Original Article (short paper)

Intra and inter-rater agreement of the Dynamic Movement Assessment ™ Agreement of the DMA™

Priscila dos Santos Bunn[1,2,3] ⓘD , Hélcio Figueiredo da Costa[2] ⓘD , Celso José da Silva Júnior[2] ⓘD ,
Saulo de Almeida Silva[2] ⓘD , Ricardo Costa Abrantes Júnior[2] ⓘD , Elirez Bezerra da Silva[2] ⓘD

[1]*Universidade do Estado do Rio de Janeiro, Instituto de Educação Física e Esporte, Progra-
ma de Pós Graduação em Ciências do Exercício e do Esporte, Rio de Janeiro, RJ, Brasil;*
[2]*Centro de Educação Física Almirante Adalberto Nunes, Laboratório de Pesquisa
em Ciências do Exercício, Rio de Janeiro, RJ, Brasil;*
[3]*Programa de Pós-graduação em Desempenho Humano Operacional, Rio de Janeiro, RJ, Brasil*

Associate Editor: Gustavo R da Mota. UFMT/Uberaba, MG, Brazil

Abstract - Aims: To investigate the inter-and intra-rater agreement of the Dynamic Movement Assessment (DMA™) risk classification. **Method**: In this study, after the anthropometric measurements were made, 17 female soccer athletes were filmed performing the six DMA™ tests (full squat, step-up, single-leg squat, jump test, test plank, and side plank). Both, major and secondary deviations, were observed during the tests. Two experienced health professionals performed video analysis using Kinovea 8.15.0 (inter-rater agreement). To assess the intra-rater agreement, the same video analysis was performed two months later. Participants were rated from 0 to 21 points and at low, medium, moderate, and high risk of developing musculoskeletal injuries. To assess the reliability of the assessment of movement patterns of DMA, the intraclass correlation coefficient (ICC) was performed with a 2-way random-effects model with an absolute agreement (inter-rater) and a 2-way mixed-effects model and consistency (intra-rater). Weighted Kappa Agreement Analysis (k_w) was performed with linear weights to assess the level of agreement related to the risk classification of DMA (high, moderate, medium, or minimum). The Analysis was performed with StatsDirect v.3 and SPSS (23.0). **Results**: Comparing the number of points between the inter-and intra-rater, the ICC was 0.91 (95% CI = 0.74-0.97) and 0.84 (95% CI = 0.59-0.94), respectively, with k_w = 0.46 (P = 0.02) intra -rater and k_w = 0.46 (P = 0.006) inter-rater (Table 9). **Conclusion**: DMA has excellent inter-and intra-rater reliability to evaluate movement patterns and classify the risk of musculoskeletal injuries.

Keywords: movement; test reproducibility; athletic injuries.

Introduction

Musculoskeletal injuries have a high incidence among participants in physical activities, and it is the predominant cause for their absence from sports activities and work[1–4]. In sports, injuries are more common on the lower limbs and trunk[5,6].

To identify individuals at high risk of injury, certain assessment strategies have been developed[7]. Some authors suggest that the presence of certain movement patterns, such as the presence of dynamic valgus, is considered the main cause of musculoskeletal injuries in athletes and practitioners of recreational or occupational activities[8,9]. Moreover, the majority of training sessions by these individuals comes associated with an increased risk of musculoskeletal injuries[8,10].

The Dynamic Movement Assessment™ (DMA™) is a tool that uses two-dimensional video analysis to measure the pattern of movement in six functional tests and classifies the risk of injury in the full squat test (FST)[11–18], step-up test (SUT)[19], single-leg squat test (SLST)[20], single-leg hop test (SLHT)[21,22], plank test (PT)[23], side plank right test (SPRT)[23] and side-plank left test (SPLT)[23]. The first four tests are based on movements and gestures present in various sports. The plank tests are designed to assess the strength of the core muscles. This method emerged as an alternative assessment because of the high cost and difficult access by the population to the bio-mechanics laboratories[24]. The use of DMA™ may assist in the identification of injury risk and the subsequent implementation of programs may reduce the risk of injury.

The DMA™ can be carried out by Physical Education professionals and Physiotherapists trained in the method, but the ability of DMA to predict the risk of injury is unknown. For scientific research, it is necessary that the instrument used to collect data is valid and reliable[25,26]. Validity refers to the ability of an instrument to measure what it is intended to measure. Reliability, on the other hand, allows the identification of potential sources of measurement error, that can compromise the evaluation. Thus, an instrument can be trusted without being valid, but it cannot be valid without being reliable[26]. The initial step to assess the validity of a method is to know its reliability since the validity coefficients are limited by the reliability of the measure. When the criteria are not perfectly reliable, the maximum possible validity is smaller than or equal to the square

root of the reliability coefficient[27]. DMA has a subjective nature and the decision of the evaluator can impact the ability to predict injuries[25]. However, the inter-and intra-rater to evaluate the movement patterns and classify the risk of injury are not known. Therefore, this study aimed to analyze the inter-and intra-rater agreement of DMA™ to evaluate the movement patterns and classify the risk of musculoskeletal injuries.

Methods

This agreement study was approved by the Ethics in Research at the Naval Hospital Marcilio Dias Committee (CAAE:94000518.0.0000.5256) and drawn from the recommendations of the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) proposed by Kottner et al., 2011[28].

Sample

Female football athletes of the Physical Education Center Almirante Adalberto Nunes from Brazilian Navy (CEFAN) were invited to participate in this study. A healthcare professional evaluated the musculoskeletal pain and complaints of the volunteers in the previous seven days using the Nordic Musculoskeletal Questionnaire[29]. Only the injury-free athletes during the seven days before the assessment were included in the study. Once the criteria above were met, the assessment of height and body mass was held, allowing the calculation of the body mass index (BMI). The fat percentage was estimated following the protocol and Pollock's equations[30].

