Journal of Musculoskeletal Pain

Title Page

Title: Inter- and intra-rater reliability of the Beighton Score compared to the Contompasis Score to assess generalised joint hypermobility

Authors:

Amy Vallis MSc, BSc, MCSP – Staff Grade Physiotherapist, East Coast Community Healthcare, Lowestoft, Norfolk.

Alison Wray MSc, BSc, MCSP – Staff Grade Physiotherapist, Norfolk and Norwich University Hospital, Norwich, Norfolk.

Toby O Smith PhD, MA, MSc, BSc, MCSP – University Lecturer, University of East Anglia, Norwich, Norfolk

Corresponding Author: Dr Toby Smith, Lecturer in Physiotherapy, School of Rehabilitation Sciences, Queen's Buildings, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ. Email: toby.smith@uea.ac.uk; Telephone: 01603 593087; Fax numbers: 01603 593166

Abstract

Objectives: Generalised Joint Hypermobility (GJH) is a common connective tissue disorder associated with a range of musculoskeletal complaints. An effective screening tool to assess GJH may influence our understanding and choice of management. Diagnosis is clinical, using tools such as the Beighton Hypermobility Score and the Contompasis Scoring System. The comparable reliability of these tools has not been previously reported. The aim of the present study was to compare the intra- and inter-rater reliability of the Beighton Score to the Contompasis Score to assess GJH.

Methods: This was an observational study assessing 36 pain-free participants; 27 females and nine males; aged 18 to 32 years. Participants were assessed in random order, by two researchers over two sessions to determine intra- and inter-rater analyses. Intraclass Correlation Coefficient (ICC) and weighted Kappa statistics were used to calculate the level of agreement.

Results: The intra- (ICC: 0.71-0.82) and inter- (ICC: 0.72-0.80) rater reliability of the Beighton Score was substantial to almost perfect. The Contompasis Score displayed substantial to almost perfect intra-rater (ICC: 0.73-0.82) reliability and moderate to substantial inter-rater (ICC: 0.58-0.62) reliability.

Conclusions: The present study provides an indication of the measurement capabilities of the Beighton and Contompasis Scores. The Beighton score appears to be superior compared to the Contompasis score particularly based on inter-rater reliability.

Keywords: Hypermobility, Beighton, Contompasis, Reliability

Introduction

Generalised Joint Hypermobility (GJH) was first recognised as a distinct pathology by Kirk and colleagues in 1967 [1]. It is a condition in which most of an individual's synovial joints have a range of motion (ROM) beyond their normal limits [2, 3]. It has been suggested that an abnormal ratio of Type III to Type I collagen results in the decreased tissue stiffness common in GJH, where thin and elastic Type III collagen becomes more prevalent within the soft tissue matrix [4]. Although GJH was once thought to be uncommon, a recent epidemiological study by Clinch and colleagues in 2011[5] reported that amongst a cohort of 6,022 children evaluated in the UK, the prevalence of GJH in girls and boys aged 14 years was 28% and 11%, respectively. The prevalence of GJH in the adult population has been estimated to be 18% to 25% [6,7].

GJH has been associated with a range of musculoskeletal complaints including, arthralgias, joint subluxations, joint dislocations and sprains, as well as early onset osteoarthritis [1, 4, 8, 9]. Due to GJH's ability to reduce joint stability, it has been linked with an increased risk of musculoskeletal injury and is proposed as a risk factor for injuries to the ankle, knee and shoulder joints [3, 10-14]. In its extremes, GJH is associated with hereditary connective tissue disorders such as Marfan's Syndrome and Ehlers-Danlos Syndrome with chronic pain and fatigue major determinants of reduced quality of life [14-17]. Due to the debilitating nature of these conditions a clear method of diagnosis is needed to ensure early and effective management [15].

