A Systematic Review of the Beighton Score Compared with Other Commonly Used Measurement Tools for Assessment and Identification of Generalised Joint Hypermobility (GJH)

Dr Malini Alexander
BSc, (Chemistry, Mathematics), MEBM, MBBS, Dip-MSc Rheumatology
University of South Wales April 2022

Abstract

This systematic review compares the validity and reliability of the Beighton Score to those of other commonly used scores for identifying generalised joint hypermobility (GJH). Inclusion criteria: English language, studies on humans, all types of study designs, publications in academic journals, publications from the year two thousand onwards, publications in print and theses. Exclusion criteria: studies not in English, studies measuring single joints only, studies published before the year 2000, cadaveric studies, papers with only abstracts available. An electronic literature search was undertaken of Pub Med/MEDLINE, Embase, Scopus, Cochrane Database, SPORT Discus, Pedro databases, followed by a manual search between August and November 2021. The final review included 73 papers. The PRISMA (2021) COSMIN (2010) guidelines and CASP (2019) criteria were used to evaluate methodological quality and bias. The Beighton Score's intra-rater and inter-rater reliability ranged between ICC 0.74-0.99 and ICC 0.72-0.98 respectively. The BS has reasonable intra-rater and inter-rater reliability, however validity cannot be accurately determined as incorporation bias was identified as a major issue in study methodology, not previously identified in the literature. Paucity of data prevented accurate assessment of other scoring systems. Urgent research is required to clarify these issues and compare the BS to other tests. No source of funding was received in in undertaking this review. The author has no conflict of interests to disclose. This review was not registered. This review has been conducted as closely as possible in accordance with the PRISMA guidelines for reporting systematic reviews.

1.0 Introduction

The subject of joint hypermobility is surprisingly complex. It is common, yet frequently overlooked and a subject of relevance to rheumatology, orthopaedic surgery, sports medicine, physiotherapy and other manual therapies, however a lack of knowledge about assessment, diagnosis and management remains a concerning issue in clinical practice.

Despite high prevalence of Generalised Joint Hypermobility (GJH) in the general population, challenges using the Beighton Score (BS) to identify GJH still exist. This leads to diagnostic problems in conditions where (GJH) is the primary feature such as Hereditable Disorder of Connective Tissue including Joint Hypermobility Spectrum Disorders and Ehlers Danlos Syndrome Disorders.

Failure to recognise GJH leads to poor patient outcomes (Kumar and Lenert, 2017). Therefore, accurate clinical assessment is of paramount importance.

Prevalence of GJH ranges between 2%-57% of the general population (Castori *et al*., 2017), but is significantly higher in African, Asian, younger individuals and females (Corten *et al*., 2020). A systematic review and meta-analysis of Global GJH prevalence in children and adolescents by Sobhani-Eragi *et al*., (2021) reported a value of 34.1% (95% CI: 33.3%–34.8%).

Joint hypermobility (JH) has been recognised since antiquity. It is depicted in dance-pose carvings on ancient temples in India (Bennink *et al*., 2013). It is described in relation to yoga poses in the Rg Veda and Yoga Sutras, written over 2000 years ago (Macdonell, 1917) (Werner, 1977). JH is honed as a necessary skill in martial arts of the far East including suo gugong an esoteric Kung Fu style requiring deliberate joint dislocations and contortionism (Thomas, 2016). Excessive joint hypermobility was documented as a state of pathology in Scythian tribes of the Eurasian Steppe by Ancient Greek physician Hippocrates in 400BCE (Brazzaventre *et al*., 2013).

Figure 1: Joint Hypermobility Depicted on the Rajarajeshvara temple in Tanjore Built circa 1000CE, Tamil Nadu, India (Bennink et al, 2013) and described in the Natya Shastra written approximately 500BCE (Lidova, 2014)

Rombaut *et al*., (2015) conducted a survey of 325 physiotherapists in Sweden, of whom 74.6% reported they had never conducted an assessment on a patient with Generalised Joint Hypermobility (GJH), or Hypermobile Ehlers Danlos Syndrome (H-EDS). Another survey by Palmer *et al*., (2016) surveyed 66 physiotherapists and reported the majority had not received any formal training in management of symptomatic joint hypermobility and had seen only 10, or fewer patients with GJH in the prior year. This discrepancy between prevalence and recognition in a clinical setting remains an ongoing issue for patients and healthcare professionals a like.

Pratt and Ball (2016) state joint range of motion (ROM) is the most common outcome measure used to evaluate the effect of treatment for musculoskeletal disease, yet its assessment remains controversial.

JH is defined as a joint capable of exceeding a normal ROM (Malek *et al*., 2021), however there is no Global consensus on what constitutes normal ROM and joint hypermobility classification itself lacks consensus (Armstrong, 2018). Use of a goniometer to assess ROM is recommended by Schlager *et al*., (2018) and the Ehlers Danlos Society (2017).

JH is not a diagnosis in itself, but can represent a feature of a systemic condition (Malek, *et al*., 2021). JH is considered a spectrum of disorders that can be classified according to Table 1 in Castori *et al*., (2017).

GJH presenting as widespread musculoskeletal pain, dislocation, tendon tears, muscle weakness, strains, or joint instability not associated with trauma, are clinical phenomena that raise clinical suspicion of symptomatic Joint Hypermobility Syndrome, Ehlers Danlos Syndrome and other conditions, associated with connective tissue weakness. Systemic disorders presenting with GJH include not only EDS, but Marfan's Syndrome and other Hereditable Disorders of Connective Tissue (HDCT) of which there are over 200 types (Tofts *et al.*, 2009). These include: Stickler Syndrome (Cattalini *et al*., 2015) and the Chromosome 2q37 Deletion Syndrome, *(Tofts et al*., 2009), several congenital myopathies (Voermans *et al*., 2009), Down's Syndrome (Foley and Killeen, 2019), sub-types of mucopolysacharoidoses (Morishita and Petty, 2011) achondroplasia (Raff and Byers, 1996) and Alport Syndrome (Jobling *et al*., 2014).

