Research Article

Intra/inter-rater and parallel-forms reliability of the three instruments for the assessment of ankle dorsiflexion range of motion and degree of asymmetry during the weight-bearing lunge test

Germán Cánovas-Ambit1*, Jorge Godinez-Leal1, José A García-Vidal1, Aitor Baño-Alcaraz1, Francesc Medina-Mirapeix1 and Rodrigo Martín-San Agustín2

1Department of Physiotherapy, Faculty of Medicine, University of Murcia, Campus de Espinardo, 30100 Murcia, Spain
2Department of Physiotherapy, University of Valencia, 46003 Valencia, Spain

Abstract

Background: This study sought to evaluate the intra- and inter-rater reliability of the three instruments for the measurement of ankle dorsiflexion (ADF) range of motion (ROM) and degree of asymmetry during the Weight-Bearing Lunge Test (WBLT) and to examine the agreement between them in asymptomatic and symptomatic subjects.

Methods: Thirty-two healthy subjects and thirty-three subjects with pathologies were measured on two occasions by two different examiners and using different methods while performing the WBLT.

Results: Intra- and inter-rater reliability of Veloflex showed good to excellent values (ICC’s > 0.88) for both the measurement of ADF-ROM and degree of asymmetry in asymptomatic and symptomatic subjects. In addition, Veloflex and the tape measure showed acceptable agreement between their values in symptomatic subjects.

Conclusion: Veloflex is reliable for the measurement of ADF-ROM and the degree of asymmetry in asymptomatic and symptomatic subjects. This system may be exchanged by the tape measure, however, exchanging Veloflex with the Dorsiflex app is not recommended.

Introduction

Maintaining appropriate ankle dorsiflexion (ADF) range of motion (ROM) is important for the performance of basic activities such as walking, running, bending, stair climbing, or jumping [1-4]. Walking or descending stairs requires a minimum ROM of 10° ADF, whereas running or sprinting requires a minimum of 20° - 30° [3]. Furthermore, a low ADF-ROM can be a risk factor for injuries such as fasciitis, foot, knee, hip, or ankle problems [5-7]. In addition, some authors have suggested that when there are asymmetries between the ROM of both ankles this is a risk factor as it can cause changes in force and balance [5,8-11].

The ADF-ROM was initially measured in the open kinetic chain [12-15], mainly in subjects with decreased ROM associated with pathological foot conditions (e.g. sprain, fractures), however, in the last twenty years the closed kinetic chain is generally used by clinicians and researchers for measuring subjects with and without those conditions [16]. Thus, the Weight-Bearing Lunge Test (WBLT) is the test that is most often used [17-21]. During the WBLT, subjects are asked to lunge forward until their knee touches the wall, without lifting the heel off the ground, thus placing the ankle in maximum dorsiflexion [4,22]. The examiner only measures at the end range. This test has shown more reliability than
other tests using the closed kinetic chain. Several authors [23-26] suggest that the lower reliability of other tests could be due to the fact that the examiner must perform the movement at the same time as the measurement.

Traditionally, several instruments such as universal goniometers, inclinometers, and tape measures have been used for measuring ADF-ROM in open and closed kinetic chains [3,22,27]. Most of these instruments have good validity and reliability when used in the WLBT both in subjects [1,18,19,28,29] and without pathological conditions [2,3,25]. Recently, technological alternatives to these methods have emerged for measuring ADF-ROM and asymmetries during WLBT, such as mobile applications [17,30-32], video analysis [33], or motion capture [34,35], which have been shown to be valid and reliable for this purpose mainly in healthy people [30,32,34], Of these, mobile applications are the most used in clinical settings since these are an easy-to-use method with a low price. However, similar to most traditional instruments, they often require the patient to hold a position while the examiner must measure and monitor the movement at the same time. Although new reliable instruments based on optoelectrical systems such as the VeloFlex [36] can provide measurements without those drawbacks, it is still necessary to examine whether they are reliable in WLBT.