Criteria for assessing GJH were first described by Carter and Wilkinson in 1964 [18] and modified by Beighton and Horan in 1969 [19], providing the Beighton Hypermobility Score, the diagnostic tool most commonly used today by rheumatologists, physiotherapist, orthopaedic surgeons and other neuro-musculoskeletal clinicians. This method is simple, requires no special equipment and takes less than a minute to complete [2]. It includes bilateral examination of fifth-finger extension, opposition of the thumb to forearm, elbow extension, knee extension and trunk flexion [19]. The tool produces an overall score between zero and nine; one point is given if the criterion is met, zero if it is not. There is no definitive agreement on the threshold for diagnosing GJH, however researchers and clinicians have generally use a score of four or five [4, 19]. The major limitation of the Beighton method is it being an 'all or nothing' test that gives no indication of the degree of hypermobility, merely an expression of wide spread laxity [8, 20].

McNerney and Johnston [21] developed a semi-quantitative modification of the Beighton Score known as the Contompasis Scoring System. This system ranges with scores from two to 70, with Contompasis scores of greater than 20 being cited as indicative of GJH [9]. It assesses each of the nine Beighton criteria, as well as hindfoot eversion at the calcaneus. Rather than providing a single positive or negative response, it gives a graded response between two and eight points for each criteria. Clinically, it is suggested that the Contompasis Score provides a more refined grading of joint hypermobility thereby providing a more accurate reflection on GJH within patient groups [20]. However, by providing a number of different responses it is more time consuming and may exhibit greater measurement variability, increasing the potential for error [22].

Given the importance of accurately diagnosing those individuals with GJH it is clinically relevant that the measurement capabilities of both these tools are fully understood. Previous evidence has examined the reliability of the Beighton Score [2, 23-25]. These have reported good to excellent intra- and inter-rater reliability (Intra-class Correlation Coefficient values: 0.84 to 0.96). However recurrent confounding variables within these study methods impact on these findings. Such limitations include: poorly controlling the time interval between measurements, goniometer measurement error, variation in verbal instructions and soft tissue warm-up before or during testing. Nonetheless, both the Beighton and Contompasis scores have been used in clinical and research practice [3,6,9,21]. It is therefore important that clinicians and researchers understand the psychometric properties of these scores when considering their future adoption. The present study considered these influencing factors, with an aim to minimize potential sources of error. No studies have previously assessed the reliability of the Contompasis Score.

The aim of this study was therefore to compare the intra- and inter-rater reliability of the Beighton Score to the Contompasis Score for the assessment of GJH. This was to be evaluated with a weighted Kappa value and Intra-class Correlation Coefficients (ICC).

Method

Recruitment

The study was undertaken within the School of Rehabilitation Sciences (RSC) at the University of East Anglia. All recruitment, consent and data collection was undertaken within the school. Posters were placed within the school to seek interested pre-registration physiotherapy or occupational therapy students to volunteer to participate in the study who met the following eligibility criteria.

Inclusion Criteria:

Individuals who have not experienced joint or muscle pain for the past three months.

Physiotherapy or Occupational Therapy students in the School of RSC, at the participating University.

Individuals who provided informed written consent.

Exclusion Criteria:

Individuals with self-reported joint pain (any part of the body) experienced over the past three months.

Thirty-six volunteers were recruited. All potentially eligible participants were asked to read a Participant Information Leaflet seven days prior to the initial data collection session. If eligible and willing to participate, all participants were asked to sign a Consent Form which was witnessed and counter-signed by a member of the research team (AV, AW, TS).

The cohort of 36 participants enrolled included 27 females and 9 males with a mean age of 22.7 years (range 18 to 32 years). A sample of 36 people was determined as optimal, based on two reasons. Firstly, previous research assessing the reliability of the Beighton Score [2] recruited 36 participants providing a statistically significant finding. This gave an indication that a sample of 36 may be sufficient to demonstrate a statistically significant correlation. Secondly, Fleiss in 1986 [26] recommended that a minimum of 15 to 20 subjects is required for estimating the reliability of a quantitative variable.