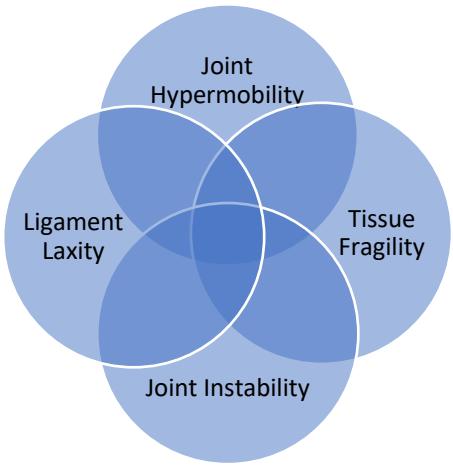
Terminology in current GJH literature is imprecise and ambiguous resulting in confusion in clinical practice. In this paper the terms used to describe hypermobility are taken from (Castori *et al*., 2017). These include the following acronyms:

- GJH – Generalised Joint Hypermobility
- JH- Joint hypermobility
- HSD – Hypermobile Spectrum Disorder
- H-EDS – Hypermobile Ehlers Danlos Syndrome
- BS – Beighton Score
- ROM – Range of motion
- HDTC – Hereditable Disorders of Connective Tissue

Use of terms such as ligament laxity, joint hypermobility, soft tissue fragility and joint instability create confusion for clinicians. Imprecise terminology creates challenges for researchers conducting literature searches.

In online health archives such as Pubmed there is no specific metadata search term for "joint hypermobility". It is classified under "Ligament Laxity" and "Joint Instability", however this is potentially misleading. These terms are not anatomically equivalent (Tofts *et al*., 2009) (Castori *et al*., 2017). Although closely related, these terms are not synonymous and are influenced by a range of anatomical and physiological mechanisms. They can be considered conceptually in Figure 3.

Figure 3. Four Closely Related, but Non-Equivalent Terms: Ligament Laxity, Joint Hypermobility, Joint Instability and Tissue Fragility



Ligament laxity contributes to joint hypermobility. Both joint instability and hypermobility are affected by other factors including variations in joint anatomy itself including socket depth, bone shape, muscle tone, strength, reduced proprioception and hormonal variations. (Knight, McCormack and Bird, 2012).

Factors contributing to joint hypermobility and joint instability are similar and are listed in table 1.

Table 1 Factors Contributing to Joint Hypermobility and Joint Instability (LibreTexts, 2020) (Kansas City Joint Experts, 2020) (Knight, McCormack and Bird, 2012) (Cypel, 2019)
Ligament laxity and collagen defects Strength and tension of joint capsule, ligaments and tendons Articular shape, size and depth Reduced muscle tone, strength and mass Reduced, or altered proprioception Previous trauma Disease processes that cause a decrease in synovial fluid Alterations in gravity and atmosphere Nutritional Deficiencies Hormonal variations including increased oestrogen and relaxin that increase ligament laxity Age and degenerative changes Patient's perception of pain at the time of examination

The BS is a modification of the 1964 Carter and Wilkinson scoring system developed by Beighton and Horan in 1969 to establish prevalence of GJH in an African population (Corten *et al*., 2020). It is the most commonly used method for identifying GJH.

There is no Global consensus, or gold standard cut-off score that defines joint hypermobility (Remvig *et al*., 2014). However, the Ehlers Danlos Society (2017) diagnostic criteria for H-EDS endorses the paper by Castori *et al*., (2017) recommending a range of age-adjusted cut off scores.

Some researchers do not recommend the BS as the primary tool for identification of GJH due to its limitations and prefer a diagnosis reliant on comprehensive full body assessment of joint range of motion (Malek *et al*., 2021) (Castori *et al*., 2017) (Grahame, 2004). Castori *et al*., (2017) and Tinkle (2020) report use of any single standardized measurement tool proves challenging.

Criticisms of using the BS to establish GJH include:

- Neither Carter and Wilkinson (1964) nor Beighton et al. (1973) provided any evidence-based justification for the selection of joints (Malek *et al.*, 2021)
- Only 4 joint sites are measured (Malek *et al*., 2021)
- Validity not adequately researched (Juul-Kristensen *et al*., 2017)
- Appropriateness for paediatric populations (Patel, *et al*., 2017)
- Inability to capture degree of hypermobility
- Developed as an epidemiological tool (Malek *et al*., 2021) (Beighton *et al*., 1973)
- Inclusion of ligament laxity measurement (Corten et al., 2020) (Castori *et al*., 2017)
- No consensus-based cut-off values (Remvig *et al*., 2014)
- Bias towards upper limb hypermobility, that might fail to capture lower limb hypermobility resulting in false negatives (Ferrari *et al*., 2005)
- Only assesses ROM in 2 dimensions. For some joints ROM occurs in multi-dimensions
- There are no consensus values for normal ROM (Soucie *et al*., 2011) (Moromizato *et al*., 2016) (Remvig *et al*., 2007) and the values chosen in the BS scoring system are based on tradition, rather than evidence

This literature review compares clinimetric properties of the Beighton Score against other scoring systems used to identify GJH, including the Hospital Del Mar/Bulbena Score, The Contompasis Score, The Rotès–Quérol Score, the Upper Limb Hypermobility Assessment Tool, The Lower Limb Hypermobility Assessment Tool, the Sasche Scale and several others including the 5pQ questionnaire.

This paper will answer the question: How does the BS compare with other widely-used methods including complete joint examination in identifying GJH?

2.0 Materials and Methods

2.1 Databases

A single researcher undertook this review as this paper is derived from a thesis for a Masters Degree in Rheumatology at the University of South Wales. There was a requirement for it to be individual work. The search strategy was conducted as closely as possible to the recommendations of Preferred Reporting Items for Systematic Reviews and Meta-Analyses PRISMA (2021) Guidelines for systematic reviews. The author did not use any software to assist in collection of the papers and used a purely manual approach.

Databases were searched during August to November 2021 and included Pub Med/ MEDLINE, Embase, Scopus, Cochrane, Database, SPORT Discus and Pedro. The following registers were searched over the same period: ANZCR, Clinicaltrials.gov, Centrewatch.com, ISRCTN, EU Clinical Trials Register.

A manual search of bibliographies and references from appropriate papers identified, Google and websites listed below were then manually searched to identify additional literature such as theses and publications unavailable in the databases listed above. Websites included: The Ehlers Danlos Society Website, The American College of Rheumatology, EULAR, UpToDate.com and Medscape. The University of South Wales library website was utilised to retrieve papers unavailable in research databases, or in online searches.

No additional literature was sourced by contacting authors or experts.

