkway with pressure sensors and its respective software can cost around 40 thousand euros depending on the length of the corridor, while photogrammetry equipment with cameras and software can easily exceed 100 thousand euros.

Table 2 - Requirements of gait assessment tools

Characteristic	Observation gait analysis	Questionnaire, Scales and clinical Test	Instrumental techniques
Equipment	+	+	+++
Supplies	-	+	++
Economic cost	+	+	+++

2.3 Objectivity of the results and statistical analysis

The objective data of a gait assessment refers to data obtained without interpretation of the evaluator (*i.e.* directly assessment of one or more dimensions of gait pattern), while in subjective data, the result is subject to the interpretation, perception, or opinion of the evaluator. In the tools available for gait assessment we find both types of results. While with instrumental assessment tools we can obtain objective measurement data, gait analysis through observation, questionnaires, or clinical scales will provide subjective information. It is for this reason that it is convenient for the health professional to know before its application that the evaluation scales are very useful instruments to qualitatively determine general states of health; however, they are somewhat subject to the subjectivity of both the patient (self-reported assessment) and the evaluator (observation-scale assessment). Furthermore, they are sometimes developed in very specific contexts (populations, regions, countries, etc.), with very important demographic and cultural biases, which make extrapolation to other communities difficult or impossible. On the other hand, the major strength of self-reported assessments is that they give an account of what the person really experiences and perceives. Moreover, the method is cheap, fast, and feasible.

Example

Subjective and objective measure of step length

Many gait assessment scales have some items related to the length of the step. One of them is the Tinetti Mobility Test (TMT), wherein the gait section, it is asked if the patient is capable of passing one foot over the other when taking the step. When we measure the length step in a person without pathology, the response may be clearly visible and the observer will undoubtedly point out that the right foot surpasses the left (or vice-versa). However, if a patient with Parkinson's is evaluated, in which one of the characteristics of the disease is walking with small steps, it may be the case that the response to the TMT item is not as clear and two researchers observe different things. On the other hand, a measuring instrument such as the instrumented walkway is capable of measuring the step length by detecting plantar pressure when the patient places the foot on the ground. With this type of instrument, the distance obtained from the step length is not influenced by the evaluator's observation.

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

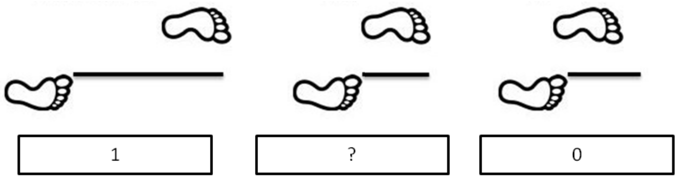

Figure 2 - Step length and height ítem from the Tinetti Mobility Test, Gait section. Three examples of step length are shown in which the assessor must mark 1 or 0 according to the criteria established by the scale. In the example of the center, it can give rise to doubts of which numbering to assign.

Figure 3 - Step length assessment with an instrumented walkway (GAITrite)

The main advantage of using objective results measured with instrumental techniques is that these allow, among other things, that the data be comparable with other data from the same patient or with results from other subjects. However, the researcher or the professional assessor should not forget that the comparison of objective data between subjects should be normalized by the most influential characteristic on this data. For example, the length of the step is also influenced by the height of the subject and therefore the length of the lower limbs, therefore the objective data of step length should be divided by the height or length of the legs. The same will happen with variables such as the reaction forces and the weight of the subjects.

Although the scales or questionnaires and the instrumental tools do not provide the same type of information (subjective versus objective), subjective measures can be highly correlated with objective measures. This is another added value to the assessment scales used in the clinical setting. A scale or questionnaire that is highly correlated with the results of the assessment using an instrumental technique, will be a valid tool to measure populations in large samples.

Example

Journal List > Age Ageing > PMC4476851

OXFORD JOURNALS Age and Ageing

Age Ageing. 2015 Jul; 44(4): 691–694.
Published online 2015 May 27. doi: 10.1093/ageing/afv062

PMCID: PMC4476851
PMID: 26018999

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

Paul Innerd,[1] Michael Catt,[1] Joanna Collerton,[2] Karen Davies,[2] Michael Trenell,[3] Thomas B. L. Kirkwood,[4] and Carol Jagger[4]

In epidemiological research, Physical Activity is commonly assessed using physical activity questionnaires due to their practicality and low cost. It would be extremely expensive and complex to assess large samples of participants with some instrumental technique. That is why there are studies that look to determine the correlation of self-reported physical activity questionnaires with objective measures such as physical activity recorded with accelerometers used in a small sample of the target population for one week. Thus, the objective of this type of study is to evaluate the validity of the subjective test as a tool for evaluating physical activity for large samples of participants.

Now, are the subjective data susceptible of being statistically analyzed? The answer is yes, both the subjective information obtained through scales or questionnaires and the objective information obtained from an instrumental technique are capable of being statistically analyzed. Clearly, walking speed, for example, is considered a quantitative variable, i.e. It is a variable that has numerical quantities as modalities with which we can do arithmetic operations (Table 3).