Currently, there is a lack of consensus in the clinical settings concerning how the ADF-ROM and asymmetries should be measured in the WLBT. Moreover, the reliability of some mobile applications and, even some traditional methods (e.g. trigonometry from two distances) often used in clinical settings were only examined in healthy patients, and therefore they can be of questionable reliability and interchange in clinical settings. Consequently, it is necessary to establish the reliability and agreement of some measurement methods in the WLBT for clinical settings. Furthermore, it is necessary to also compare reliability and agreement between these methods in healthy subjects [22].

Thus, the two main aims of this study were (1) to evaluate the intra- and inter-rater reliability of one traditional (trigonometry method) and two technological instruments (VeloFlex and the DorsiFlex app) for the measurement of the ankle DF-ROM and the degree of asymmetry in the WBLT and (2) to examine the agreement between scores from these alternate forms of ROM measurement in subjects with and without pathological conditions.

Materials & methods

Participants and study design

Thirty-two asymptomatic subjects without any pathological foot condition and 33 subjects with foot pathology voluntarily participated in the study. In this paper, the latter were called symptomatic subjects because most of them reported pain and/or limited ROM. Subjects were recruited through advertising on the Martial Arts Federation of Murcia. The specific inclusion criteria for asymptomatic subjects were: (1) age between 18 and 40 years; (2) not having undergone surgery on the lower limb; (3) not having suffered episodes of pain in the lower limb two months before data collection. Symptomatic subjects who reported some type of pathology (e.g. sprain, fractures) were included in this group the inclusion criteria were: (1) age over 18 years; (2) presenting at least one week of the evolution of the pathology. Those who presented bilateral pathology were excluded. Furthermore, those with possible neurological alert signs, vascular problems, or relevant medical history were excluded from both groups. All participants received a detailed explanation of the study procedures and signed an informed consent form. The study was approved by the research ethics committee of the University of Murcia (2236/2018).

All participants were measured over two identical sessions. For the asymptomatic subjects, a one-week interval was applied between sessions, and for the symptomatic subjects, measurements were repeated after one hour. Both sessions took place at the same location, under similar environmental conditions, and at the same time of day. During each session, the movement was measured with three different instruments: a tape measure, the DorsiFlex app, and VeloFlex. To examine inter-rater reliability, one examiner (Examiner A) with experience using the three measurement methods, carried out measurements in both sessions with all measurement instruments, whereas, to examine inter-rater reliability, another examiner (Examiner B), who was inexperienced in its use, performed the same measurements only in the second session. The examiners’ order of measurement was randomized in this session to avoid possible confounding factors. To examine the agreement between the three instruments, we used a parallel-forms reliability design using data from the first session.

The degree of asymmetry is the result of subtracting the DF-ROM of one ankle minus the DF-ROM of another ankle. This result obtained is used to examine the reliability. For the group of symptomatic subjects, the ankle with and without pathology was compared. For the group of asymptomatic subjects, the dominant leg was compared with the non-dominant one. It was a prior assumed that variability between ankles and average difference within this last group would be lower than in symptomatic subjects.

Instrumentation

VeloFlex is an optoelectronic device capable of recording data on the position and range of motion of any joint. It consists of three elements, which are: a camera, a laptop, and some markers. The markers are reflective stickers that adhere to the skin of the participants. These are placed at reference points so that the camera can measure movement, by recording movement both dynamically and statically. The camera is placed 1 - 1.5 meters from the participant and at
the height of the joint being measured. The camera is able to record 0.1-degree increments. Veloﬂex has been validated to measure the ROM of the lower limb and upper limb joints [36,37]. In this study, version 1.0 of Veloﬂex was used, by the Instruments and Sports technology SL company in Murcia, Spain.