Prior to recruitment ethical approval was granted by the Faculty of Medicine and Health Sciences, Research Ethics Committee at the University of East Anglia, Norfolk (Reference: 2011-2013-26).

Data Collection

The assessing team consisted of two researchers (AV, AW) who, prior to data collection, familiarised themselves with the original publications detailing both scores [19, 21]. A teaching session was provided to both researchers prior to data collection to standardise the assessment of GJH and eliminate subjectivity as much as possible. This was led by the senior author (TS). Only once all three researchers reached agreement that each technique was standardised across the two researchers, was testing commenced.

All participants were examined on two occasions. This was undertaken at the same time of day, one week apart. There were two separate, curtained-off examination areas in the same room, one for each researcher. Each participant was firstly assessed by one researcher followed by the other. The order of examinations was randomised to avoid systematic bias.

The Beighton Score assesses hypermobility by asking the participant to stretch own limbs to end of range with the researcher observing this movement. The Contompasis Score requires the participant to perform the same movements, with the addition of calcaneal eversion. For the Contompasis Score, joint angles were measured using a standard goniometer with two-degree increments, following the guidelines provided by Norkin and White in 2003 [27].

A description of each score is presented in Table 1.

Data Analysis

Analysis was conducted using SPSS version 18.0 for Windows (IBM, New York, USA). Intra- and inter-rater reliability was determined using both ICC for continuous data and weighted Kappa statistics for categorical data. Through this, comparisons between each reviewer's first and second scores for each test were compared using the weighted Kappa for each individual item. Secondly a comparison for the first assessment of GHJ using the two assessment was made comparing Reviewer 1 to Reviewer 2's findings using the using the weighted Kappa for each individual item. Total scores for Beighton and Contompasis Score were assessed using the ICC for both intra- and inter-rater reliability assessments.

Data was presented as single measures ICC and weighted Kappa values with 95% confidence intervals (CI) and p-values. A p-value of less than 0.05 was considered statistically significant. Weighted Kappa and ICC values were interpreted using Landis and Koch's interpretation of agreement where: less than 0.20 indicates slight agreement, 0.21 to 0.40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 substantial and greater than 0.81 indicates almost perfect agreement [28].

Results

Beighton Score

Data on Beighton Score intra- and inter-rater reliability are summarised in Table 2. The results suggest that the Beighton Score presented with substantial agreement in both intra- and inter-rater reliability. Intra-rater reliability of the total score showed almost perfect agreement in Researcher 1 (ICC: 0.82; 95% CI: 0.67 to 0.90) and substantial agreement in Researcher 2 (ICC: 0.71; 95% CI: 0.50 to 0.84). Inter-rater reliability showed substantial agreement between researchers during Session 1 (ICC: 0.72; 95% CI: 0.51 to 0.84) and Session 2 (ICC: 0.80; 95% CI: 0.64 to 0.89).

Contompasis Score

Data on Contompasis Score intra- and inter-rater reliability are summarised in Table 3. Although the Contompasis Score presented with substantial intra-rater reliability, it presented with only moderate inter-rater reliability. Intra-rater reliability of the total score showed almost perfect agreement in Researcher 2 (ICC: 0.82; 95% CI: 0.67 to 0.90) and substantial agreement in Researcher 1 (ICC: 0.73; 95% CI: 0.53 to 0.85). Inter-rater reliability showed substantial agreement between researchers during Session 1 (ICC: 0.62; 95% CI: 0.37 to 0.79) and moderate agreement during Session 2 (ICC: 0.58; 95% CI 0.31 to 0.76).