Table 3 - Objectivity of the results and statistical analysis

Characteristic	Observation gait analysis	Questionnaire, Scales and clinical Test	Instrumental techniques
Objectivity	Subjective information	Subjective information	Objective information
Statistical analysis	No (a priori)	Yes	Yes
Type of variable in statistical analysis	-	Qualitative, semi-quantitative and quantitative (depending on the answer to analyze)	Quantitative

On the other hand, from a questionnaire or a scale, we can analyze its final score. If this final score is a number, it will be statistically analyzed as a quantitative variable, because it is what we will introduce in the statistical program (a number), but is still being subjective information. However, if what we submit to analysis is each question on the scale or questionnaire, we can obtain different statistical variables (Table 3), depending on the type of answers we can find:

1. The Dynamic Parkinson Gait Scale has questions where the assessor must assign a score between 0 and 5 on the performance of the patient walking. For example, in item 1 relating to how the patient walks seven meters forward, the evaluator can rate this characteristic from normal (0 points) to "unable to initiate a step forward or fail" (5 points). Although it is still categorical modalities within the question, these possible answer have an order: one is worse than zero, two is worse than one, three is worse than two, and so on. This is a 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

2. The Tinetti Mobility Test has questions where the observer must verify a characteristic of the gait in the patient and assign 0 point if what they observe is a negative or altered response or 1 point if the gait characteristic evaluated is presented as positive or normal. In other words, this type of question or scale item is a yes / no answer. When analyzed statistically, we would consider it as a qualitative, categorical variable with two modalities: 0 = characteristic of altered gait and 1 = characteristic of normal gait.

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

Finally, when the observer performs an analysis of the patient's gait through the naked-eye and obtains a set of ideas or characteristics about the gait pattern that she/he observes, this information is subjective but since it is not standardized, it could not be analyze using conventional statistical methods (Table 3).

2.4 Validity

Before being considered suitable, the instruments must offer accurate, valid and interpretable data for the population's health assessment. Moreover, the measures are supposed to provide scientifically robust results. The performance of results of these measures comes from the reliability and validity of instruments.

Validity refers to the fact that a tool measures exactly what it proposes to measure or, in other words, validity refers to the accuracy of the measurement (Figure 4). Validity is not an instrument characteristic and must be determined regarding a specific matter, once it refers to a defined population. The measurement properties – validity and reliability – are not totally independent. An instrument that is not reliable cannot be valid; however, a reliable instrument, can, sometimes, be invalid. Thus, a high reliability does not ensure an instrument validity.

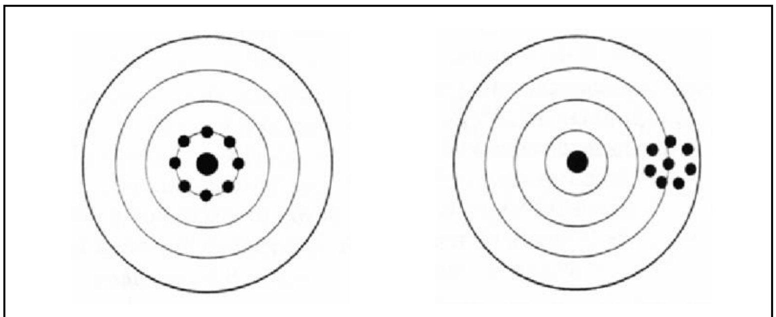

Figure 4 - In the target of the left, the shots hit the place they were supposed to and were consistent, right in the target center. In the target of the right, the shots were reliable, hitting the same point; however, none has hit the center of the target, not being considered valid (Image from De Souza A. et al. 2017)

Validity of assessment instruments requires several sources of evidence to build the case that the instrument measures what it is supposed to measure. Determining validity can be viewed as constructing an evidence-based argument regarding how well a tool measures what it is supposed to do. Evidence can be assembled to support, or not support, a specific use of the assessment tool. That is why in the bibliography we find several studies that seek to study the validity of a tool to measure gait in a certain target population (Gail M. Sullivan 2011).

What is the procedure to measure the validity of a tool?

The technique or tool whose validity is intended to be studied must be compared with a "*gold standard*" tool whose validity has already been extensively tested to measure a specific outcome (concurrent validity). In this type of study, the aim is to answer the question: does tool *A* measure as precisely as tool *B* does human gait? This is usually analyzed using the Pearson or Spearman Correlation Coefficient (r). It is considered excellent when the Correlation Coefficient between the measurement instrument and the reference standard is above 0.6, adequate between 0.59 and 0.31 and poor below 0.3.

Example 1: validity of scale

Diane M. Wrisley and Neeraj A. Kumar (2010) determined the validity of the Functional Gait Assessment (FGA) for classifying fall risk and predicting unexplained falls in community-dwelling older adults. For that purpose, thirty-five older adults aged 60 to 90 years completed the Activities specific Balance Confidence Scale (ABC), the Berg Balance Scale (BBS), the Dynamic Gait Index (DGI), the Timed "Up & Go" Test (TUG), in addition to the tool to be validated (FGA). Spearman correlation coefficients were used to determine concurrent validity among the ABC, BBS, TUG, DGI, and FGA. The FGA correlated with the ABC ($r = .053$, $P<.001$), BBS ($r = .84$, $P<.001$), and TUG ($r = .84$, $P<.001$). The authors determined that the FGA with a cutoff score of 22/30 is effective in classifying fall risk in older adults and predicting unexplained falls in community-dwelling older adults.