Studies were grouped according to experimental and quasi-experimental studies, narrative and systematic reviews and grey literature with expert opinion.

2.2 Inclusion Criteria:

English language
Humans
Studies on adults and children
All types of study designs
All publications in academic journals (including expert opinion)
Studies published from the year 2000 onwards
Publications in print
Theses

2.3 Exclusion Criteria:

Studies not in English, or not translated into English
Studies measuring a single joint only
Studies using the scoring system to measure a clinical presentation other than GJH, or H-EDS
Studies published before the year 2000
Grey literature
Cadaveric Studies
Papers for which only abstracts are available

2.3 Search Terms

Metadata search terms were not used as some databases do not distinguish between joint hypermobility, laxity and instability. No filters were used in the search. Each database was searched electronically for the following terms:

- Assessment of Joint Hypermobility
- Measurement of Joint Hypermobility
- Validity of {inset scoring system} in Joint Hypermobility
- Reliability of {inset scoring system} in Joint Hypermobility
- Reproducibility of {inset scoring system} in Joint Hypermobility
- Accuracy of {inset scoring system} in Joint Hypermobility

2.4 Outcome measures and measures of association

The clinimetric properties assessed were:
• Inter-rater reliability
• Intra-rater reliability
• Validity
• Sensitivity
• Specificity

All forms of measurement of strength of association reported in the literature were included.

2.5 Search Strategy and Data Extraction

No filters were used in the searches. Each search term was manually entered generating a number of papers. For each search the number was added up to gain the total number for each database. Each search was independently screened and papers whose titles obviously met the eligibility criteria, or exclusion criteria were saved into desktop folders, or excluded from the search accordingly.

This process generated a very high number of duplicates each time an individual database was searched. Once the search was complete, the collected papers were then screened more closely to assess eligibility and exclusion criteria and duplicates were removed and deleted.

This resulted in a final list of papers for the electronic search component. For the manual search, narrative and systematic reviews and papers from the EDS Society website were used to search bibliographies of relevant studies. These papers were searched for in Google and, or sourced from the University of South Wales Library and downloaded into folders. Searching in the University library catalogue and Google generated additional papers of interest that were screened for inclusion and exclusion criteria. Duplicates were removed.

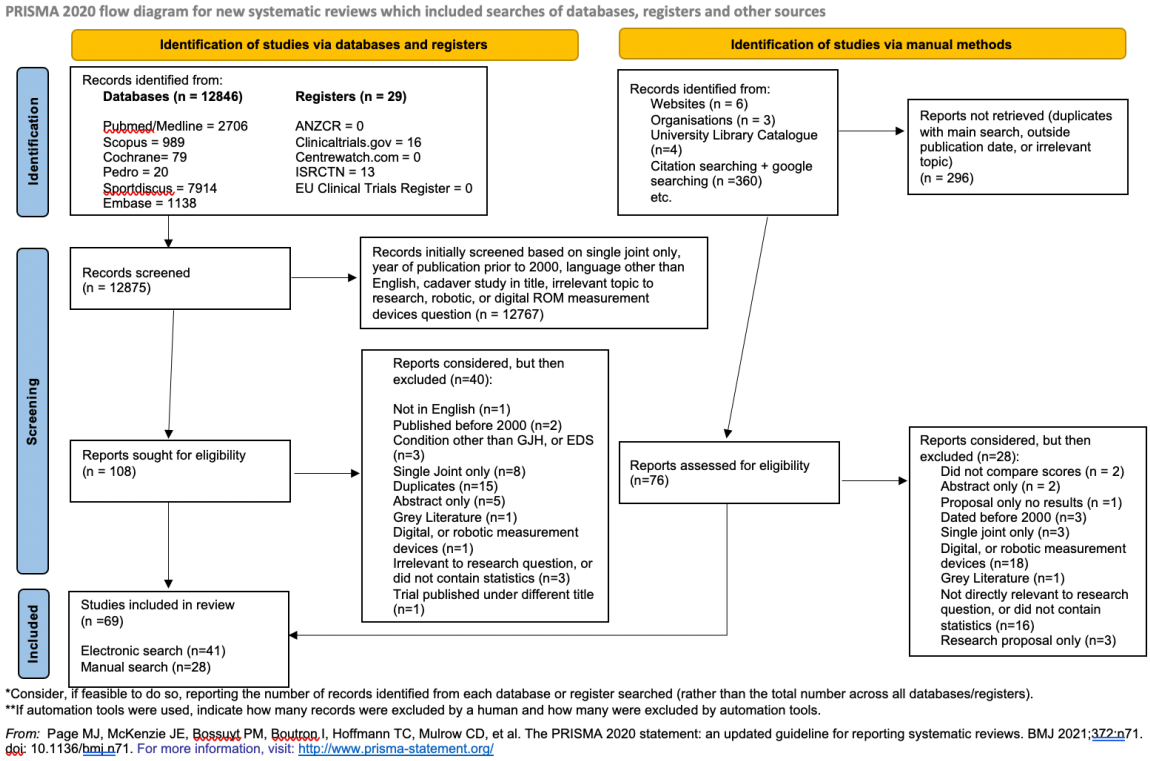
Papers not meeting inclusion criteria, or that met exclusion criteria were deleted. This resulted in the final selection of review papers. To access each paper, if it was not available in the database where it was originally cited, then a record was made and upon completion of the review these papers were searched for in other sources such as Research Gate, Academia.edu, Deepdyve, the University of South Wales Library databases, Google Scholar and Google. If free-access full text papers could not be sourced via this process, it was assumed only the abstract was available and such studies were excluded in accordance with exclusion criteria.

This review did not cover emails, other private correspondence, or letters and errata. There are no regulatory reviews relevant to GJH as far as the author is aware.

A number of measures were reported in the literature for assessment of validity and reliability. All measures were included in the results.
GJH has been considered as a diagnosis and therefore the Critical Appraisal Skills Programme (2019) (CASP) for diagnostic tests was used to assess quality of papers along with The Joanna Briggs Levels of Evidence for Diagnosis (2013) to rate methodological quality of studies. Additionally, use of a goniometer, or other assessment tools was included as part of the review as it is regarded as best practice when assessing ROM (Pratt and Ball, 2016) (Shamsi *et al*., 2019) (Mejia-Hernandez *et al*., 2018).