Comparison of Beighton vs. Contompasis Score

Intra-rater reliability - The Beighton Score shows greatest intra-rater reliability when used to assess the spine (weighted Kappa: 0.87 to 100) and thumb (weighted Kappa: 0.85 to 0.92). It showed least intra-rater reliability when used to assess the finger (weighted Kappa: 0.00 to 0.59). The Contompasis Score showed greatest intra-rater reliability when used to assess the spine (weighted Kappa: 0.82 to 0.91) and for the assessment of the elbow by Researcher 2 (weighted Kappa: 1.00). It demonstrated least intra-rater reliability when used to assess the ankle (weighted Kappa: 0.22 to 0.54), and the elbow (weighted Kappa: 0.41) in Researcher 1.

Inter-rater reliability - The Beighton Score showed greatest inter-rater reliability when used to assess the spine (weighted Kappa: 0.81 to 0.93) and thumb (weighted Kappa: 0.85 to 0.94). It showed least inter-rater reliability when used to assess the elbow (weighted Kappa: 0.38 to 0.58) and finger (weighted Kappa: 0.07 to 0.52). The Contompasis Score demonstrated greatest inter-rater reliability when used to assess the spine (weighted Kappa: 0.89 to 0.92). It showed poorest inter-rater reliability in the other criteria, the least being evident in the elbow (weighted Kappa: 0.38 to 0.68), ankle (weighted Kappa: 0.09 to 0.29) and thumb assessments (weighted Kappa: 0.00 to 0.38).

Discussion

The present study provides the first report comparing the reliability of the Beighton and Contompasis Scores for assessing GJH. The findings suggest that both the Beighton Score and Contompasis Score possess good intra-rater reliability but moderate inter-rater reliability. The Contompasis Score demonstrated poorer inter-rater reliability compared to the Beighton Score with only moderate levels of agreement between the two reviewers.

Whilst there was little variation in the intra-rater reliability of the total Beighton Score compared to the total Contompasis Score, variation was shown between the researchers. Researcher 1 displayed greater intra-rater reliability in the Beighton Score, whilst Researcher 2 displayed greater intra-rater reliability in the Contompasis Score. This variance may be due to Researcher 2 having more experience in goniometric joint assessment. Previous research suggests that inexperienced therapists have lower intra- (ICC: 0.59 versus 0.72) and inter- (ICC: 0.12 versus 0.28) rater reliability when performing goniometric assessments, due to small systematic errors in alignment and identification of landmarks [29, 30]. It is not possible to determine whether this was a principle factor in this study given that we did not assess goniometry skills *per se*. Furthermore, the standardisation of the assessment method prior to testing was aimed to minimise this potential variability. Nonetheless, this is one hypothesis which may account for this variability.

Grahame [20] labelled the Beighton Score an 'all-or-nothing' test, citing this as a limitation as it provides no indication of the degree of hypermobility. The Contompasis Score sought to overcome this limitation by providing a semi-quantitative scoring system. However, by providing a variety of different responses rather than a dichotomous response opinion, it is more time consuming and has been suggested to exhibit greater measurement variability [20]. The present study supports this latter claim, as the Contompasis Score displayed poorer inter-rater reliability compared to the Beighton Score. This could again be explained by the requirement of goniometric joint assessment, as least agreement is seen in the criteria that use a goniometer, i.e. the elbow, knee and ankle. It could also be explained by the Contompasis Score utilising a graded response system, with minimal increments between each score. Taking the ankle criterion as an example, the responses include, two (0°-2° of eversion), four (3°-5°), five (6°-10°) six (11°-15°), seven (>15°). This gives little margin for error when using a tool which is based on observer interpretation, and is impractical based on research which has shown that inter-rater goniometric measures fall within 7° to 9° of each other [31, 32].

McNerney and Johnston [21] recognised the difficulty in finding a method of joint evaluation with no inherent error, as the values obtained will largely depend on the observer and method of assessment. They concluded that the aim must be to minimise the error as much as possible. Therefore, it could be suggested to only use researchers experienced in the use of goniometric joint assessment, and to follow a strictly controlled method. However, this fails to reflect the pragmatism required to generalise findings to clinical practice where multiple clinicians work with varying levels of experience. Consequently, further assessment of these tools by clinicians with difference experiences and skills-sets in joint assessment using goniometry, may be warranted to further explore this potential source of variability.