Example 2: validity of three-dimensional gait analysis on treadmill

Pinto R**.** et al. (2020) determined the validity of three-dimensional (3D) gait biomechanics derived from treadmill-based systems. For this, the researchers examined concurrent validity by estimating the associations between treadmill-based and overground (gold standard) measures in participants with knee osteoarthritis. Treadmill walking speed was matched to self-selected overground speed. Marker set, knee angle and moment calculations were consistent for both systems. Variables calculated from knee angle and moment gait waveforms during stance were evaluated using Pearson correlations (r) among other statistics parameters. The results showed that Pearson correlations between treadmill and overground systems ranged from 0.56-to-0.97. Although highly associated, there were substantial differences in the moments, emphasizing they cannot be used interchangeably. This suggest that frontal and sagittal plane knee angles and moments in patients with knee osteoarthritis evaluated using a treadmill-based system are valid.

Now that we have defined validity and how it works, it is worth asking: *what types of tools have the most validity to measure gait or a specific characteristic of gait? Scales / clinical tests or the instrumental techniques?* The answer is clear, instrumental measurement techniques are more valid for measuring gait than scales or clinical tests because they are more precise instruments to measure a certain variable of the gait. It is worth mentioning that even within the instrumental techniques for gait analysis, there are some more precise than others. These precision differences are related by the measurement error of the equipment implicit in the methodology itself. Figure 5 shows a comparison of the instrumental techniques available to measure the spatiotemporal parameters of gait such as speed, step length, cadence, duration of the gait cycle and gait phases.

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

Figure 5 - Comparison of the common technologies used to measured spatiotemporal gait parameters (Moissenet F. and Armand S. 2016). For each instrumental technique, the degree of precision and the cost of the technique are mentioned.

2.5 Reliability

On the other hand, reliability is the ability to reproduce a consistent result in time and space, or from different observers (Figure 6), presenting aspects on coherence, stability, equivalence and homogeneity. It is one of the quality criteria of an instrument (de Souza et al. 2017). Reliability is also concerned with repeatability. For example, a scale or test is said to be reliable if repeat measurement made by it under constant conditions will give the same result (Taherdoost H. 2016).

It is important to highlight that the reliability is not a fixed property of a questionnaire. On the contrary, reliability relies on the function of the instrument, of the population in which it is used, on the circumstances, on the context; that is, the same instrument may not be considered reliable under different conditions. Reliability estimates are affected by several aspects of the assessment environment (raters, sample characteristics, type of instrument, administration method) and by the statistical method used. Therefore, the results of a research using measurement instruments can only be interpreted when the assessment conditions and the statistical approach are clearly presented. In conclusion, reliability refers to whether an assessment instrument gives the same results each time it is used in the same setting with the same type of subjects (Sullivan G.M. 2011).

The choice of the statistical tests used to assess reliability may vary, depending on what in intended to be measured. There are three important reliability criteria of great interest for researchers: 1) stability, 2) internal consistency and 2) equivalence.

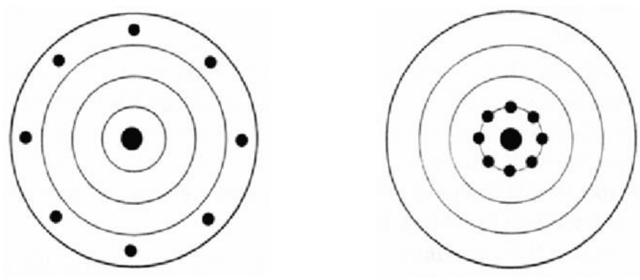

Figure 6 - In the target of the left, the shots hit not reliable, because the points hit are not located in a specific place, but were spread throughout the whole target. In the target of the right, the shots were reliable, hitting the same point (Image from De Souza A. et al. 2017).

What is the procedure to measure the reliability of an instrument?

As mentioned above, the statistical procedure for measuring reliability depends on what is intended to measure.

1) *Stability*: stability measures how similar the results are when measured at two different times, that is, it estimates the consistency of measurement repetition. Stability assessment can be performed using *test-retest* method, i.e. the procedure consists of applying the same measurement at two different times. The intra-class correlation coefficient (ICC) is one of the most used tests to estimate continuous variables stability, because it takes into account the measurement errors. For the results interpretation, minimum values of 0.70 are considered satisfactory (de Souza et al. 2017).

2) *Internal consistency*: the internal consistency (or homogeneity) shows if all subparts of an instrument measure the same characteristic. This is an important measure property for instruments that assess a single construct that use a variety of items. A low internal consistency may indicate that the items measure different constructs (de Souza et al. 2017). In the internal consistency analysis, the difference among the answers of the items from a construct is calculated. Cronbach alpha is the test frequently used to calculate the correlation values among the answers of an assessment tool. A high reliability estimate should be as close to 1 as possible (Sullivan G.M. 2011).

3) *Equivalence*: equivalence is the concordance degree of two or more observers regarding an instrument score. The most common way of assessing the equivalence is the inter-observer reliability, which involves the independent participation of two or more raters. Kappa coefficient is a measure used to assess inter-observers, applied to category variables and has a maximum value of 1. The higher the Kappa value is, the higher the concordance between the raters will be (Sullivan G.M. 2011).