For the trigonometric method, two tape measures measuring 150 cm were used to measure the distance from the heel to the wall (placed on the ground), as well as from the knee to the ground (placed vertically on the wall). Once the wall-heel distance and the floor-knee distance were calculated, through trigonometric formulas the DF-ROM is obtained, which is the data to be compared in this study. To obtain this angle, the following formula is used: [4,22].

\[
\tan^{-1} \frac{\text{Floor - knee distance}}{\text{Wall - heel distance}}
\]

The Dorsiﬂex application is one of the most downloaded applications for Apple devices, which assesses WBLT by measuring the tibia’s inclination when performing the ankle dorsiflexion movement in a closed kinetic chain. To do so, the electronic device on which the application has been downloaded is placed vertically, attached to the anterior aspect of the tibia, just below the anterior tuberosity [31]. Once it is in this position, the start button is pressed to measure the movement. The accuracy of the tibial tilt measurement by the Dorsiﬂex application was 0.17. As the results of the application show the tibial inclination, to calculate the DF-ROM, it is necessary to calculate the complementary angle by subtracting the degrees obtained from a right angle of 90°.

Procedure

A single movement, the WBLT, was examined in each session. This movement was performed with the dominant and non-dominant lower extremities, randomly designating the order of measurement, both in the asymptomatic and non-asymptomatic groups.

For each participant, three measurements were made, of which the mean was obtained for data analysis. Previously, the subjects performed five full movements to familiarize themselves with these movements and as a warm-up. Symptomatic subjects were explained that movements should not cause pain.

To prepare participants for the test, the examiners explained how the movement should be performed, before to beginning the measurement. To proceed with the WBLT, the participant had to align the big toe and heel with a tape measure placed on the floor. Once aligned, the movement was carried out noting the exact point where, without taking the heel off the ground, the participant could touch a vertical tape placed on the wall with his knee [38]. In the event that the participant failed to touch the wall in the first attempt, the participant had to approach the wall centimeter by centimeter, to ensure that the movement was carried out in its maximum range. In addition, the position of the supporting foot was standardized, and the entire sole of the foot must always be in contact with the ground [39].

The participants had to hold the final position of the WBLT for a few seconds to enable measurement with the tape measure and to place the Dorsiﬂex on the tibia. In the case of the Veloﬂex, the DF-ROM was automatically recorded. Upon completion of the measurements, the participants returned to the starting position and rested for 60 seconds to repeat the procedure. For data collection with Veloﬂex, adhesive markers were placed prior to the movement so that the camera could capture the position. The placement of the markers consisted of one on the malleolus of the fibula and the other on the head of the fibula (Figure 1). Once the movement had been carried out, using the horizontal as a reference, the degrees of ankle dorsiflexion was determined. Before performing the WBLT, one measuring tape was placed on the floor and the other vertically on the wall. When the final position was reached, the measurements obtained both vertically and horizontally were recorded, once these data were obtained, the degree of DF-ROM was calculated with a trigonometric formula [22]. The dorsiflexion measurement was performed by placing the electronic device just below the anterior tibial tuberosity to indicate the inclination of the tibia in that position [31].

Analysis

All statistical analyses were carried out with the SPSS software, version 19.0. The information from the demographic and anthropometric data of the participants, as well as the means of the range of mobility of the WBLT, were analyzed using the means and standard deviations. To evaluate the intra- and inter-examiner reliability of the different instruments, the intraclass correlation coefficients (ICC) were calculated. The reliability was classified as excellent (ICC > 0.90), good (ICC = 0.76 – 0.90), moderate (ICC = 0.51 – 0.75), and poor (ICC < 0.50) [40]. The standardized measurement error (SEM) was also analyzed, which was calculated with the formula SD × √1-ICC and the percentage that this represented.
of the mean, i.e., the SEM%. The minimum detectable change (MDC) was calculated with the formula SEM × 1.96 × √2. For the reliability of the asymptomatic group, the measurements of the dominant leg were used and for the symptomatic group, the measurements of the leg with the pathological ankle were used.