The sample size was calculated after conducting a pilot study with six participants. The aim was to calculate the observed agreement and the chance agreement of the DMA[31]. After calculating the scores achieved by the assessors, it was possible to prepare a 2x2 table with the number of concordant and discordant answers associated with the minimum, medium, high, or moderate risk of musculoskeletal injuries of participants. Afterward, the Kappa Index (k) was calculated according to equation (1)[31]. Considering $k = 0$ as null hypothesis[31], data were used to obtain the sample size recommended in the literature to assess agreement in studies where the outcome is an ordinal categorical variable[31]. The result obtained in the pilot was a $k = 0.70$. Considering the accuracy of the test being 80%, an α error of 0.05, calculations were made based on a sample of 17 participants.

Procedure for the Dynamic Movement Assessment™

The individuals were invited to participate in the study during the team's preseason. Before starting the tests, a health care professional investigated the data: date of birth, sports category, history of musculoskeletal injuries or recent surgeries, and musculoskeletal pain complaints in the previous seven days. If criteria were met, the athletes would then be evaluated using the DMA™.

The DMA™ tests were filmed by two physical therapists trained in the DMA™ method, with six years of experience and having performed more than a hundred tests. The participants did not do physical activities on the assessment day. As recommended by the DMA™, the athletes were not familiarized with the method. Initially, they were given verbal directions for each test regarding the execution of the exercise and there were a few repetitions so that all participants received the same information regarding the movement in each functional test.

All tests were filmed with a digital camera (Nikon Coolpix P600 16.1-megapixel and 60 images per second, Japan) positioned on a tripod leveled both horizontally and vertically, with a distance of 3.04 meters from the participant. The DMA™ consisted of the following functional tests, in order, with no gaps between them, as previously described by Nessler & Haile[24]: FST, SUT, SLST, SLHT, PT, SPRT, and SPLT (Figure 1). The tests were performed in this sequence without any gaps between them. The participants were asked to inform if they felt any pain prior to the test and any that came along with the movements. If any pain related to the test was reported, the movement would be finalized and given a zero score[24]. As some of the movement patterns analyzed are not viewed from the frontal plane (PT, SPRT, and SPLT) the two evaluators were positioned perpendicular to the camera during these tests. Participants performed 10 repetitions on each side for the unilateral tests (SUT, SLST, and SLHT), 20 repetitions of FST (10 with their front to the camera and 10 with their back), and maintained the plank test for 60 seconds (PT, SPRT, and SPLT).

Video analysis and score calculation

The analysis of the videos was done using Kinovea software (v. 8.15.0). The image was calibrated using a red piece of wood 30 cm in length. This material was positioned next to the participant during all tests (Figure 1). Each of the six tests was rated from zero to three points, three points being the best score a participant could get in each test. Then, the possibilities of classification were: three points (normal movement); two points (few deviations); one point (many deviations or movement patterns associated with a high risk of injury); or zero (if any pain was associated with the test).

The movement patterns observed in the first four tests were: the magnitude and direction of the lateral deviation (FST), the height of the heel elevation off the ground (FST), the total of times when the balance was lost (FST, SUT, SLST, and SLHT), excessive flexion of the trunk (FST, SUT, SLST, and SLHT), reported pain (all tests), hip adduction (FST, SUT, SLST, and SLHT), hip elevation and pelvic rotation/inclination (FST), hip circumduction or reduced ankle dorsiflexion (SUT), Trendelenburg (SUT, SLST, and SLHT), Retrotrendelenburg (SUT, SLST, and SLHT), corkscrew movement, characterized by depression and posterior rotation of the pelvis (SUT, SLST, and SLHT), semi-flexion of the support limb (SUT, SLST, and SLHT), reduced peak height (SUT, SLST, and SLHT) and hip adduction (SUT, SLST, and SLHT)[11,13,22,24].

The tests used to assess the resistance of the trunk muscles

were the plank and the side planks[23,32]. In these tests, the depression of the pelvis higher than 2.54 cm, winged scapula, trunk rotation, lower trunk rotation, hip elevation, increased lordosis, and oscillations were observed[24].

Table 1 details the DMA risk classification criteria. Tests 2, 3, 4, and 6 are unilateral. Thus, the final score is calculated by using the lowest score between the right and left sides. Risk rating increases by one risk category if the following factors