Example 1: Stability reliability

Instrumental technique

Inter-rater test-retest (differents assessors)

Meng L. *et al.* (2020) investigated whether less reliance on manually identifying anatomical landmarks could improve inter-assessor reliability of joint kinematics compared to three kinematics gait models. This aim of study was raised because a major source of error in reliability of gait analysis arises from the palpation of anatomical landmarks. The hypothesis of the study was that Strathclyde functional cluster model (SFCM) would obtain greater inter-assessor reliability than the anatomical models: Plug-in gait (PIG) and the Human body model gait (HBM2). To demostrate this, 10 participants completed three trial conducted by different assessor on nonconsecutive days. In each session, the assessor applied the combined marker set on the participants. Then, a static trial was recorded for 5 s with the participant standing in a natural posture. After this, each participant walked on the treadmill at his/her comfortable speed. Thirty seconds of data were captured after two-minute familiarisation period and marked trajectories were captured by 12 cameras at a sampling rate of 100 Hz. The intraclass correlation coefficient across gait cycles were used to compare the inter-assesspr reliability and the value of ICC (α) was interpreted into four leveles of reability: $\alpha < 0.5$ = poor, $0.5 < \alpha < 0.75$ = moderate, $0.75 < \alpha < 0.9$ = good, and $\alpha > 0.9$ = excellent. Results demonstrate that all models showed a „good to excellent" inter-assessor reliability for all flexion/extension angles and hip ab/adduction angle but performed „poor to moderate" inter-assesor reliability for other non-sagittal angles. However, the SFCM obtained higher reliability with less variation compared with the anatomical models. The results demonstrate that the SFCM may be more beneficial for less experienced assessors.

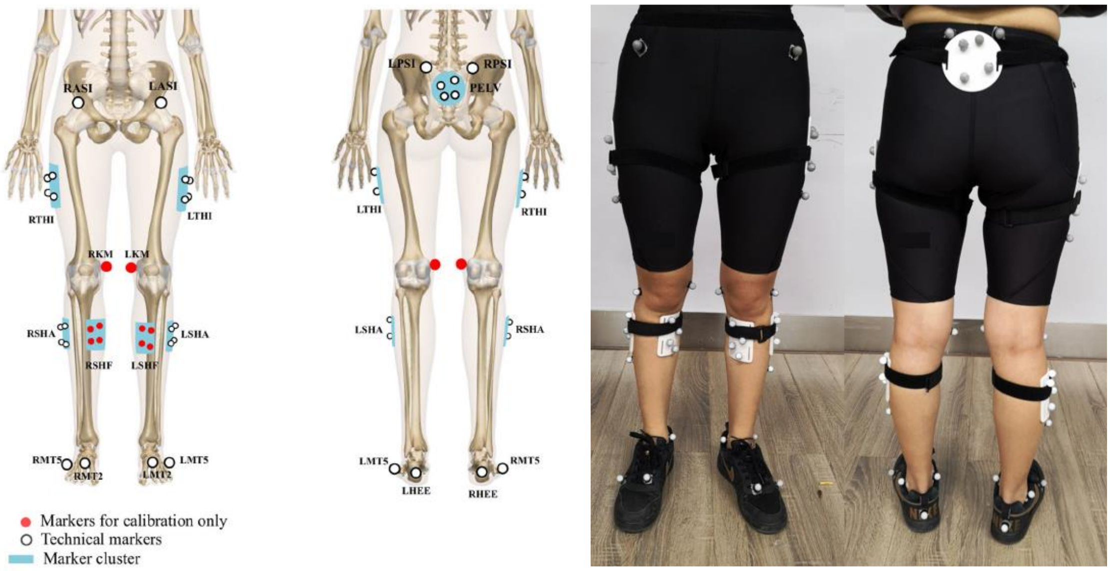

Figure 7 – Marker layout os strathclyde functional cluster model (left) and a combined marker set-up for the SFCM, PiG and HBM2 model (rigth).

Intra-rater test-retest (differents assessment times)

Geerse D. et al. (2020) determinated the test-retest reliability of the Microsoft's HoloLens, a mixed-realitt headset that provides, besides holograms, rich position data of the head, which can be used to quantify different tasks like walking. To study what are the limits of agreement of derived spatiotemporal gait parameters over repetition, 23 healthy young adults that walking at slow, comfortable, and fast speeds was recluted, as well as 24 people with Parkinson's disease that walking at self-selected speed. The HoloLens provides position and orientation data of the head in the environment in three directions, and the position data were used to derive spatiotemporal gait parameters. Test-retest reliability was evaluated with the intraclass correlation coefficient for absolute agreement. ICCs above 0.60 and 0.75 represented good and excellent agreement, respectivaly. The ICCs indicated and excellent test-retest reliability for both systems for healthy yount adults and people with Parkinson's disease.