A note was made on whether studies referenced recommended Consensus-based Standards for selection of health Measurement Instruments (COSMIN, 2010) Guidelines, or any other similar standards. Systematic and narrative reviews were assessed against the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), 2021 guidelines (Page *et al*., 2021). These were summarised in a separate table included in Supplementary Material C. A high degree of heterogeneity in results was anticipated, therefore a meta-analysis was not conducted.

Figure 2: PRISMA FLOW DIAGRAM



This review has been conducted as closely as possible in accordance with the PRISMA Statement: An Updated Guideline for following a Systematic Review (Page, *et al*., 2021) as well as the PRISMA S Statement an Extension to the Prisma Statement for Reporting Literature Searches in Systematic Reviews (Rethlefsen *et al*., 2021). This information is available in supplemental material F.

3.0 Results

The final review consisted of 73 papers. 56 papers of experimental and quasi-experimental studies, 3 systematic reviews, 9 narrative reviews, and 5 grey literature and expert opinion. Selected papers from initial screening, but subsequently excluded from the final analysis were recorded and are available from the authors upon request in supplementary material B.

Results are summarised in tables. Table 3.1 provides statistical results of reliability and validity in papers that directly assessed scoring systems. Table 3.1 includes colour coded information on goniometer use, cohort and patient characteristics. Additional information includes whether assessors were blinded and whether any studies discussed their protocols with reference to the COSMIN (2010), or other relevant guidelines.

The Joanna Briggs Level of Evidence (2013) for each study is reported. The majority were rated as a Level 2b-3b. Many studies did not report whether consecutive patients were assessed. If not reported it was assumed they were not consecutive patient studies.

If a paper did not explicitly state goniometer use, it was assumed none was. If a paper did not explicitly state whether blinding of researchers, or participants was undertaken, it was assumed no blinding occurred. Blinding in the context of these research scores is with regards to participants knowing which scoring system is being used, or whether examiners were aware of each other's scoring points, or what patients had been scored as previously.

Table 3.2 is a summary of the narrative and systematic review findings. Table 3.3 is table of expert opinion, or grey literature sources included in the study.

The tables where methodological quality was assessed according to the CASP (2019) diagnostic tests and the PRISMA (2021) checklist for reporting Systematic Reviews are included in Supplementary Material C and D respectively.

Table 3.1 Summary of Reliability and Validity Statistics For Quasi-Experimental and Experimental Studies

Key to Table interpretation

Paediatric = blue
Pregnant = pink
Ballet or athlete study cohorts = yellow
EDS or hypermobile cohorts = green
Healthy adult cohorts, adult cohorts with other conditions = clear

Goniometer use = purple (Not necessary if study used only questionnaire)
Blinded = grey

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics			Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
			Inter-rater reliability	Intra-rater reliability	Validity		
Ahlqvist *et al.*, (2020)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**5pQ** **N/A**	N/A	N/A	N/A	Women with GJH also had higher odds of PGP during the entire pregnancy (adjusted odds ratio (aOR) 1.27: 95% CI 1.11–1.47) and in trimester 1 (aOR 1.54: 95% CI 1.20–1.96), but the associations were not statistically significant in trimester 2 (aOR 1.24: 95% CI 0.82–1.88) or trimester 3 (aOR 1.20: 95% CI 0.99–1.45). The odds of PGP in pregnancy increased with increasing numbers of positive answers to the 5PQ (p for linear trend < 0.001) for the entire pregnancy and in trimester 1 (p for linear trend < 0.001), but not in trimesters 2 or 3 (p = 0.13 and p = 0.06, respectively	**N**
	2455 Swedish-speaking pregnant women were consecutively recruited at their first visit for registration in the national antenatal screening programme, mean age 29 years (SD 5 years) (Sweden)	**Pelvic girdle pain during pregnancy**					**N**
Ahn *et al.*, (2019)	**RCT** **JBI LEVEL 1B**	**KRSP**	Inter-rater reliability ICC 0.846(95% CI 0.686-0.931) to 0.986 (95% CI0.972-0.994)	N/A	N/A	These results indicate that the Korean protocol can be the reference standard for measuring ROM in clinical settings as an alternative to goniometers	N
	91 healthy participants aged >65 31 males 60 females aged ≥65 years (Korea)	**ROM (using goniometer)**	Inter-rater reliability ICC range 0.643 (95% CI 0.486–0.783) to -0.078 (95% CI -0.296–0.494)				**N**
Antonio and Magalhaes (2018)	**Cross Sectional Survey** **JBI LEVEL 2B**	**BS** **SF36** **N**	N/A	N/A	Cut off ≥ 4 Prevalence of GJH 26.8%	Rho correlations between BS and SF-36 0.1 to 0.3	N
	388 healthy volunteers 299 females 89 males Age range 18 -25 years medical and physiotherapy courses (Brazil)						N
Aslan *et al.*, (2006)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **Y**	Inter-rater reliability ICC = 0.82	Intra-rater reliability 0.92	N/A	BS is a good score for identifying GJH in healthy subjects	N
	A total of 72 undergraduate physical therapy students, aged 18 to 25 (29 females and 43 males) (Age range female 20.2±0.9 male 20.5±1.5) (Poland)						Y
Armstrong (2018)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **Brighton Critera** **Y**	N/A	Intrarater reliability ICC of 0.992 (95% CI 0.979-0.997)	Removal of lumbar flexion from the BS resulted in no change in "not hypermobile" (NH) scores across the three classifications in male rugby and an increase of 5% (n = 2) in male controls (BE and SB). In netball players, the B classification (0-2) increased by 9.8% (n = 6)	in comparison to the BS and BE classification increase of 1.6% (n = 1) while female rugby remained similar at 3% (n = 2) (BE and SB) and 8% (n = 5) (B). Female dancers demonstrated large changes in the B classification 'moderately hypermobile" (MH) (3-4) with an increase of 33.3% (n = 14) in contrast to a decrease -4.8% (n = -2) and increase of 11.9% (n = 5) respectively in the BE (≥4) and SB "hypermobile" (4-6) classifications. This highlights classification system variation and influence of lumbar flexion inclusion	N
	Female and male rugby players, female netball players, female dancers and male and female age matched controls (UK)						N