are observed: prior anterior cruciate ligament (ACL) injury or high-risk sports practice (such as basketball, soccer, volleyball, football, cheerleading). Inter-rater reliability and intra-rater reliability movement patterns were observed on video. The only exception was the plank tests, in which in addition to the video analysis, the evaluator observed the participant to detect deviations that cannot be seen from the frontal plane, such as rotations of the pelvis or trunk and a winged scapula.
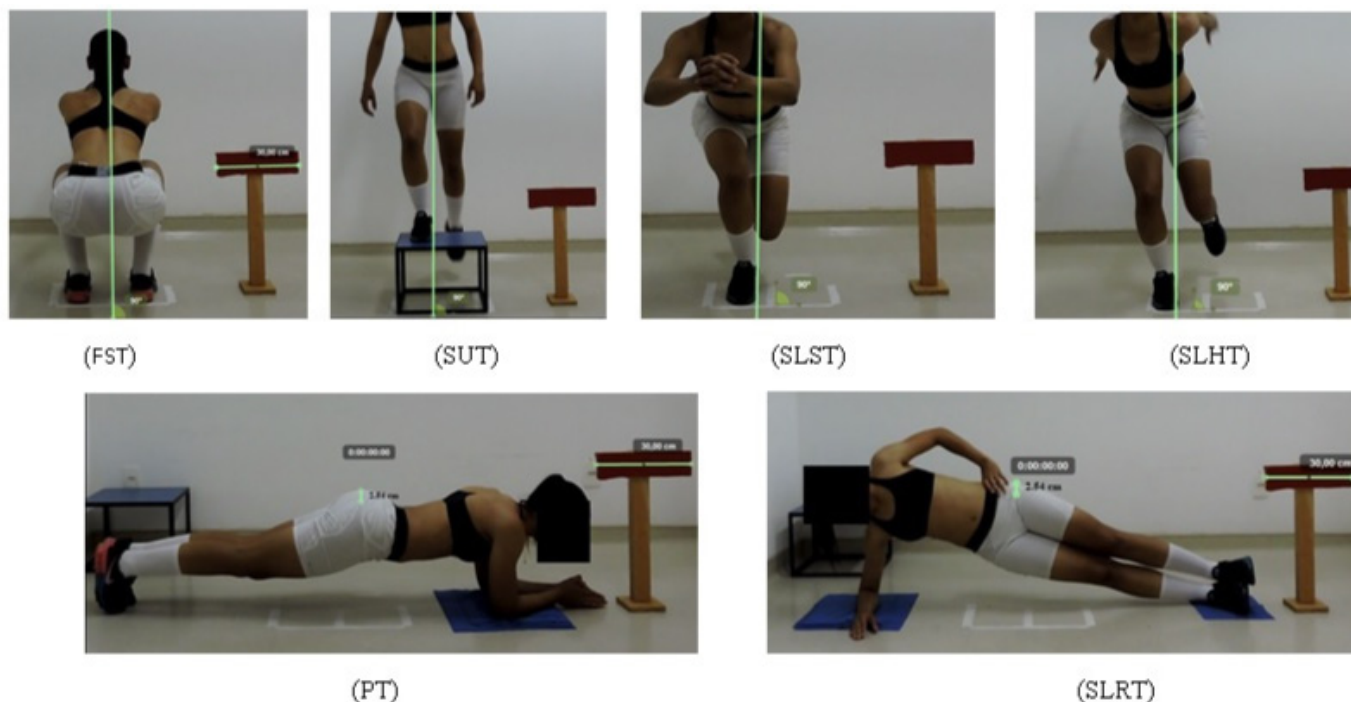

Figure 1 - The protocol of DMA. Full squat test (FST); Step-up test (SUT); Single-leg squat test (SLST); Single leg hop test (SLHT); Plank test (PT); Side plank test (SPRT).

Table 1 - Calculation of the final score DMA ™ (24).

Risk rating	Characteristics
4 - High	A collective score of 0-3 points on tests FST, SUT, and SLST + collective score of 3 points on tests PT, SPRT, and SPLT; and/or LOB in the tests FST and SUT;
3 - Moderate	A collective score of 4-5 points on tests FST, SUT, and SLST + collective score of 4 points or better on tests PT, SPRT, and SPLT; and/or LOB in two of the 3 test FST, SUT, or SLST;
2 - Medium	A collective score of 6-7 points on tests FST, SUT, and SLST + collective score of 5 points or better on tests PT, SPRT and SPLT; and/or LOB in one of the 3 test FST, SUT, or SLST;
1 - Minimum	A collective score of 8-9 points on tests FST, SU, T, and SLST + collective score of 6 points or better on tests PT, S P, T and SPLT; and/or LOB in one of the 3 test FST, S,UT or SLST;

FST = full squat test, SUT=step-up test; SLST=single leg squat test; PT=plank test; SPRT=side plank right test; SPLT=side plank left test; LOB = loss of balance.

Intra-rater agreement

Evaluator 1 did not attend the screening and was unaware of anamnesis data from the participants. He was responsible for filming the functional tests and analyzing the videos on the same day (phase: DMA1). To assess the intra-rater agreement, the videos were analyzed again after two months (phase: DMA2). The motion analysis procedure was done using the biomechanical analysis software Kinovea version 8.15.0.

Inter-rater agreement

The videos obtained during the first stage of evaluation of the participants (DMA1) were analyzed independently by a second examiner, who also did not know the history of the participants.

The results obtained from the analysis of movement patterns, as well as the risk rating, were compared by an independent statistician. Thus, it was possible to verify the inter-rater agreement. The participants had access to the results, which were delivered to the team's physiotherapist.

Statistical analysis

Initially, 4x4 tables were constructed in order to investigate the intra-and inter-rater agreement. The movement patterns were analyzed using seven tests from the DMA and were given a final score ranging from 0 (zero) to 21 (twenty-one) points (interval). As explained in Table 1, the classification ranged from a minimum, mean, moderate, or high risk of injury (ordinal). To assess the reliability of the DMA™, we used the intraclass correlation coefficient (ICC) preceded by the confidence intervals for Bland–Altman Limits of Agreement for the exclusion of outliers, as well as a typical error of measurement in regard to the number of points obtained. For intra-rater reliability, the ICC (3,1) was performed with a 2-way mixed-effects model and consistency. To evaluate the inter-rater reliability, the ICC (2,1) was performed with a 2-way random-effects model with absolute agreement[33]. To assess agreement DMA™ to classify the risk of injury as low, medium, high or moderate, the Kappa Index was used with linear weights (kw)[34,35]. The intra-and inter-rater agreement was verified by the number of consistent responses, i.e. the number of cases in which the result is the same between evaluators, excluding those by chance[31,34], according to equation (1):

$$kw = \frac{Po - Pa}{1 - Pa}$$

where k_w = Kappa with linear weights; Po = Proportion of the observed agreement; Pa = Proportion of chance agreement.