To measure the intra- and inter-examiner reliability of the asymmetries, the results obtained when comparing the dominant leg with the non-dominant leg in the asymptomatic group were used, and in the symptomatic group, the ankle with and without pathology was compared. Because calculations of the ICC can be questionable when there is a lack of variability among subjects’ scores, these can be verified by examining the significance of the between-subjects variance, as recommended [40,41]. When this was not significant, ICCs and associated statistics were not calculated.

For the analysis of the agreement between instruments both for DF-ROM and degree of asymmetry, Bland Altman plots were used. Furthermore, we calculated the upper and lower limits of agreement (LoA) and the mean of the difference between instruments (absolute difference and percentages).

Results

Participant characteristics

Thirty-two asymptomatic subjects (15 males) were recruited with a mean age of 28.66 years and a BMI of 23.9 kg/m². For the group of symptomatic subjects, 33 subjects were recruited (19 males) with a mean age of 40.5 years and a BMI of 25.7 kg/m². Symptomatic subjects had a mean pain of 3.12 on the VAS and meaningful evolution of pain of 3.12 between instruments (absolute difference and percentages).

Intra/inter-rater and parallel-forms reliability of the three instruments for the assessment of ankle dorsiflexion range of motion and degree of asymmetry during the weight-bearing lunge test

Intra/inter-rater and parallel-forms reliability of the three instruments for the assessment of ankle dorsiflexion range of motion and degree of asymmetry during the weight-bearing lunge test

Discussion

The main findings of this study were that Veloflex (1) is a reliable tool to measure the ADF-ROM and the degree of asymmetry between legs in the WBLT and (2) can be interchangeable with the tape measure but not with Dorsiflex by showing an acceptable agreement only with the first. Furthermore, whereas the tape measure and Dorsiflex are reliable for the measurement of ADF-ROM during WBLT in both symptomatic and asymptomatic subjects, the agreement between these is poor and therefore both tools cannot be interchangeable.

The intra- and inter-rater reliability of the Veloflex to measure ADF-ROM during the WBLT proved to be excellent...
Table 1: Intra- and inter-rater reliability of the three instruments for DF-ROM and degree of asymmetry in symptomatic and asymptomatic subjects.

<table>
<thead>
<tr>
<th>Measurement variable</th>
<th>Test (SD)/ Retest (SD)</th>
<th>ICC (95% CI)</th>
<th>SEM</th>
<th>MCD</th>
<th>Test (SD)</th>
<th>ICC (95% CI)</th>
<th>SEM</th>
<th>MCD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptomatic subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Veloflex</td>
<td>44.12° (6.07°)/ 44.68° (6.09°)</td>
<td>0.95 (0.90 to 0.97)</td>
<td>1.18° (2.66%)</td>
<td>3.27°</td>
<td>44.02° (5.42°)</td>
<td>0.88 (0.76 to 0.94)</td>
<td>1.73° (3.94%)</td>
<td>4.81°</td>
</tr>
<tr>
<td>Tape Measure</td>
<td>44.68° (4.99°)/ 44.63° (5.59°)</td>
<td>0.98 (0.97 to 0.99)</td>
<td>0.57° (1.29%)</td>
<td>1.59°</td>
<td>44.30° (5.41°)</td>
<td>0.98 (0.97 to 0.99)</td>
<td>0.58° (1.32%)</td>
<td>1.63°</td>
</tr>
<tr>
<td>Dorsiflex</td>
<td>44.54° (5.13°)/ 44.56° (4.82°)</td>
<td>0.98 (0.96 to 0.99)</td>
<td>0.66° (1.48%)</td>
<td>1.83°</td>
<td>44.39° (5.33°)</td>
<td>0.94 (0.89 to 0.97)</td>
<td>1.17° (2.63%)</td>
<td>3.24°</td>
</tr>
<tr>
<td><strong>Asymptomatic subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Veloflex</td>
<td>5.98° (5.38°)/ 5.41° (6.39)</td>
<td>0.95 (0.89 to 0.97)</td>
<td>1.33° (24.7%)</td>
<td>3.68°</td>
<td>5.34° (5.72°)</td>
<td>0.90 (0.76 to 0.94)</td>
<td>1.66° (34.47%)</td>
<td>5.15°</td>
</tr>
<tr>
<td>Tape measure</td>
<td>5.67° (5.37°)/ 5.87° (5.89°)</td>
<td>0.97 (0.94-0.98)</td>
<td>0.93 (16.4%)</td>
<td>2.57°</td>
<td>5.56° (5.97°)</td>
<td>0.96 (0.92 to 0.98)</td>
<td>1.11° (19.76%)</td>
<td>3.07°</td>
</tr>
<tr>
<td>Dorsiflex</td>
<td>5.18° (6.29°)/ 4.74° (6.69°)</td>
<td>0.955 (0.90 to 0.97)</td>
<td>1.86 (37.5%)</td>
<td>5.15°</td>
<td>5.20° (6.41°)</td>
<td>0.922 (0.84 to 0.96)</td>
<td>1.75 (33.71%)</td>
<td>4.85°</td>
</tr>
</tbody>
</table>