			Trial 1	Trial 2	
			Mean ± SD	Mean ± SD	$ICC_{(A,1)}$
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
		FWS	176.3 ± 21.8	174.7 ± 22.8	0.930
	PD	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
		FWS	86.2 ± 8.0	85.5 ± 8.3	0.928
	PD	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
		FWS	124.1 ± 10.4	123.7 ± 9.8	0.953
	PD	CWS	109.6 ± 7.6	110.5 ± 7.0	0.778

Figure 8 - Test-retest reliability for spatiotemporal gait parameters of instructed slow walking speed (SWS), comfortable walking speed (CWS) and fast walking speed (FWS) conditions in healthy young adults (HYA) and CWS in people with Parkinson's disease (PD) derived from HoloLens data (Geerse D. et al. 2020).

As expected, in the reliability analyzes of an instrumental technique, the ICC values are high, since the measurement techniques are usually of high precision. Factors such as following a strict assessment protocol help maintain high test-retest reliability. If the titrator uses an instrumental measurement technique without following the same instrumentation protocol in a standardized way for the different assessment times, it is likely that the ICC values were not those observed in the example study. It can also be observed that the reliability of the measurement technique decreases in the cadence of the participants with parkinson's disease. This occurs because this group of patients changes the cadence depending on the length of the stride they perform, which is altered by the disease. This example is useful to point out that the reliability of the measurement, even though it is an instrumental technique, varies in different populations.

Clinical test

Reliability analysis are even more valued by clinical staff in commonly used clinical tests. Hee-jae Kim et al. (2016) examined the reliability (Intra-rater test-retest) of gait speed measured at various distances and paces in elderly Koreans. Fifty-four female participants ≥70 years of age were recruited from a local retirement community. Gait speed was assessed at 4, 6 and 10 meters, and at usual- and fast-pace walking mode. Participants were instructed to walk from a standing start at a pace that was normal and comfortable for them or to walk as fast as they could until they reached the end of the marked path. A trained tester walked behind the participant and stopped timing when the participant's foot contacted the floor at the end of the walking course. This mean that gait velocity was calculated with the distance traveled and the time it took the participants to travel it. The results is showed in Figure 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*

Figure 9 - Reliability of walking test (ICC) for different distance and pace.

Higher ICC values were observed at the longest walking distance of 10 meters compared to 4 and 6 meters. In addition, ICC values of gait test at maximal speed were higher than that at the normal pace. The autors mentioned that, although the walking test at a maximal pace over a longer distance has better reliability in elderly individuals, test distance and pace have to be considered according to the purpose of the measurements and the clinical health conditions of participants, rather than by the criterion of a high level of reliability.

Example 2: Internal consistency reliability

Clinical test

Unlike instrumental tools, with which we measure a specific variable such as gait speed, stride length or joint movement angle, clinical tests allow us to assess gait from a more global performance. This is the case of The Functional Gait Assessment (FGA), a measure of walking balance ability, was developed to eliminate the ceiling effect observed in the

Dynamic Gait Index (DGI). Three presumably more difficult tasks were added and 1 easier task was removed from the original 8 DGI tasks (Beninato M. and Ludlow L., 2016). The items on the FGA are: 1) Gait on level surface, 2) Change in gait speed, 3) Gait with horizontaal head turns, 4) Gait with vertical head turns, 5) Gait and pivot turn, 6) Step over obstacle, 7) Gait with narrow base of support, 8) Gait with eyes closed, 9) Ambulating backwards, and 10) Steps. In each item, the evaluator can give a score from 0 to 3, obtaining a maximum score of 30 points in total.

Wrisley D. et al (2004), evaluated the internal consistency of data obtaning with the FGA when used with people with vestibular disorders. Six patients with vestibular disorders completed the FGA twice, with an hour's rest between sessions. Internal consistency, or the homogeneity, of items included in the FGA was determined using the Cronbach alpha. This assessment was performed across both testing sessions and within each of the tests. The Cronbach alpha value vary from 0 to 1, the consistency is high when the alpha value is greater than 0.8.

The FGA demonstrated internal consistency within and across both FGA test trials for each patient. Cronbach alpha values were .81 and .77 for individual trials 1 and 2, respectively. The Cronbach alpha was .79 across both trials. This is one of the analyzes performed to validate an item-based measurement instrument. The advantage of this type of test is that we obtain a score of the overall performance of patients in an area. Thus obtaining a much more functional measure than a single outcomes measured with an instrumental technique.

Example 3: Equivalence reliability

Clinical test

In the same study by Wrisley D. et al (2004), they also evaluated the equivalence or the concordance degree of two or more observers regarding to Functional Gait Assessment (FGA), through the Kappa coefficient. For this propuse, seven physical therapists from various practice settings and 3 physical therapist students volunteered to participate. The setup for this assessment was that appears in Figure 10, where all the raters observed the same patients performance.

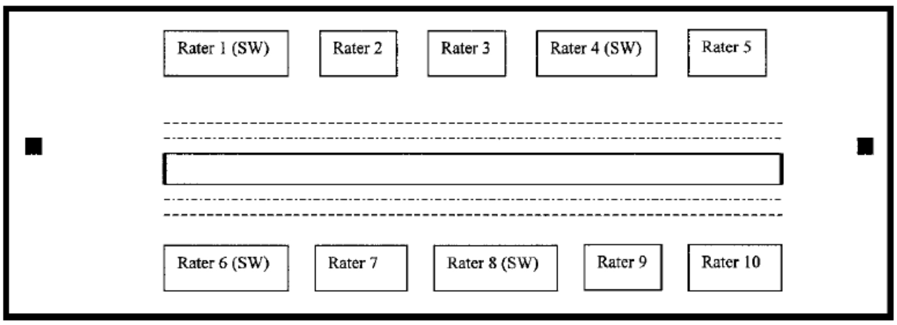

Figure 10 – Sepup of walkway and walkway markings, plus the position of the raters for administration the test. The black squares represent cones indicating starting and stropping points. The raters SW used stropwatch during the test.