The Beighton Score displayed poor intra- and inter-rater reliability in the finger criteria. The inaccuracy of visual estimates of finger angles has been previously documented as 25% when compared to computer-based joint assessment [33]. Nonetheless the Contompasis Score displayed poorer intra- and inter-rater reliability compared to the Beighton score. Given this findings, it is suggested that this criterion of the Contompasis Score would require refinement if reliability is to be improved. Accordingly, this tool may benefit from more distinct categories, as it is difficult to accurately distinguish between responses.

The principle limitation of the present study was the use of a pain-free population. Although a subset of the cohort were ‘clinically’ asymptomatic but demonstrated GJH on assessment, future work should include a patient population, as it is within this demographic that the tools are intended for use and knowledge of how they perform in a clinical setting would be relevant to expand the evidence-base. Specifically, investigating the reliability of these tools with people who present with Ehlers-Danlos syndrome and fibromyalgia would be particularly valuable given the potential severity of joint hypermobility and soft tissue pain which these people experience. Furthermore, since a patient-population may present with greater clinical variability most notably in joint pain and hypermobility [34,35], such a study would need to be sufficiently powered to ensure rigorous data can be gained.

Accordingly, a cohort of 36 may be insufficient and recruitment of a larger sample size may be warranted. In such a symptomatic population, it would be appropriate to then assess how the classification of GJH compares for the Beighton to Contompasis Score. Using cut-off points for clinical GJH diagnosis, it would then be possible to determine the precision of the measurement through sensitivity, specificity and likelihood ratios, which could have great clinical value.

A second limitation which could not be controlled was that of potential participant variability. The protocol stipulated that a one-week interval was stipulated between the first and second testing to minimise the recall of both assessor and participants to the testing procedure. This was also stipulated to reduce the potential for physiological variability [36]. Furthermore strategies to minimise participant (and their behaviour) variability included undertaking assessments at the same time of day, and by asking participants to exercise in the same way 48 hours prior to the testing procedures to minimise both circadian variability and possible fatigue acting as confounders [37,38]. However it was not possible to control with certainty all circumstances within the human body such as pain/muscle ache, fatigue, strenuous exercise as well as perception of the testing procedure which may have had an influence on the findings of this study.

Conclusion

Our results show that the intra- and inter-rater reliability of the Beighton Score was substantial to almost perfect. In comparison the Contompasis Score displays substantial to almost perfect intra-rater reliability and moderate to substantial inter-rater reliability in a healthy pain-free population. In both instances, intra-rater reliability was greater than inter-rater reliability. Based on these findings, the Beighton Score appears superior to the Contompasis score.

Declarations

Ethical Approval: Ethical Approval was sought and gained from the Faculty of Medicine and Health Sciences Research Ethics Committee (Reference: 2011-2013-26)

Funding: No funds were received for the conduct of this study.

Conflicts of Interests: The authors report no conflicts of interest.