The presence of kw values equal to 1 (maximum value)

presents a perfect agreement between evaluators (Table 3). The levels of agreement were[34]: k_w =0 (absence of agreement); k_w = 0-0.19 (poor agreement); k_w = 0.20-0.39 (slight agreement); k_w =0.40-0.59 (moderate agreement); k_w=0.60-0.79 (substantial agreement); k_w=0.80-1.00 (almost perfect agreement)[34]. The analysis of kw, ICC and typical error of measurement were done, respectively, in StatsDirect software (v.3), SPSS 23.0 and Excel.

To compare the agreement between pairs of DMA scores, we developed a 4x4 table. Kappa analysis was done by comparing the scores of four DMA (high, moderate, medium, and low risk of injury). Finally, the prevalence and the bias ratio of the study were calculated.

The equation (2) was used to calculate the prevalence:

$$Prevalence = \frac{|a - d|}{n}$$

where a = number of positive concordant cases between the evaluators 1 and 2; d = number of negative concordance cases between the evaluators 1 and 2; n = sample size. Values close to 1.0 (one) indicate higher prevalence and values closer to 0 (zero) indicate low or absent prevalence.

The bias was calculated using the following equation (3):

$$Bias = \frac{b - c}{n}$$

where b = number of conflicting positive cases between evaluators 1 and 2; c = number of conflicting negative cases between evaluators 1 and 2; n = sample size. The closer to 1.0 (one) the greater the bias and the closer to 0 (zero) the smaller the bias.

It was considered as significant values of $P \leq 0.05$.

Results

A total of 38 athletes who play women's professional football were invited to participate in the study, but only 17 met the inclusion criteria and volunteered. Sample characteristics are shown in Table 2 and the evaluation results for each participant are detailed in Table 3. The number of concordant results for the study of the inter-and intra-rater agreement is, respectively, in Tables 4 and 5. Figures 2 and 3 show calculations of confidence intervals for Bland–Altman Limits of Agreement. The correlation values inter-and intra-rater by calculating the ICC, through comparison of scores ranging from 0 (zero) to 21 (twenty-one) of DMA are in Table 8. Comparing the scores of four DMA (high, moderate, medium, and minimum risk of injury), k values were low (Table 9). Therefore, a comparative table of the Kappa Index and DMA scores was constructed (Table 10), in order to verify which scores, show the lowest agreement that none of the participants had minimal risk. Furthermore, there was no significant difference

between the "medium" and "moderate" scores (k = 0.34 and P = 0.14) but the "high risk" was significantly different (k = 1.0; P = 0.000). The prevalence and bias values were respectively 0.82 and 0.06. There was no significant difference between the

"medium" and "moderate" scores (k = 0.34 and P = 0.14) but the "high" score was significantly different from the "medium" (k = 1.0; P = 0.000). The prevalence values and bias rate for the study were respectively 0.82 and 0.06.
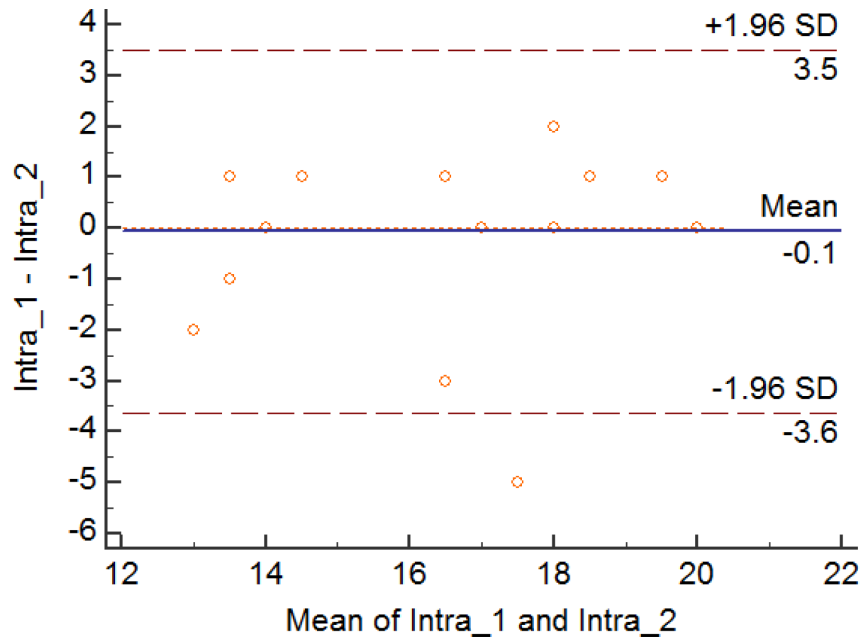

Figure 2 - Bland & Altman Confidence limits. One outlier was excluded. Reviewer Intra_1 = 1 at first; Reviewer Intra_2 = 1 in the second time, two months later.
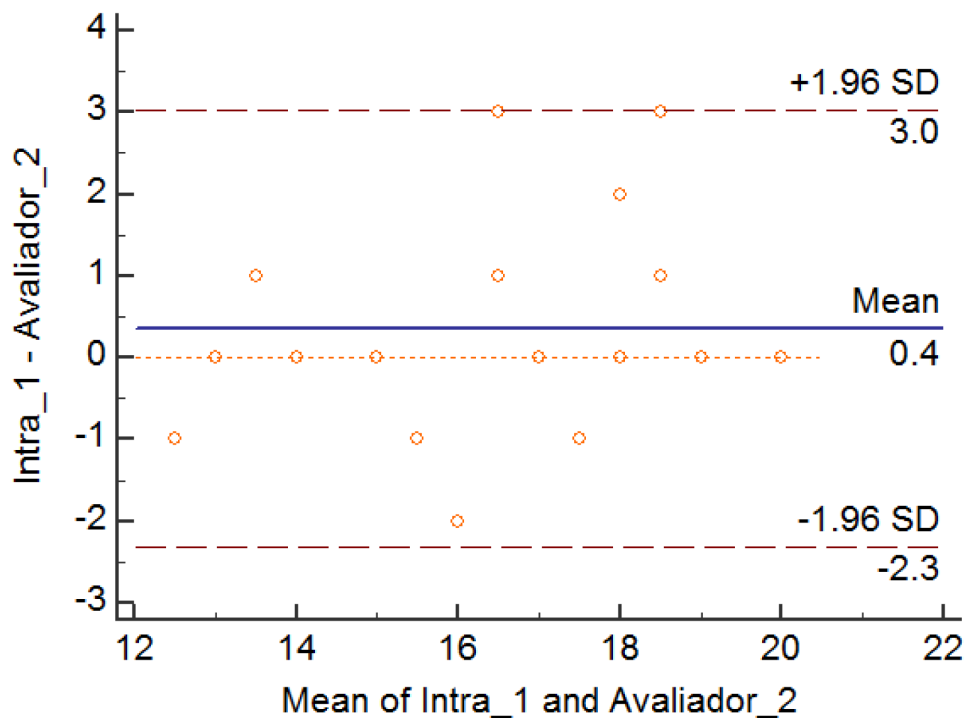