The kappa statictic was used for evaluate the interrater agreement (between raters) for individual FGA items and the FGA total. Figure 11 contains percentages of agreement and Kappa values for each item and total FGA scores. Cohen suggested the Kappa result be interpreted as follows: values ≤ 0 as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement (McHugh M. 2012).

	FGA Item										
	1	2	3	4	5	6	7	8	9	10	Total
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

Figure 11 - Interrater Reliability for Individual and Total Functional Gait Assessment (FGA) Items.

Agreement was determined among 10 clinicians who measured 6 patients on the first measurement trial. As there are 45 possible clinician pair agreements per subject, 6 subjects created 270 possible agreements for each FGA item and the total FGA scores. Values in Figure 11 represent mean agreement and kappa across the 45 clinician pairs for each item. Mean test-retest agreement across all clinician pairs, in our opinion, was moderate or better for all items except items 2 (“change in gait speed”) and 5 (“Gait and pivot turn”).

The results about interater reliability of the study by Wrisley D. et al (2004) are useful to exemplify that the scales and tests that are developed by observing the evaluator, can vary in the result according to what the evaluator observes, which is considered a disadvantage of clinical scales and tests with respect to instrumental tools.

2.6 Sensitivity to change and Responsiveness

Sensitivity to change is defined as the ability of an instrument to measure change in state, regardless of whether the change is relevant or meaningful to the decision-maker. Although necessary, sensitivity to change has been described as insufficient for assessing change and establishing treatment effectiveness. A test may be sensitive to a state or diagnosis, but whether it is meaningfull or important cannot be deduced from this property alone.

The importance of this property of gait measuring instruments is related to the evaluation of the impacts of programs and treatments in clinical science. If an instrument is not sensitive enough, it will not be able to evaluate the effect of a certain pathology or treatment on human gait. Such sensitivity is especially relevant in applied settings where program or treatment effects are often not particularly strong, and measurement conditions can be quite variable (Lipsey M. 1983). If an instrument is not sensitive enough to detect a change (effect size) in a certain condition or treatment, large samples are frequently needed to achieve adequate statistical power in order to detect the effect of these interventions.

On the other hand, *responsiveness* is defines as the ability of an instrument to measure a meaningful or clinically important change in a clinical state and has been advocated as an essential property of instruments designed to measure change and effectiveness of interventions. Similar to reliability and validity, responsiveness is not considered a generalizable property and should be assessed for each population and purpose for which the measure is used.

A clinically meaningful or important change can be defined and therefore evaluated from the perspective of the patient, his or her proxy, society, or the health professional. It implies a change that is noticeably, appreciably different that is of value to the patient (or physician). This change may allow the individual to perform some essential task or to perform tasks more efficiently or with less pain or difficulty. These changes also should exceed variation that can be attributed to chance. Jaeschke *et al*. (1989) suggested that a clinically meaningful change could be defined as the minimal important difference. This could be defined as the smallest difference in score in the domain of interest that a patient perceives as a change and that would mandate, in the absence of side effects and excessive costs, modification in the patient's management.

That is how, responsiveness is commonly reported through the minimally important difference (MID) or minimal clinically important difference (MCID) estimate, whereby a change score on a measure should equal or exceed its MID estimate to be considered important. Estimating MID of measures enhances the interpretation of change score, establishing benchmarks to help determine the meaningfulness of change.

Practical approach

When assessing the clinical utility of therapies intended to improve subjective outcomes, the amount of improvement that is important to patients must be determined. The minimal clinically importnat difference (MCID) is the smallest benefit of value to patients. The MCID is a patient-centered concept, capturing both the magnitude of the improvement and also the value patients place on the change. In other words, the MCID defines the smallest amount an outcome must change to be meaningful to patients (McGlothlin A. *et al* 2014).

In 2014, Bohannon R. et al. developed a Systematic Review about Minimal clinically important difference for change in comfortable gait speed of adults with pathology. The conclusion of the study was that changes in gait speed of 0.10 to 0.20 ms−1 may be important across multiple patient groups like people with stroke, hip fracture, multiple sclerosis among others. Therefore, if we want to measure any intervention or evolution in people with the pathologies mentioned in the previous study, we must use a measuring equipment with sufficient sensitivity to record changes in gait speed such as those indicated in the study by Bohannon R. *et al*., that is, capable of registering velocity changes of 0.1 ms-1.

In summary, healthcare professionals when choosing a measurement instrument, should take into account that sensitivity of equipment must be sufficient to measure minimal clinically important difference in the variable they wish to observe in a given population.

In relation to sensitivity, biomechanical assessment techniques have a greater sensitivity than clinical assessment scales to detect changes on gait features. This is because biomechanical measuring instruments are much more accurate equipment. To understand this concept, look at Figure 12.