Degree of asymmetry

<table>
<thead>
<tr>
<th>Measurement variable</th>
<th>Mean Difference (%) / Effect Size®</th>
<th>LoA-/LoA+</th>
<th>Mean Difference (%) / Effect Size®</th>
<th>LoA-/LoA+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptomatic subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF-ROM</td>
<td>0.56° (1.26%) / 0.11</td>
<td>-3.30° / 2.18°</td>
<td>0.42° (0.95%) / 0.08</td>
<td>-6.41° / 5.56°</td>
</tr>
<tr>
<td>Degree of asymmetry</td>
<td>0.31° (5.18%) / 0.05</td>
<td>-4.82° / 4.18°</td>
<td>0.80° (13.37%) / 0.14</td>
<td>-9.09° / 7.47°</td>
</tr>
<tr>
<td><strong>Asymptomatic subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF-ROM</td>
<td>5.1° (9.82%) / 0.96</td>
<td>1.46° / 8.83°</td>
<td>6.26° (12.06%) / 1.19</td>
<td>1.58° / 10.93°</td>
</tr>
<tr>
<td>Degree of asymmetry</td>
<td>0.18° (15%) / 0.07</td>
<td>-3.44° / 3.81°</td>
<td>0.43° (35.83%) / 0.17</td>
<td>-3.38° / 4.52°</td>
</tr>
</tbody>
</table>

Table 2: Agreement of Veloflex with the tape measure and the Dorsiflex app for DF-ROM and degree of asymmetry in symptomatic and asymptomatic subjects.

<table>
<thead>
<tr>
<th>Measurement variable</th>
<th>Veloflex with the tape Measure</th>
<th>Veloflex with Dorsiflex</th>
<th>LoA-/LoA+</th>
<th>LoA-/LoA+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptomatic subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF-ROM</td>
<td>0.56° (1.26%) / 0.11</td>
<td>-3.30° / 2.18°</td>
<td>0.42° (0.95%) / 0.08</td>
<td>-6.41° / 5.56°</td>
</tr>
<tr>
<td>Degree of asymmetry</td>
<td>0.31° (5.18%) / 0.05</td>
<td>-4.82° / 4.18°</td>
<td>0.80° (13.37%) / 0.14</td>
<td>-9.09° / 7.47°</td>
</tr>
<tr>
<td><strong>Asymptomatic subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF-ROM</td>
<td>5.1° (9.82%) / 0.96</td>
<td>1.46° / 8.83°</td>
<td>6.26° (12.06%) / 1.19</td>
<td>1.58° / 10.93°</td>
</tr>
<tr>
<td>Degree of asymmetry</td>
<td>0.18° (15%) / 0.07</td>
<td>-3.44° / 3.81°</td>
<td>0.43° (35.83%) / 0.17</td>
<td>-3.38° / 4.52°</td>
</tr>
</tbody>
</table>
Intra/inter-rater and parallel-forms reliability of the three instruments for the assessment of ankle dorsiflexion range of motion and degree of asymmetry during the weight-bearing lunge test