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Armstrong and Greig (2018)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** Sixty-five female rugby players, 38 male rugby players, 61 netball players, 42 female dancers, 40 male controls and 40 female controls (UK)	**BS** **Y**	N/A	Intra-rater reliability ICC 0.992 (CI 0.979-0.997)	N/A	Significant differences existed for group and gender analysis at the left and right 5th metacarpophalangeal joints, left and right thumb, left and right elbow and lumbar spine ($p < 0.001$). Lumbar flexion demonstrated significant x2 values and large effect sizes for all groups. This effect size was reduced to a moderate effect size when male against female analysis was performed and joint hypermobility was greater in females in comparison to males	N N
Bale *et al*., (2014)	**Cross Sectional Cohort Study** **JBI LEVEL 3B** 120 Children aged 5–16 years old were recruited from secondary paediatric care	**BS** **Brighton Criteria** **N**			(28.3%) were classified as having JHS using the Brighton criteria. Of these 58.8% were female. Of the 34, 28 (82.4%) met both major criteria one of which is Beighton score >4. The commonest site of reported pain was in the muscles of the legs (27.5%). 64 (53.3%) children met at least one minor criteria however of these, 28 (43.4%) scored for arthralgia in 1–3 joints.	The other 6 met classification due to one major criteria (Beighton >4) with 2 minor criteria (2 (5.8%) due to arthralgia and skin extensibility, 1 (2.9%) due to arthralgia and dislocations, 1(2.9%) due to arthralgia and dislocations, 1(2.9%) due to skin extensibility and eye signs (myopia), 1 (2.9%) due to skin extensibility and dislocations and 1 (2.9%) due to arthralgia, skin extensibility and multiple dislocations). Of the children meeting criteria for JHS, 12 (33.3%) had a family history. A further 6 cases of JHS could be classified by the Brighton criteria if their family history was confirmed The commonest site of reported pain was in the muscles of the legs (27.5%). 64 (53.3%) children met at least one minor criteria however of these, 28 (43.4%) scored for arthralgia in 1–3 joints.	N N
Ballenger, Moore-Clingenpeel and Oberle (2021)	**Case Control Study** **JBI LEVEL 3B** 50 paediatric participants average age 16.1, Female 26 males 24 recruited from the same paediatric rheumatology clinic as well as a paediatric ophthalmology clinic and a paediatric dermatology clinic at an academic children's hospital. Ethics approval was obtained from the Nationwide Children's Institutional Review Board (USA)	**BS** **Ultrasound findings** **Y**				H + P knees were more likely to have positive findings noted on MSUS (94% vs. 70% of H-P and 74% of NP knees, p = 0.043). Patellar tendon hyperemia was more common in H + P knees (52%, vs. 19% among H-P and 23% among NP, p = 0.025). Participants who reported taking scheduled non-steroidal anti-inflammatory drugs (NSAIDs) had an increased risk of synovial effusion (RR = 1.83, 95% CI = 1.07–2.30, p = 0.026) and a trend towards increased risk of a higher synovial effusion/ hypertrophy quantitative score (RR = 1.77, 95% CI = 0.92–3.38, p = 0.086)	N Y
Bevilaqua, Maillard and Ferrari (2019)	**Retrospective Cohort Study** **JBI LEVEL 2B** 30 paediatric patients eage from 3 to 12 years (mean = 7.02, ±2.47 SD). Of the 30 subjects assessed, 21 (30%) were classified as hypermobile. Overall, the prevalence of JH in this study was higher in boys than in girls (UK)	**Hospital Del Mar Criteria** **N**	Inter-rater reliability Kappa = 0.81	N/A	(70%) were classified as hypermobile, 11 boys (52.3%) and 10 girls (47.6%). Nine children (30%) were classified as a non-hypermobile, three boys (33.3%) and six girls (66.6%)	The Hospital Del Mar Criteria has good inter-rater reliability. Cut off scores need to be established	N Y
Boyle, Witt and Rieger Krugh (2003)	**Cross Sectional Cohort Study** **JBI LEVEL 2A** Students from the Chapel Hill High School soccer team and the University of North Carolina at Chapel Hill physical therapy program. Forty-two (intrarater) and 36 (interrater) female volunteers, aged 15 to 45 years (mean age was 25.4 6 4.2 years) (USA)	**BS** **Y**	Inter-rater reliability Spearman rho = 0.87 ($p<0.0001$)	Intra-rater reliability Spearman rho = 0.86 ($p<0.0001$)	N/A		N Y

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Bulbena *et al.*, (2014)	**Cross Sectional Cohort Study** **JB LEVEL 2B**	**Screening Questionaire to Detect Hypermobility** **SQCH**	Inter-rater reliability ICC= 0.961 CI (0.922-0.980)	N/A	GJH 44.3% PPV 92.79% NPV 46.22% "Wrongly classified 2.48%" The area under the ROC curve of the 7 self-assessed criteria (SQ-CH) is 0.861.Sensitivity and specificity percentages for the 2/3 cut-off point in relation to the Hospital del Mar criteria are 78.0% and 75.8% respectively, obtaining a percentage of 65.75% correctly classified as positive and 85.2% correctly classified as negative. Likelihood ratios are positive and significant for each item according to the cut-off point in relation to the total score in SQ-CH	The cut-off point proposed for this sample population is 2/3 (from 0-2 lack of JHS and from 3-7 JHS) with a sensitivity of 0.78 and a specificity of 0.24.	N
		Hospital Del Mar **N**	N/A	N/A	For presence of GJH 37.3% (rho = 0.745; p<0.001) Sensitivity 78% specificity 75.8% 2/3 cut-off point True positives 65.75% True negatives 85.2%		
	158 patients between 18 and 60 years old were consecutively recruited from an anxiety outpatient unit belonging to a general university hospital. 142 females, 16 males (Spain)	**(5pQ)**	N/A	N/A	For presence of GJH (rho = 0.857; p<0.001)		Y
Calhil *et al.*, (2021)	**Cross sectional** **JBI LEVEL 2B**	**BS** **N**	N/A	N/A	Prevalence of GJH = 13.3%	Marshal test good for ruling out GJH, but must be used with caution.	N
	204 healthy patients 2-18 years old presenting to a general paediatric orthopaedic clinic (111 female, 93 male) average age 10.7 years (USA)	**Marshall test**	N/A	N/A	PPV = 34% NPV = 99% Sensitivity =99% Specificity =67% Positive likelihood ratio = 3.3		N
Chan *et al.*,(2018)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS**	N/A	N/A	Cut off point ≥ 5/9) GJH=72%	72% of dancers have GJH according to BS, while 38% and 42% met have GJH when using LLAS cut-point on the left and right respectively. Between the scoring methods of BS and LLAS there is a disagreement of 48% and 46% respectively, with Beighton classifying more participants as having GJH. A higher Beighton cut-point, e.g. 6/9, to identify true GJH amongst dancers maybe warranted.	N
	85 dancers from two dance institutions 74 female 26 male Mean age 21.2 (Australia)	**LLAS**			Cut off point ≥ 7/12 GJH= 38% and 42% met the LLAS cut-point on the left and right respectively		
		Villefranche vs Brighton Score	54% disagreement	N/A	The Villefranche identified more dancers with JHS/EDS-HT than the Brighton (84% vs 31%, p < 0.001)		N
Cherni *et al.*, (2019)	**Case Control Study** **JBI LEVEL 3B**	**BS** **Extensometer**	N/A	N/A	Laxity of the metacarpophalangeal joint increased by 11% from the first to the second trimester of pregnancy and stabilized until delivery. The Beighton score was significantly higher in the second trimester of pregnancy (p < 0.05)	Moderate correlation was observed between the results given by the extensometer and the Beighton score in both the cases and the control group at first trimester (r = 0.60, p < 0.05) but none was found for the two hip and lumbar flexibility tests The chosen clinical tests don't seem appropriate to be used alone in pregnant women	N
	33 pregnant females over the age of 18, BMI <40 with the control group was included to have a baseline value of the laxity before pregnancy (Canada)						N