References

1. Kirk JA, Ansell BM, Bywaters EGL: The hypermobility syndrome: musculoskeletal complaints associated with generalized joint hypermobility. Ann Rheumat Dis 26: 419-425, 1967.
2. Boyle KL, Witt P, Riegger-Krugh C: Intrarater and interrater reliability of the Beighton and Horan joint mobility index. J Athlet Train 38: 281-285; 2003.
3. Pacey V, Nicholson LL, Adams RD, Munn J, Munns CF: Generalized joint hypermobility and risk of lower limb joint injury during sport: A systematic review with meta-analysis. Am J Sport Med 38: 1487-1497, 2003.
4. Russek LN: Hypermobility syndrome. Phys Therap 79: 591-599, 2003.
5. Clinch J, Deere K, Sayers A, Palmer S, Riddoch C, Tobias JH, Clark EM: Epidemiology of generalized joint laxity (hypermobility) in fourteen-year-old children from the UK: A population-based evaluation. Arthr Rheumat 63: 2819-2827, 2003.
6. Mulvey MR, Macfarlane GJ, Beasley M, Symmons DP, Lovell K, Keeley P, Woby S, McBeth J: Modest association of joint hypermobility with disabling and limiting musculoskeletal pain: results from a large-scale general population-based survey. Arthritis Care Res 65: 1325-1333, 2013.
7. Vaishya R, Hasija R: Joint hypermobility and anterior cruciate ligament injury. J Orthop Surg 21: 182-184, 2013.
8. Middleditch A: Management of the hypermobile adolescent. Hypermobility syndrome: recognition and management for physiotherapists. Edited by Keer R, Grahame R.. Butterworth Heinemann, London, 2003, 51-66.
9. Mishra MB, Ryan P, Atkinson P, Taylor H, Bell J, Calver D, Fogelman I, Child A, Jackson G, Chambers JB, Grahame R: Extra-articular features of benign joint hypermobility syndrome. Brit J Rheumat 35: 861-866, 1996.
10. Beynnon BD, Murphy DF, Alosa DM: Predictive factors for lateral ankle sprains: A literature review. J Athlet Train 37: 376-380, 2002.
11. Cameron KL, Duffey ML, DeBerardino TM, Stoneman PD, Jones CJ, Owens BD: Association of generalized joint hypermobility with a history of glenohumeral joint instability. J Athlet Train 45: 253-258, 2010.
12. Collinge R, Simmonds JV: Hypermobility, injury rate and rehabilitation in a professional football squad: A preliminary study. Phys Therap Sport 10: 91-96, 2009.
13. Nilsson C, Wykman A, Leanderson J: Spinal sagittal mobility and joint laxity in young ballet dancers: a comparative study between first-year students at the Swedish Ballet School and a control group. Knee Surg, Sport Traumatol, Arthrosc 1: 206-208, 2003.

14. Ramesh R, Von Arx O, Azzopardi T, Schranz PJ: The risk of anterior cruciate ligament rupture with generalised joint laxity. J Bone Joint Surg 87: 800-803, 2006.
15. Castori M, Camerota F, Celletti C, Grammatico P, Padua L. Quality of life in the classic and hypermobility types of Ehlers-Danlos syndrome. Ann Neurol 67: 145–146, 2010.
16. Celletti C, Castori M, Grammatico P, Camerota F: Evaluation of lower limb disability in joint hypermobility syndrome. Rheumatol Internat 32: 2577–2581, 2012.
17. Voermans NC, Knoop H: Both pain and fatigue are important possible determinants of disability in patients with the Ehlers-Danlos Syndrome hypermobility type. Disabil Rehabil 33: 706-707, 2010.
18. Carter C, Wilkinson J: Persistent joint laxity and congenital dislocation of the hip. J Bone Joint Surg 46: 40-45, 1964.
19. Beighton P, Horan F: Orthopaedic aspects of the Ehlers-Danlos Syndrome. J Bone Joint Surg 51: 444-453, 1969.
20. Grahame, R. (2003) Hypermobility and Hypermobility Syndrome. Hypermobility syndrome: recognition and management for physiotherapists. Edited by Keer R, Grahame R.. Butterworth Heinemann, London, 2003, 1-14.
21. McNerney JE, Johnston WB: Generalised ligamentous laxity, hallux abducto valgus and the first metatarsal cuneiform joint. J Am Podiatry Assoc 69: 69-82; 1979.
22. Grahame R, Pyeritz RE: The Marfan Syndrome: Joint and skin manifestations are prevalent and correlated. Brit J Rheumatol 34: 126-131, 1995.
23. Evans AM, Rome K, Peet L: The foot posture index, ankle lunge test, Beighton scale and the lower limb assessment score in healthy children: a reliability study. J Orofac Orthop 68: 342-352, 2007.
24. Hirsch C, Hirsch M, John MT, Bock JJ: Reliability of the Beighton Hypermobility Index to determinate the general joint laxity performed by dentists. J Foot Ankle Res 5: 1, 2012.
25. Juul-Kristensen B, Røgind H, Jensen DV, Remvig L: Inter-examiner reproducibility of tests and criteria for generalized joint hypermobility and benign joint hypermobility syndrome. Rheumatology 46: 1835-1841, 2007.
26. Fleiss JL: The design and analysis of clinical experiments. NY: Wiley and Sons, 1986.
27. Norkin CC, White DJ. Measurement of joint motion: A guide to goniometry. 3rd Edition. F. A. Davis Company, Philadelphia, 2003.
28. Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 33: 159-174, 1977.
29. Elveru RA, Rothstein JM, Lamb RL: Goniometric reliability in a clinical setting: Subtalar and ankle joint measurements. Phys Therap 68: 672-677, 1988.