reliable tools for measuring ADF-ROM in both symptomatic and asymptomatic subjects. This is consistent with previous studies using the tape measure for both populations [3,16,18] as well as for the DorsiFlex app in asymptomatic subjects [31]. Thus, our study is the first to address the reliability of the DorsiFlex app for symptomatic subjects, showing that in the face of reduced ADF-ROMs, it is a reliable tool. Previously, Balsalobre, et al. showed excellent levels of reliability for the DorsiFlex app, as well as for identifying asymmetries between legs in asymptomatic subjects [31]. Thus, our study is the first to address the reliability of the DorsiFlex app for symptomatic subjects, showing that in the face of reduced ADF-ROMs, it is a reliable tool. Previously, Balsalobre, et al. showed excellent levels of reliability for the DorsiFlex app, as well as for identifying asymmetries between legs in asymptomatic subjects. We cannot determine this reliability because our sample of asymptomatic subjects was very homogeneous for asymmetry. This homogeneity was to be expected because usually, these subjects have low levels of asymmetry. Our findings suggest that reliability analysis in healthy people is not recommended for the degree of asymmetry of these tools between legs.

As all three alternate forms of ROM measurement had good to excellent reliability, and even two of them (the tape measure and Veloflex) can be interchangeable, our study provides relevant implications for practice. First, all three instruments can be used in clinical settings and in symptomatic subjects. Second, regarding the three instruments, the traditional method can be recommended as the first choice as it requires lower cost, equipment, and training than the two technological methods. Third, when professionals need to interchange the tape measure with any of the two technological instruments examined, the Veloflex is recommended because, according to our study, it provides lower limits of agreement than DorsiFlex. Four, technological methods should not be interchanged because they provide the highest limits of agreement. A possible reason for the latter could be the disposition of the mobile device; although the position is standardized, it is difficult to always place the smartphone at the same point. It would also be advisable to evaluate the reliability using other smartphone models. In addition, the application depends on the internal processor of the mobile which could generate some degree of failure in the process.

This study provides valuable information on the reliability of the Veloflex and other instruments despite several limitations. First, intra- and inter-rater reliability was performed by two raters, a procedure that, according to the literature, is usually enough to study reliability properties, however, the participation of three or more evaluators could have provided even more reliable information. Second, the WBLT is measured with the subject holding a position, which may limit the ability to extrapolate the results to dynamic movement or walking. Future studies should address these limitations.

Figure 2: Bland-Altman agreements Bland-Altman plots for Veloflex and the tape measure (A and C) or DorsiFlex (B and D) during WBLT: (A and B) DF-ROM. (C and D) degree of asymmetry.
Conclusion

Veloflex proved to be a reliable intra- and inter-rater tool for the measurement of DF-ROM using the WBLT in both symptomatic and asymptomatic subjects, and also for the measurement of the degree of asymmetry in symptomatic subjects, furthermore, these measurements are very close to those obtained using the tape measure. Therefore, both instruments are interchangeable. In addition, the tape measure and the Dorsiflex app proved to have intra-inter-rater reliability for the measurement of DF-ROM in both symptomatic and asymptomatic subjects, although the exchange of Dorsiflex for Veloflex or the tape measure is not recommended due to differences in the degree of agreement.

Acknowledgment

The authors thank members of the Martial Arts Federation of Murcia for their participation in the study.

References

24. Krause DA, Cloud BA, Forster LA, Schranks JA, Holland JH. Measurement of ankle dorsiflexion: a comparison of active and