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Cooper *et al.*, (2018)	**Cross sectional Cohort Study** **JBI LEVEL 2B** 50 volunteer participants 22 male 28 female Median age 49 (UK)	**Self Reported line drawings** **Modified BS cut off >=4** **Y**	Inter-rater reliability k= 1.00; (95% CI 1.00-1.00) Self-reported participant-repeatability k =0.91 (95% CI 0.74-1.00)	N/A	Sensitivity of 0.87 (95% CI 0.81, 0.91) Specificity of 0.99 (95% CI 0.98, 1.00)	The self-reported instrument provides a valid and reliable assessment of the presence of generalised joint hypermobility and may have practical use in epidemiological studies	N Y
Cypel (2019)	**Retrospective Case Control Study** **JBI LEVEL 3B** 110 cases of known EDS. Mean age 30 years (range,6–69 years) overall, 32.5 years (range, 6–66 years) in the 87 females,and 22 years (range, 6–69 years) in the 23 males 100 healthy controls Mean age 48 years 50 females 50 males (France)	**Glenohumeral abduction** **+ BS** **Y**	N/A	N/A	Sensitivity 96.4% Specificity 92.5% glenohumeral hypermobility	Increased gleno-humeral abduction may be sufficient to demonstrate joint hypermobilityand to suggest EDS in patients whose personal and family history is consistent with this diagnosis	N N
Czaprowski et al (2015)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 249 paediatric patients (136 females and 113 males) Average age 11.7/11.8 SD 0.7/0.8 (Poland)	**BS** **Straight leg raise (SLR)** **Thomas test for one- (O-JHF)** **two-joint (T-JHF) hip flexors** **fingertip-to-floor (FTF) and lateral trunk flexion (LTF) tests** **N**			A positive result of BS was observed only in 8 females and in 3 males with GJH (22% and 16% all of females and males with GJH, respectively). We suspect that this finding may be due to the influence of both, mobility of the spine and hamstring flexibility (that was low in GJH groups with the mean range of hip flexion 57.3°±7.7 and 53.1°±9.3 for females and males with GJH, respectively) on FTF test result	Clinical examination of the pelvic-hip complex muscles and trunk flexibility by use of SLR, O-JHF, T-JHF, FTF and LTF revealed to be insufficient in diagnosing GJH in children aged 10–13 years. Thus, the Beighton scale should be considered a standard element of physiotherapeutic examination of the musculoskeletal system in children and youth.	N Y
Evans, Rome and Peet (2012)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** Participant characteristics: 30 healthy asymptomatic children aged 7-15 years children were recruited as a convenience sample from the Auckland University of Technology podiatry clinic and from staff associated with this clinic (New Zealand)	**BS** **N** **LLAS**	Inter-rater reliability ICC = 0.73 Inter-rater reliability ICC = 0.78	Intra-rater reliability ICC = 0.96-0.98 Intra-rater reliability ICC = 0.90-0.98	N/A		N N
Farmer *et al.*, (2010)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 250 healthy volunteers 131 female, with a mean age of 39 years (range 18–89 yrs) and mean body mass index (BMI) 25.4 kg/m2 (range 16.7–45.7 kg/m2). Ethnically, 78.4% were Caucasian, 13.6% were South Asian, 5.6% were Afro-Caribbean, and 2.4% were East Asian (Japanese and Chinese). The mean Beighton score was 1.8 (range 0–8). Forty-four out of 250 (17.6%) of the cohort had a Beighton score ≥ 2 (UK)	**5pQ** **BS** **Skin Hyperextensibility** **N/A (skin measured with a harpenden caliper)** **CSES (Corrected Skin Extensibility Score)**	N/A Inter-rater reliability ICC = 0.96 Kappa = 0.83	N/A	Cut off ≥ 4/9 Sensitivity = 0.853 Specificity = 0.85 AUC = 0.90 (95% CI 0.84–0.96) Sensitivity = 0.72 Specificity = 0.75 AUC = 0.86	The mean Corrected Extensibility Score was 23.84%/mm in participants found to be hypermobile versus 13.55%/mm in the normal mobility group ($p < 0.0001$). CSES sensitivity was 0.72, specificity 0.75. The κ value for interobserver variability was 0.83.	N Y