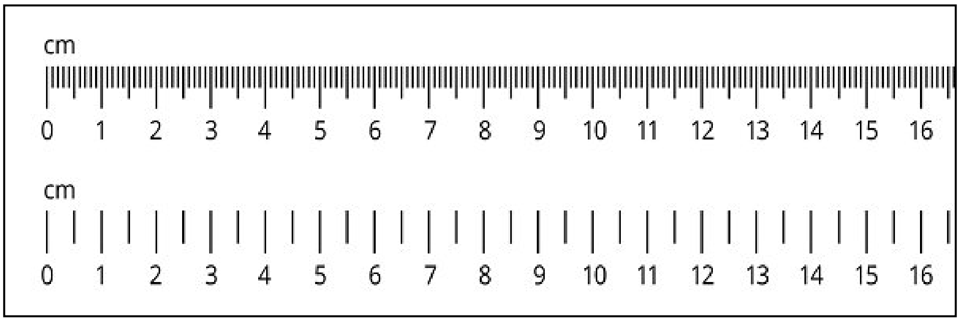

Figure 12 – The image shows two rulers. The upper ruler is graduated in mm and the lower ruler is graduated in cm, and their accuracies are respectively ± 1 mm and ± 1 cm = ± 10 mm. Since the former is more accurate than the latter, the upper ruler is more sensitive than the lower ruler.

The upper ruler is accurate to 1mm and the lower ruler in the figure is accurate to 1cm (10mm). Since the upper ruler is more precise than the lower ruler, it is in turn more sensitive to length measure. This is also the same case among biomechanical assessment instrument to measure human gait. In figure 5, we can see the precision of the differente biomechanical instruments to measure spatio-temporal parameters during human gait. Photogrammetry systems and optoelectronic cameras, are the most accurate equipment when compared to other systemas, therefore, they are also more sensitive to the measurement of spatio-temporal variables than walkway instrumented with pressure sensors, intertial sensors or a stopwatch (Moissenet F. *et al*.). Unfortunately, the scientific evidence on the sensitivity of measurement equipment is limited.

Example: sensitivity among clincial test of gait in people with spinal cord injury

Just as biomechanical assessment instruments differ in their sensitivity to measure certain outcomes, clinical scales also have this feature. Jackson A. et al. (2008) analyzed the utility in clinical practice of current outcome measures used as indicators of improvement in gait and ambulation in the spinal cord injured population. Specifically, they evaluated the following gait and ambulation measures for: the Walking Index for Spinal Cord Injury II (WISCIII), 50-Foot Walk Test (50FTWT), 6-Minute Walk Test (6MWT), 10-Meter Walk Test (10MWT), and Functional Independence Measure-Locomotor (FIM-L). The results are show in Table 4.

Table 4 – Sensitivity to change in locomotion performance of people with spinal cord injury

Clinical test	Sensitivity informed
Walking Index for Spinal Cord Injury II (WISCI II)	- Good sensitivity to change in patients with more impaired gait. - Not incorporate elements of speed or endurance.
50-Foot Walk Test (50FTWT)	- Subject who cannot walk 50 feet are unable to participate in the measurement (floor effects). - Those patients who can walk the distance at a fast pace may also be able to walk a greater distance at that same speed (ceiling effect).
6-Minute Walk Test (6MWT)	- Good sensitivity to change in subjects with better ambulatory function in acute injury and 6 months post injury (while the WISCI II and Lower Extremity Motor Score did not). - In patients with stroke the test showed less sensitivity to change than the 12 MWT. - Sensitivity for the 6MWT is affected by a floor effect among patients who cannot walk for 6 minutes. - Upright resting is allowed during the test, but if the subject sits, then he/she is disqualified. Similarly, a ceiling effect is seen in patients who can continue walking beyond 6 minutes at the same pace.
10-Meter Walk Test (10MWT)	- More sensitivity in patients with greater than 90% recovery of lower extremity motor score at 6 and 12-months post injury (ASIA D). - Better sensitivity than WISCI II particularly in subjects who have less impairment due to the ceiling effect of that test. - Its sensitivity to change in ambulation is more comparable to other gait speed tests such as the 6MWT because similar traits are being measured. - Responsiveness to change has also been shown in the stroke population using a modified 6-meter distance as a variation of the 10MWT. - Better sensitivity than the Lower Extremity Motor Score for a longer period (6 month). - The outcomes measure is less sensitive to change in locomotion in chronic SCI where there is little change in strength.
Functional Independence Measure-Locomotor (FIM-L)	- Poor sensitivity to show a change in one subjects' score pre and post training despite their mode of ambulation changing from wheelchair to walking. - A ceiling effect is revealed in that the test shows poor sensitivity to change in subjects with better walking abilities. Also, scoring of the FIM-L does not account for the speed with which test subjects complete the required distances.

The conclusion of the study was that in people with spinal cord injury, the 10-Meter Walk Test and Walking Index for Spinal Cord Injury II are the most valid and clinically useful test in evaluating improvements in gait and ambulation, due, between other things, to its good sensitivity to assess changes in ambulation in this group of patients with high recovery and with more altered gait, respectively.