Figure 3 - Bland & Altman Confidence limits. No outliers. Reviewer Intra_1 = 1 at first; Reviewer Avaliador_2 = 2 at first.

Table 2 - Characteristics of participants included in the study

Variable		Result
Age years		26.7 ± 5.8
BMI (kg / m2)		23.2 ± 1.0
% C		25.6 ± 3.9
QNSO * - n (%)	Neck	4 (23.5)
	Shoulders	3 (17.6)
	Back (top)	2 (11.7)
	Elbows	0 (0)
	Wrists/hands	2 (11.7)
	Back (lower)	5 (29.4)
	Hips/thighs	12 (70.6)
	Knee	6 (35.3)
	Ankle/foot	9 (52.9)

BMI = body mass index; % F = percentage of body fat. QNSO = Nordic Musculoskeletal Questionnaire; * Each participant could indicate more than one painful area in QNSO. - n (%) = total number and percentage of participants with complaints of pain in the region.

Table 3 - Total points and injury risk rating evaluation of the 17 participants.

Participant	Evaluator 1 (phase 1)		Evaluator 1 (phase 2)		Evaluator 2	
	C	I	C	I	C	I
1	Moderate	17	Moderate	16	Medium	17
2	Moderate	17	Moderate	16	Medium	16
3	Moderate	14	High	14	High	14
4	Medium	20	Medium	20	Medium	17
5	Medium	19	Moderate	18	Medium	18
6	Moderate	19	Moderate	17	Moderate	17
7	Moderate	13	Moderate	14	Moderate	13
8	Moderate	15	Moderate	18	Moderate	17
9	Moderate	17	Moderate	17	Medium	18
10	Moderate	15	Medium	20	Moderate	16
11	Moderate	18	Medium	18	Medium	18
12	Medium	19	Moderate	17	Medium	19
13	Moderate	15	Moderate	14	Moderate	15
14	Medium	20	Medium	19	Medium	20
15	High	12	High	14	High	13
16	Moderate	18	Moderate	18	Moderate	15
17	Medium	14	Moderate	13	Moderate	13

Legend: C = categorical; I = interval.

Table 4 - 4x4 table (Intra-rater agreement)

		Evaluator 1 (phase 1)				
	Risk	Minimum	Medium	Moderate	High	Total
	Minimum*	0	-	-	-	0
	Medium	-	2	2	1	5
Evaluator 1 (phase 2)	Moderate	-	3	8	-	11
	High	-	-	-	1	1
	Total	0	5	10	2	17

* There were no participants with minimal risk.

Table 5 - Table 4x4 (inter-rater agreement)

		Evaluator 1				
	Risk	Minimum	Medium	Moderate	High	Total
	Minimum*	0	-	-	-	0
	Medium	-	4	4	1	9
Evaluator 2	Moderate	-	1	6		7
	High	-			1	1
	Total		5	10	2	17

* There were no participants with minimal risk.

Table 6 - ICC inter-and intra-rater reliability of the scores from 0 to 21 points of DMA (*n = 16 and n = 17 for intra and inter-rater evaluation, respectively)

Agreement	Number of points	A & B	ETM	ETM%	ICC	95%
					Intra-rater	
Phase 1	16.6 ± 2.6	3.6 -3.5	1.0	0.1	0.84[a] / (0.91)[b]	0.59 to 0.94[a]/0.74 to 0.97[b]
Phase 2	16.6 ± 2.3					
					Inter-raters	
Evaluator 1	16.6 ± 2.6	2.3 - 3.0	1.0	0.1	0.83[a] /(0.91)[b]	0.59 to 0.93[a] / 0.74 to 0.97[b]
Evaluator 2	16.2 ± 2.1					

ICC = intraclass correlation coefficient; TME = typical measurement error; There was a deletion of a participant, the result of which was out of the Bland-Altman confidence limits. [a] The estimator is the same, whether the interaction effect is present or not. [b] This estimate is computed assuming the interaction effect is absent because it is not estimable otherwise.

Table 7 - Inter-and intra-rater comparing the scores of four DMA (high, moderate, mild, and minimum) of 17 participants

Agreement	Number of concordant answers	Percentage of agreement	kw	95%	P
Intra-rater	11	64.70	0.46	0.05 to 0.86	0.02
Inter-raters	13	76.47	0.46	0.09 to 0.82	0.006

Kw = Kappa with linear weights. Prevalence = 0.82; bias ratio = 0.06.