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Glans *et al.*, (2020)	**Cross Sectional Survey** **JBI LEVEL 2B** 328 participants aged 18–65 years fluent in the Swedish language (Sweden)	**BS -self reported (Translated into Swedish)** **N/A**	N/A	Intra-rater reliability ICC = 0.92 (95% CI 0.85, 0.96)	sensitivity = 91% (95% CI 72, 99%) Specificity = 75% (95% CI 69,72) 80%)ROC = 0.87 (95% CI 0.79, 0.95)	Intra-rater reliability ICC = 0.92 (95% CI 0.85, 0.96)	N Y
Hakim And Grahame (2003)	328 participants aged 18–65 years fluent in the Swedish language (Sweden) 212 Consecutive patients from the Joint Hypermobility Clinic 30 were hypermobile, 182 had joint hypermobility syndrome (UK)	**5pQ** **N/A**			Cut off score >2 Hypermobile:: Sensitivity =84%Specificity 89% HSD: Sensitivity=80%Specificity = 84%	The questionnaire correctly identified 84% of all case control studies	Y
Hamonet and Brock (2015)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** Participant Characteristics: 232 patients with known EDS, recruited at random, from an EDS specialty clinic aged 2 to 70 Female average age (84% of individuals) of 32 years (SD: 17.5) Males average age (16% of individuals) of 20 years (SD: 14.1) (France)	**Retraction of hamstrings as an indicator of GJH** **N**	N/A	N/A	87.5% of sample had retraction of hamstrings Retractions of the triceps surae were found in 90.9% of patients, and retractions of the soles of feet were observed in 95.9% of patients The impact of retraction on the Beighton palms-on-the-floor test is very great indeed 97.8% of patients who present a retraction of the hamstrings of over 45° cannot perform this manoeuvre	The presence of muscle and tendon retractions in the posterior muscle compartments of the lower limbs and the soles of the feet constitute clinical features of Ehlers-Danlos syndrome. They should be addressed with a view to prevention and treatment, mainly through physical therapy	N N
Hansen *et al.*, (2002)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** A total of 188 98 female, 90male from sports clubs in Copenhagen aged 9-13 years (Denmark)	**Compared Parents answers to physicians' assessment using pictures and questionaires to arrive at a BS measurement** **N**	varying kappa values for the individual componenents of BS between physicians and between physicians and parents. Kappa ranges from <0.44 to 1.0 Lowest agreements was for elbow and knee hyperextensibility Highest kappa ranges were between the 2 experienced physicians, lowest between these physicians and parents	N/A	N/A	The study illustrates that skilled physicians have a low inter observer variability when testing children for joint hypermobility. Untrained parents were unable to identify hypermobility in their children using the BS	N N
Heidbreder *et al.*, (2008)	**Case Control Study** **JBI LEVEL 3B** 17 patients with classical EDS, five patients with vascular type EDS, 17 patients with sCAD without known connective tissue disease and 29 healthy control individuals 38 EDS patients 22 Females 16 males Age range 6-76 29 healthy controles 13 females 16 males Mean age 33.1 (Germany)	**Skin hyperextensibility** **N/A**		Significance of differences in skin extensibility between groups (Student's *t*-test) Group 1 / Group 2 / *p*-value Controls / Vascular type EDS / 0.593 Controls / Classical type EDS / <0.0001 Vascular type EDS / Classical type EDS / <0.0001 Controls / sCAD / 0.367 Classical type EDS / sCAD / <0.0001 Vascular type EDS / sCAD / 0.111 Each of the four groups is compared with each of the other groups. EDS, Ehlers–Danlos syndrome; sCAD, spontaneous cervical artery dissection. The method assessed is capable of detecting increased skin extensibility in classical type EDS. Increased skin extensibility is not present in sCAD patients	Significance of differences in joint hyperextensibility point values between groups (ANOVA) Group 1 / Group 2 / *p*-value Controls / Vascular type EDS / 0.015 Controls / Classical type EDS / <0.0001 Vascular type EDS / Classical type EDS / 0.57 Controls / sCAD / 0.91 Classical type EDS / sCAD / <0.0001 Vascular type EDS / sCAD / 0.016 Each of the four groups is compared with each of the other groups. EDS, Ehlers–Danlos syndrome; sCAD, spontaneous cervical artery dissection.		N Y

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Jannson *et al.*, (2004)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **N**	N/A	N/A	The largest difference concerning cut-off point is found at the age of 15 y. According to this method, girls were considered to be hypermobile if 8 manoeuvres were performed correctly. The cut-off for boys at the same age would be at 6 manoeuvres, if correctly performed. *Table 1*. The distribution (%) of children with different cut-off points for general joint laxity estimated by the Beighton score. Figures in bold indicate the closest upper limit for 95 % of the study population (*n* = 1845). Beighton score Age (y) Gender ≥4 ≥5 ≥6 ≥7 ≥8 9 9 Boys 37.6 20.2 16.1 5.6 **5.0** 0.3 Girls 47.9 26.4 21.8 7.3 **5.7** 1.2 12 Boys 21.0 8.6 **6.2** 0.6 0.3 0.0 Girls 37.8 19.9 14.9 **5.4** 3.0 0.9 15 Boys 15.5 7.9 **7.2** 2.8 1.7 0.3 Girls 53.0 33.4 24.0 13.9 **7.4** 2.0		N
	1845 children were clinically examined concerning general joint laxity in grades three (n = 573; 317 boys, 256 girls), six (n = 703; 349 boys, 354 girls) and nine (n = 569, 284 boys, 285 girls). The mean ages of the children were 9 y (95% CI: 8.99–9.10) in grade 3, 12 y (95% CI: 11.93– 12.03) in grade 6 and 15 y (95% CI: 14.93–15.05) in grade 9 (Sweden)						N
Johnson wand and Simmonds (2019)	**Cross Sectional Cohort Study** **JBI LEVEL 2A**	**LLAS** **N**	N/A	N/A	cut off ≥4/10 Sensitivity = 67% Specificity =94% Positive Predictive Value (50%) Negative Predictive Value (97%).		N
	Participant characteristics: Thirty-six male, professional footballers aged between 18 and 37 years old (UK)						N
Junge *et al.*, (2013)	**Longitudinal Prospective Cohort Study** **JBI LEVEL 2A**	**BS (Method A)** **Y**	kappa = 0.49-0.94 Method A	N/A	Cut off ≥5 Both methods need to be tested for their predictive validity at higher cut-off levels, e.g. ≥6 and ≥ 7		N
	1300 children in the Municipality of Svendborg aged (7-8 years) and fourth grade (10-12 years) (Denmark)	**BS (Method B)**	Kappa= 0.30-0.84 Method B				Y
Juul-Kristensen (2007)	**Case Control Study** **JBI LEVEL 3B**	**BS** **N**	For the total Beighton score, either currently or historically, ICC was 0.91. For GJH, inter-rater reliability ICC =0.88 kappa = 0.74	N/A	N/A	Authors state: Further research on the validity of tests and criteria for GJH and BJHS is urgently needed. In spite of a different cut-off level used for the diagnosis of GJH, the present reproducibility of the criterion (0.74) was at the same level as previously reported (0.78)	N
	Patients aged between 18 and 71 yrs. Cases and controls for phases 1 and 2 were randomly selected from files of patients previously referred to the out-patient clinic of the Department of Medical Orthopaedics and Rehabilitation (Denmark)	**Brighton Criteria**	Inter-rater reliability ICC = 0.93 kappa = 0.84				Y
		Rote`s-Que´rol	Kappa between 0.31 – 0.80 currently, or historically				
Kulik and Gebska (2018)	**Cross-Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **N**	N/A	N/A	Prevalence of GJH in sample = 34.3% Mann-Whitney U test, p = 0.085	The occurrence of joint hypermobility in children using the Beighton score is greater than using the Brighton criteria. The Beighton score and Brighton's criteria are not well correlated, so a standardized method for diagnosing hypermobility should be developed	N
	The study covered a group of 102 students (60 boys, 42 girls) aged between 6 to 11 years (Poland)	**Brighton Criteria**			Prevalence of GJH in sample = 0.98%		N
Kwon *et al.*, (2013)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **Y**	N/A	N/A	Cut off = 4/9 GJH present in 238 girls (58.9%) and 97 women (36.5%)	Recommend setting 5/9 as cut off in paediatric patients	N
	404 healthy female paediatric patients and 266 adult female patients Age range 6-12 in girls and 24-50 in adult females (Korea)						