In the previous example, we have seen that the sensitivity of the gait assessment tests is affected by two phenomena called the ceiling effect and the floor effect, which are related to the feasibility of the test itself to be performed by the target people.

2.7 Floor and ceiling effect

If the range of function covered by a measure is less than the range experienced by patients, the measure may lack responsiveness. The potential for floor and ceiling effects is often assessed by examining response patterns. If there are spikes at the highest or lowest response option this is often interpreted as evidence of ceiling or floor effects, respectively. However, when using measures to assess the effectiveness of interventions prospective evidence of the performance of a measure is more important than whether or not there are spikes (Feeny DH. *et al*. 2013).

If the participants' scores cluster toward the high end (or best possible score) of the measure/instrument (ceiling effect) or in the down end (floor effect), changes experienced by patients in gait performance can be biased. This is due to patients could be "worse off" than the measure could capture or "better off" than the instrument can measure.

If we compare the instrumental assessment techniques with the clinical assessment scales, we will observe that the instrumental techniques have a floor effect, while the clinical gait assessment scales have a ceiling effect. The reason for this difference is the ease or difficulty with which each of them can be performed by patients. The assessment scales are defined following structured questionnaires, making categorical or discrete classifications that reduce sensitivity with respect to other clinical assessment instruments or technologies. In other words, small changes that occur in the functional capacity of the patient as a result of the intervention of the professional, are very difficult to identify. This fact implies, depending on the design of the scale, a ceiling effect. However, people with severe gait impairment will not be able to perform the complex assessment protocols required by many of the biomechanical instrumental techniques, so the most sensitive assessment for this type of patient will be the clinical assessment scales.

On the other hand, the instrumental techniques of biomechanical assessment have a floor effect because the assessment protocols require a greater demand from the patient to be carried out. That is, in the majority of times, the assessment of the gait of people with severe impairment in ambulation is not possible with instrumental techniques, so the floor effect may limit entry of severely injured individuals until they can walk with the with the required instrumentation.

Example

Although the studies that analyze the statistical properties of valuation instruments are limited, Middleton A. and Fritz S. (2013), reported evidence about accurate assessment of gait in older adults since it is an important aspect of clinical practice for clinicians working with this population. Selected measures are as follows: Gait: gait speed, gait symmetry, gait endurance, adaptability of gait, dual task performance during gait, and self-reported confidence during gait. Figure 13 provides a summary of outcome measures included in the article as a quick reference for clinicians.

Measure	Assessing	Predictive abilities	Scale[a]	MDC_{95}
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

Figure 13 - Summary table of gait outcomes measures. Minimal detectable change (MDC), not applicable (n/a).

The minimal detectable change (MDC) provides clinicians with a reference value to determine if true change has occurred. The MDC95 quantifies the smallest amount of change required to surpass measurement error and variability with a 95 % confidence level. When reported in the literature, MDC scores for outcome measures can be valuable tools for clinicians. In Figure 13 it is possible to observe the minimum detectable change for the gait outcomes most used in clinical practice, measured by the instruments mentioned in the first column.

This study is useful to explain that patients who perform below the minimum detectable change in each of the mentioned variables, it is preferable that they be measured with other instruments that allow detecting changes below the referenced value as a minimum detectable change.

3. Key ideas

- The medical staff have to know the methodological characteristics and statistical properties at the time of choosing a gait assessment tool. This is necessary to avoid methodological errors and biases in the measured results.

- Regarding usability, clinical scales and tests have the advantage that they are possible to develop in a short time, they do not require specialized training from the rater and they can be used in any context such as in clinical practice.

- The equipment required to use clinical tests and scales is much less and accessible than the equipment needed to perform a gait assessment with biomechanical assessment instruments.

- The most important quality of instrumental biomechanical assessment techniques is that they provide objective data obtained without interpretation of the evaluator (i.e. directly assessment of one or more dimensions of gait pattern), so their use is mainly in the research area. On the contrary, the information obtained through scales and clinical tests is influenced by the interpretation and perception of the evaluator.

- The high precision of the instrumental measurement techniques gives them the quality of being more valid to measure a gait characteristic than the scales or clinical tests.

- The reliability is usually better in biomechanical instruments because the repeatability of the measurement does not depend on the observer but on other factors, such as performing the measurement with a standardized protocol.

- The more accurate a measuring instrument is, the more sensitive to change the instrument will be. The sensitivity of equipment must be sufficient to measure minimal clinically important difference in the outcomes that professional intent to observe in a given population.

- The clinical scales and tests have a greater tendency to have a ceiling effect, that is, the participants' scores cluster toward the high end (or best possible score) of the measure / instrument. On the other hand, the instrumental techniques have a greater floor effect, where the participants' scores cluster toward the down end. This is due to patients could be “better off” than the measure could capture or “worse off” than the instrument can measure.

4. References

[1] Gutierrez-Clavería M, Beroíza T, Cartagena C, Caviedes I, Céspedes J, Gutiérrez-Navas M, Oyarzún M, Palacios S, Schönffeldt P. Guidelines for the six-minute walk test. RevChil Enf Respir2009; 25: 15-24.

[2] Innerd P, Catt M, Collerton J, Davies K, Trenell M, Kirkwood T, Jagger C. A comparision 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 biomedcal 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.