30. Fish DR, Wingate L: Sources of goniometric error at the elbow. Phys Therap 65: 1666-1670, 1988.
31. Burr N, Pratt AL, Stott D: Inter-rater and intra-rater reliability when measuring interphalangeal joints: comparison between three hand-held goniometers. Physiotherapy 89: 641-652, 2003.
32. Ellis B, Bruton A: A study to compare the reliability of composite finger flexion with goniometry for measurement of range of motion in the hand. Clin Rehabil 16: 562-570, 2002.
33. Rose V, Nduka CC, Pereira JA, Pickford MA, Belcher HJCR: Visual estimation of finger angles: do we need goniometers. J Hand Surg 27: 382-384, 2002.
34. Sperotto F, Balzarin M, Parolin M, Monteforte N, Vittadello F, Zulian F: Joint hypermobility, growing pain and obesity are mutually exclusive as causes of musculoskeletal pain in schoolchildren. Clin Exp Rheumatol 32: 131-136, 2014.
35. Schmid S, Luder G, Mueller Mebes C, Stettler M, Stutz U, Ziswiler HR, Radlinger L: Neuromechanical gait adaptations in women with joint hypermobility--an exploratory study. Clin Biomech 28: 1020-1025, 2013.
36. Matheson LA, Duffy S, Maroof A, Gibbons R, Duffy C, Roth J: Intra- and inter-rater reliability of jumping mechanography muscle function assessments. J Musculoskelet Neuronal Interact 13: 480-486, 2013.
37. Coldwells A, Atkinson G, Reilly T: Sources of variation in back and leg dynamometry. Ergonomics 37: 79-86, 1994.
38. Domenech MA, Sizer PS, Dedrick GS, McGalliard MK, Brismee JM: The deep neck flexor endurance test: normative data scores in healthy adults. PM R 3: 105-110, 2011.

Table 1: Outline of the Beighton and Contompasis scoring criteria.

Criteria	Beighton	Contompasis
Thumb Opposition to forearm	1=Yes 0=No	2= Separated by 30-75° 4= Touches forearm 5= Digs into forearm 6= Pushed beyond forearm
5th Finger Dorsi-flexion	1=Yes (Beyond 90°) 0=No	2= Between 30-85° 4= Between 90-100° 5= Between 100-120° 6= Beyond 120°
Elbow Hyper-extension	1=Yes 0=No	2= Between 0-5° 4= Between 10-15° 5= Between 16-20° 6= Beyond 20°
Knee Hyper-extension	1=Yes 0=No	2= Between 0-5° 4= Between 10-15° 5= Between 16-20° 6= Beyond 20°
Spine Trunk flexion with feet together without bending the knees	1=Yes (Palms flat on floor) 0=No	2= No contact 4= Fingertips touching 5= Fingers touching 6= Palms flat 7=Wrists touching 8=Forearms touching
Ankle Degree of Calcaneal eversion	N/A	2= 0-2° 4= 3-5° 5= 6-10° 6= 11-15° 7= Beyond 15°