Table 8 - Kappa Index comparative table among the DMA injury risk ratings

Risk	Minimum	Medium	Moderate	High
Minimum*	x	-	-	-
Medium	-	x	K = 0.35; P = 0.14 **	**K = 1.0; P = 0.00 ***
Moderate	-	K = 0.35; P = 0.14 **	x	K = 0.60; P = 0.06 #
High	-	**K = 1.0; P = 0.000 ***	K = 0.60; P = 0.06 #	x

* No participant was classified as minimal risk; ** n = 15; *** n = 5; # n = 8; K= Kappa index

Discussion

The results of the present study showed that there were 16.6 ± 2.6 and 16.6 ± 2.3 points in phase 1 and 2 (intra-rater agreement), respectively, resulting in a significant ICC of 0.84 (95% CI = 0.59 to 0.94). The evaluator 2 quantified 16.6 ± 2.6 points (inter-rater agreement). Thus, the mean value of the ICC inter-rater was 0.91 (95% CI = 0.74-0.97) (Table 6). Moreover, compared to the high-risk classification categories, medium, moderate and minimum risk presented intra-raters k_w of 0.46 (P = 0.02) and inter-rater values of k = 0.46 (P = 0.006) (Table 7). Therefore, there was a moderate agreement between values. If we consider the total sum of points (0-21), the DMA has relatively high reliability, but it shifts to moderate when the risk of injury is classified (Tables 6 and 7).

The DMA was originally developed to classify the risk of an individual developing musculoskeletal injuries[24]. The score is obtained after quantification of the points given in each of the seven tests and can vary from 0 (zero) to 21 (twenty-one) points. Since categorizing risk takes into account the total sum of the points in FST, SUT, and SLST (tests 1, 2, and 3) as the first criteria for individuals to be categorized into risk groups, the total sum of the score values yields classification into high risk (up to 3 points), moderate (4 to 5 points), medium (6 to 7 points) and minimum (8 to 9 points). Moreover, the total sum of points in the three plank tests (PT, SPRT, and SPLT) should be up to three points (for high risk), more than 4 points (for moderate), 5 points (for medium), and 6 points (for minimum). In addition to fulfilling the classification criteria, if the participant practiced high-risk sports or reported prior anterior cruciate ligament injury, the risk level was increased. Thus, an athlete could be given the same total score for the two different evaluators or even if evaluated at two different times but be classified into different risk categories.

This can be better exemplified using the results obtained in this study: of the 17 participants, six were classified in different categories between the first and second evaluations (intra-rater agreement), but the total score (between 0 and 21) of two participants was equal; two participants showed a difference of two points and two other participants showed a difference of two points or more. Between evaluators 1 and 2 (inter-rater agreement), seven participants were classified into different injury risk categories, but the total score (between 0 and 21) of two participants was equal; three had a difference of one point and 2 participants had a difference of two points. Thus, among 5 and 6 participants, whose points were the same or differed by only one point, respectively, were classified into different risk categories (Table 2). Considering, for example, participant number 1 received 16 points by the first evaluator and 17 points by the second evaluator. This difference has no impact on the ICC. However, the athlete was classified as moderate and high-risk respectively by evaluators 1 and 2. Thus, a difference of one point in tests 1, 2, and 3 can result in a classification into different categories, which is reflected in high ICC values and moderate kw values.

Analyzing the comparative table between the pairs of scores (Table 8), even though there is no significant difference between medium and moderate-risk categories (k = 0.34 and P = 0.14), the high-risk category is different from the moderate risk category (k = 1.0; P = 0.000). On the other hand, the classification interval showed good reliability intra-and inter-rater (Table 6).

In evaluation tests of movement patterns, low-reliability values are due to the subjectivity of the tests[25]. In such cases, evaluators use qualitative decisions to perform the classification test. Therefore, discrepancies may occur between evaluators or between multiple tests performed by the same evaluator. Although qualitative, functional tests are often helpful, especially when it comes to assessment protocols of individuals with previous injuries[36].

In the DMA, the main differences that may be potential sources of disagreement are a lateral deviation of the pelvis (FST)[37], presence and magnitude of hip adduction (FST, SLST, and SLHT)[25] and positioning the pelvis (PT, SPRT, and SPLT). Reliability values of hip adduction range from poor to substantial[38]. We did not find studies that evaluated the plank using the two-dimensional analysis. However, in these tests, a potential source of bias can arise from determining the starting position of the pelvis in a two-dimensional evaluation software. Furthermore, some deviations were analyzed live (presence of winged scapula, pelvis rotation or trunk oscillations, etc.) and, therefore, are potential sources of disagreement.

Using more than two categories of standard movement evaluation protocols were previously performed, as proposed by the DMA, and was mentioned as a threat to the reliability[39], because it can make the assessment more complex. The subjectivity of the tests may have contributed to a moderate agreement. A clear guidance regarding the evaluation protocol is also recommended[40]. Kappa index was used, since the DMA results is an ordinal variable, with more than two categories[35].

The interpretation of the Kappa index is not straightforward, as other factors can influence the magnitude of the coefficient or interpretation that can be placed in a given magnitude. Non-independent classifiers, prevalence, and bias can influence the magnitude of kappa[31]. The analysis was performed independently, so as not to influence the results. On the other hand, a disease in the population impacts agreement assessed by K in cases where the disease is very common or rare, making the professionals responsible for diagnosing the increasing trend of a positive or negative diagnosis[31]. In the present study, there was a prevalence of 0.82 (Table 7), which could contribute to a final higher agreement. On the other hand, there are situations where the evaluators disagree with the proportion of positive (or negative) cases and that is reflected in the difference between the cells of the 2x2 Table showing the disagreements in regard to positive and negative cases[41]. In this study, a bias ratio of 0.06 (Table 7) does not interfere with kappa values.