Table 2: Intra- and inter-rater statistical values for the Beighton Score

	Intra-Rater				Inter-Rater			
	Researcher 1		Researcher 2		Researcher 1		Researcher 2	
Total Score	0.82	(0.67-0.90)	0.71	(0.50-0.84)	0.72	(0.51-0.84)	0.80	(0.64-0.89)
1 (spine)	0.87	(< 0.001)	1.00	(< 0.001)	0.81	(< 0.001)	0.93	(< 0.001)
2 (knee, left)	0.80	(< 0.001)	0.57	(< 0.001)	0.75	(< 0.001)	0.75	(< 0.001)
3 (knee, right)	0.80	(< 0.001)	0.74	(< 0.001)	0.82	(< 0.001)	0.70	(< 0.001)
4 (elbow, left)	0.60	(< 0.001)	0.70	(< 0.001)	0.38	(0.023)	0.58	(< 0.001)
5 (elbow, right)	0.60	(< 0.001)	0.65	(< 0.001)	0.44	(0.009)	0.58	(< 0.001)
6 (thumb, left)	0.87	(< 0.001)	0.87	(< 0.001)	0.93	(< 0.001)	0.94	(< 0.001)
7 (thumb, right)	0.92	(< 0.001)	0.85	(< 0.001)	0.85	(< 0.001)	0.92	(< 0.001)
8 (finger, left)	0.51	(0.002)	0.00	(0.85)	0.37	(0.02)	0.16	(0.27)
9 (finger, right)	0.59	(< 0.001)	0.20	(0.21)	0.52	(0.002)	0.07	(0.62)

* total score (single measure ICC and 95% confidence interval) and the individual criteria (weighted Kappa statistic and p-value)

Table 3: Intra- and inter-rater statistical values for the Contompasis Score

	Intra-Rater				Inter-Rater			
	Researcher 1		**Researcher 2**		**Researcher 1**		**Researcher 2**	
Total Score	0.73	(0.53-0.85)	0.82	(0.67-0.90)	0.62	(0.37-0.79)	0.58	(0.31-0.76)
1 (spine)	0.82	(0.68-0.91)	0.91	(0.83-0.95)	0.89	(0.80-0.94)	0.92	(0.85-0.96)
2 (knee, left)	0.96	(0.93-0.98)	0.90	(0.82-0.95)	0.95	(0.91-0.98)	0.92	(0.84-0.96)
3 (knee, right)	0.54	(0.26-0.74)	0.77	(0.60-0.88)	0.55	(0.27-0.74)	0.45	(0.15-0.68)
4 (elbow, left)	0.64	(0.40-0.80)	0.75	(0.56-0.86)	0.68	(0.45-0.82)	0.41	(0.10-0.65)
5 (elbow, right)	0.41	(0.10-0.65)	1.00	(1.00-1.00)	0.32	(0.00-0.58)	0.38	(0.07-0.63)
6 (thumb, left)	0.41	(0.10-0.65)	1.00	(1.00-1.00)	0.32	(0.00-0.58)	0.38	(0.07-0.63)
7 (thumb, right)	0.80	(0.64-0.89)	0.49	(0.20-0.70)	0.13	(0.00-0.44)	0.00	(0.00-0.27)
8 (finger, left)	0.80	(0.64-0.89)	0.39	(0.08-0.64)	0.08	(0.00-0.39)	0.00	(0.00-0.27)
9 (finger, right)	0.92	(0.85-0.96)	0.98	(0.97-0.99)	0.87	(0.76-0.93)	0.96	(0.92-0.98)
10 (ankle, left)	0.40	(0.09-0.64)	0.22	(0.00-0.51)	0.11	(0.00-0.42)	0.09	(0.00-0.40)
11 (ankle, right)	0.54	(0.27-0.74)	0.28	(0.00-0.55)	0.25	(0.00-0.53)	0.29	(0.00-0.56)

* total score (single measure ICC and 95% confidence interval)