Limitations and strengths of the study

It is important to assess if the reliability of the method varied depending on the experience of the evaluators. This study was conducted using experienced evaluators. The evaluators of this study underwent formal training with DMA and had five

years of experience at the time of the study. Therefore, these results cannot be extrapolated to inexperienced individuals without formal training in the method. To increase the clinical applicability of the method, a next step could be the evaluation of live DMA without the 2D evaluation. However, new studies should be conducted to test the reliability. Moreover, it was not possible to assess the internal consistency reliability with a reassessment within two days, because the same evaluator remembers the results of the test. However, an evaluation of the stability of values was conducted.

Conclusion

In this study, DMA™ displayed excellent inter-and intra-rater reliability to measure the movement patterns and classify the risk of musculoskeletal injuries when the number of points obtained in the seven tests are considered. This reliability decreases to moderate when considering the risk of injury yielded by the method. Further studies are suggested to compare expert and novice evaluators. Therefore, caution should be exercised when generalizing the results obtained in this study. Further research is suggested in order to verify the validity of DMA to classify the risk of injury.

References

1. Taanila H, Suni J, Pihlajamäki H, Mattila VM, Ohrankämmen O, Vuorinen P, et al. Musculoskeletal disorders in physically active conscripts: a one-year follow-up study in the Finnish Defence Forces. BMC Musculoskeletal Disorders. [Online] 2009;10(1): 89. Available at: doi:10.1186/1471-2474-10-89

2. O'Brien J, Finch CF. The Implementation of Musculoskeletal Injury-Prevention Exercise Programmes in Team Ball Sports: A Systematic Review Employing the RE-AIM Framework. Sports Medicine. [Online] 2014; (October 2015): 1305–1318. Available at: doi:10.1007/s40279-014-0208-4

3. McCall A, Davison M, Andersen TE, Beasley I, Bizzini M, Dupont G, et al. Injury prevention strategies at the FIFA 2014 World Cup: perceptions and practices of the physicians from the 32 participating national teams. Br. J. Sports Med. [Online] 2015;49(9): 603–608. Available at: doi:10.1136/bjsports-2015-094747

4. Paterno M V, Taylor-Haas JA, Myer GD, Hewett TE. Prevention of Overuse Sports Injuries in the Young Athlete. Orthop Clin N Am. [Online] 2013;44(4): 553–564. Available at: doi:10.1016/j.ocl.2013.06.009.Prevention

5. Rosa B, Asperti A, Helito C, Demange M, Fernandes T, Hernandez A. Epidemiology of sports injuries on collegiate athletes at a single center. Acta Ortop Bras. 2014;22(6): 321–324.

6. Bueno AM, Pilgaard M, Hulme A, Forsberg P, Ramskov D, Damsted C, et al. Injury prevalence across sports : a descriptive analysis on a representative sample of the Danish population. Inj. Epidemiol. 2018;5(6): 1–8.

7. Onate JA, Everhart JS, Clifton DR, Best TM, Borchers JR, Chaudhari AMW. Physical Exam Risk Factors for Lower Extremity Injury in High School Athletes. Clin J Sport Med.

8. Rombaldi AJ, Silva MC Da, Barbosa MT, Pinto RC, Azevedo MR, Hallal PC, et al. Prevalência e fatores associados à ocorrência de lesões durante a prática de atividade física. Rev Bras Med Esporte. [Online] 2014;20(3): 190–194. Available at: doi:10.1590/1517-86922014200301709

9. Marar M, McIlvain NM, Fields SK, Comstock RD. Epidemiology of Concussions Among United States High School Athletes in 20 Sports. Am J Sports Med. [Online] 2012;40(4): 747–755. Available at: doi:10.1177/0363546511435626

10. Taanila H, Suni JH, Kannus P, Pihlajamäki H, Ruohola J-P, Viskari J, et al. Risk factors of acute and overuse musculoskeletal injuries among young conscripts: a population-based cohort study. Bmc Musculoskel Dis. [Online]; 2015;16(1): 104. Available at: doi:10.1186/s12891-015-0557-7

11. Lee J, KoH D, Kim K. The kinematics of the lower leg in the sagittal plane during downward squatting in persons with pronated feet. J. Phys. Ther. Sci. 2015;27: 285–287.

12. Kim S, Kwon O, Park K, Jeon I, Weon J. Lower Extremity Strength and the Range of Motion in Relation to Squat Depth. J. Hum. Kinet. [Online] 2015;45(1): 59–69. Available at: doi:10.1515/hukin-2015-0007

13. Powers CM. The Influence of Abnormal Hip Mechanics on Knee Injury: A Biomechanical Perspective. J Orthop Sports Phys Ther. [Online] 2010;40(2): 42–51. Available at: doi:10.2519/jospt.2010.3337

14. Clifton DR, Grooms DR, Onate JA. Overhead Deep Squat Performance Predicts Functional Movement Screen™ Score. Int J Sports Phys Ther. [Online] 2015;10(5): 622–627. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4595915&tool=pmcentrez&rendertype=abstract

15. Butler RJ, Contreras M, Burton LC, Plisky PJ, Goode A, Kiesel K. Modifiable risk factors predict injuries in firefighters during training academies. Work. [Online] 2013;46(1): 11–17. Available at: doi:10.3233/WOR-121545

16. Bushman T, Grier T, Canham-Chervak M, Anderson M, North W, Jones B. Pain on Functional Movement Screen Tests and Injury Risk. J Strength Cond Res. 2015;29(11): 65–70.

17. Zalai D, Panics G, Bobak P, Csáki I, Hamar P. Quality of functional movement patterns and injury examination in elite-level male professional football players. Acta Physiol Hung. [Online] 2015;102(1): 34–42. Available at: doi:10.1556/APhysiol.101.2014.010

18. Hotta T, Nishiguchi S, Fukutani N, Tashiro Y, Adaki D, Morino S, et al. Functional Movement Screen for predicting running injuries in 18- to 24-year-old competitive male runners. J Strength Cond Res. 2015;29(10): 2808–2815.

19. Lewis CL, Foch E, Luko MM, Loverro KL, Khuu A. Differences in Lower Extremity and Trunk Kinematics between Single Leg Squat and Step Down Tasks. Plos One. [Online] 2015;10(5): e0126258. Available at: doi:10.1371/journal.pone.0126258

20. Ross GB, Curran S. Quantifying frontal plane knee motion during single-limb squats: reliability and validity of 2-dimensional measures. Int J Sports Phys Ther. 2014;(7).

21. McLean SG, Walker K, Ford KR, Myer GD, Hewett TE, van den Bogert a J. Evaluation of a two-dimensional analysis method as a

screening and evaluation tool for anterior cruciate ligament injury. Br. J. Sports Med. [Online] 2005;39(6): 355–362. Available at doi:10.1136/bjsm.2005.018598

22. Augustsson J, Thomee R, Linden C, Folkesson M, Tranberg R, Karlsson J. Single-leg hop testing following fatiguing exercise: reliability and biomechanical analysis. Scand J Med Sci Sports?. [Online] 2006;16(2): 111–120. Available at: doi:10.1111/j.1600-0838.2005.00446.x

23. Tong TK, Wu S, Nie J. Sport-specific endurance plank test for the evaluation of global core muscle function. Phys Ther Sport. [Online] 2014;15(1): 58–63. Available at: doi:10.1016/j.ptsp.2013.03.003

24. Nessler TD, Haile E. Dynamic Movement Assessment: Prevent Injury and Enhance Performance Kindle Edition. Publiwide, USA, 2014. 2014; 2014.

25. McCunn R, Aus der Fünten K, Fullagar HHK, McKeown I, Meyer T. Reliability and Association with Injury of Movement Screens: A Critical Review. Sports Medicine. [Online] 2015; 1–19. Available at: doi:10.1007/s40279-015-0453-1

26. Kimbelin C, Wintersteins A. Validity and reliability of measurement instruments used in research Carole. Am J Health-Syst Pharm. [Online] 2008;65(1): 2276–2284. Available at: doi:10.2146/ajhp070364

27. Kasten R, Nevo B. Exploring the Relationship Between Interrater Correlations and Validity of Peer Ratings. Hum. Perform. [Online] 2008;21: 180–197. Available at: doi:10.1080/08959280801917792

28. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. [Online] 2011;64(1): 96–106. Available at: doi:10.1016/j.jclinepi.2010.03.002

29. Pinheiro FA, Tróccoli BT, Carvalho CV De. Validação do Questionário Nórdico de Sintomas Osteomusculares como medida de morbidade. Rev. Saúde Públ. [Online] 2002;36(3): 307–312. Available at: doi:10.1590/S0034-89102002000300008

30. Jackson AS, Pollock ML. Generalized equations for predicting body density of men. Br. J. Nutr. [Online] 1978;40(3): 497–504. Available at: doi:10.1079/BJN19780152

31. Sim J, Wright CC. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. Phys Ther. [Online] 2005;85(3): 257–268. Available at: doi:15733050

32. Zazulak BT, Hewett TE, Reeves NP, Goldberg B, Cholewicki J. Deficits in Neuromuscular Control of the Trunk Predict Knee Injury Risk: A Prospective Biomechanical-Epidemiologic Study. Am J Sport Med. [Online] 2007;35(7): 1123–1130. Available at: doi:10.1177/0363546507301585

33. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. [Online] Elsevier B.V.; 2016;15(2): 155–163. Available at: doi:10.1016/j.jcm.2016.02.012

34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. [Online] 1977;33(1): 159–174. Available at: doi:10.2307/2529310

35. Estrela C. Metodologia Científica: Ciência, Ensino, Pesquisa. Artes Médicas; 2018. 190 p.

36. Haitz K, Shultz R, Hodgins M, Matheson GO. Test-retest and interrater reliability of the functional lower extremity evaluation. J Orthop Sports Phys Ther. [Online] 2014;44(12): 947–954. Available at: doi:10.2519/jospt.2014.4809

37. Kilpikoski S, Airaksinen O. Interexaminer Reliability of Low Back Pain Assessment Using the McKenzie Method. 2002;27(8): 207–214.

38. Munro A, Herrington L, Carolan M. Reliability of 2-Dimensional Video Assessment of Frontal-Plane Dynamic Knee Valgus During Common Athletic Screening Tasks. J. Sport Rehabil. 2012;21: 7–11.

39. Chmielewski TL, Hodges SCSMJ, Horodyski MHSM, Bishop MD, Conrad CBP, Tillman MESM. Investigation of Clinician Agreement in Evaluating Movement Quality During Unilateral Lower Extremity Functional Tasks: A Comparison of 2 Rating Methods. J Orthop Sport Phys. 2007;37(3): 122–129. Available at: doi:10.2519/jospt.2007.2457

40. Ageberg E, Bennell KL, Hunt MA, Simic M, Roos EM, Creaby MW. Validity and inter-rater reliability of mediolateral knee motion observed during a single-limb mini squat. BMC Musculoskelet. Disord. [Online] England; 2010;11: 265. Available at: doi:10.1186/1471-2474-11-265

41. Byrt T, Bishop J, Carlin JB. Bias, prevalence, and kappa. J Clin Epidemiol. 1993;46(5):423–9.

Corresponding author

Priscila dos Santos Bunn
Avenida Brasil, 10.590, Penha, Rio de Janeiro, RJ, Brazil, CEP: 21.012-350.
Email: priscilabunn@yahoo.com.br