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Lamari, Chueire and Cordeirdo (2005)	**Cross-Sectional Cohort Study** **JBI LEVEL 2B** 1,120 healthy asymptomatic children of mixed racial background (534 boys, 47.7%; 586 girls, 52.3%; age range: 4-7 years) (Sao Paulo, Brazil)	**BS** **N**	N/A	N/A	Cut off> 4 GJH present in 64.6% of the children	Recommend additional methodological parameters and criteria to characterize joint mobility in addition to BS	N N
McCormack, Grahame and Hakim (2003)	**Case control Study** **JBI LEVEL 3B** 149 dance students, 85 from the Lower School and 64 from the Upper School, and 71 professional ballet dancers were recruited from the Royal Ballet School and the Royal Ballet Company, London. 36 pupils from a local secondary school and 31 adults working at The Royal Opera House, London (home of the Royal Ballet) were recruited as controls for the senior student and professional dance cohorts, respectively (UK)	**BS + Contompasis Score + Brighton Criteria** **Y**	*Table 2.* Prevalence and odds ratios for hypermobility (Beighton score ≥ 4/9) and BJHS, and the means for Contompasis score and right 5th finger dorsiflexion. Variable — Lower School[a] — Upper School — Upper School Controls — Company — Company Controls Mean Beighton score (range) Female — 5.6 (1–9) — 7.5 (3–9) — 4.3 (0–9) — 7.1 (3–9) — 3.1 (0–7) Male — 4.3 (1–8) — 6.2 (1–9) — 2.5 (0–9) — 6.4 (3–9) — 3.1 (0–7)[a] Prevalence of hypermobility (Beighton score ≥ 4/9) (%) Female — 35/47 (74) — 33/35 (94) — 13/21 (62) — 41/43 (95) — 8/16 (50) Male — 23/38 (82) — 24/29 (83) — 3/15 (20) — 23/28 (82) — 6/15 (40) Combined OR for hypermobility — — — 11.3 (4.1–31.2), p = 0.001 — 11.1 (3.8–31.8), p = 0.3 — Prevalence of BJHS (%) Female — 22/47 (47) — 16/35 (46) — 4/21 (19) — 11/43 (26) — 2/16 (13) Male — 17/38 (45) — 10/29 (35) — 1/15 (7) — 10/28 (36) — 2/15 (15) Combined OR for BJHS — — — 3.9 (1.3–11.3) p = 0.01 — 1.7 (0.6–4.7) p = 0.3 — Nonparametric t test of Contompasis Score (95% CI) Female — — 41.5 (39.2–43.8) — 28.7 (25.2–32.2) — 38.2 (36.4–40.1) — 26.1 (22.5–29.6) Mean difference — — 12.8 (8.6–16.7) — — 12.1 (8.2–16.1) — Male — — 36.0 (33.5–38.6) — 23.7 (19.4–28.0) — 35.7 (33.1–38.4) — 24.3 (21.3–27.3) Mean difference — — 12.3 (7.4–17.1) — — 11.4 (7.6–15.3) — Prevalence of hypermobility (Contompasis score ≥ 26) (%) Female — — 38 (100) — 12/21 (57) — 43 (100) — 9/16 (56) Male — — 27/29 (93) — 3/15 (20) — 28 (100) — 6/15 (40) Nonparametric t test of right 5th finger dorsiflexion (95% CI) Female — — 88 (77–85) — 68 (60–76) — 90 (87–93) — 74 (63–84) Mean difference — — 21 (12–29) — — 16 (8–24) — Male — — 90 (88–92) — 62 (50–74) — 89 (85–94) — 67 (55–79) Mean difference — — 28 (16–40) — — 22 (12–33) — [a] Observation only and no controls. Scoring systems not validated for children.	An OR of 11.0 (95% CI 3.3–31.8) was found for hypermobility in dancers for both the ballet school and the professional company Odds ratios for BJHS in student dancers were significant, OR = 3.9 (95% CI 1.3–11.3), but not so in professional dancers: OR = 1.7 (95% CI 0.6–4.7)	N N
McGillis *et al.*, (2019)	**Cross sectional cohort study** **JBI LEVEL 2A** A sample of 298 patients referred to the Goodhope Ehlers Danlos clinic, 156 with a previous diagnosis of EDS (Canada)	**BS** **N**	Only 46% (n = 51) of HSD/LJH patients scored a BS of 4/9 as assessed by an EDS practitioner (SD = 2.12). In the H-EDS group the average BS was 6/9 (SD = 1.45). The HSD/LJH patients were reviewed for a comparative analysis of BS as assessed by the referring physician compared to the objective assessment by the EDS specialist using a goniometer. 82% (n = 91) of patients had a BS completed by their referring physician on initial referral. It was found that the average BS on referral was 6/9 (SD = 2.17). 81% (n = 74) of referring physicians assessed the BS as higher than the assessment of the EDS practitioner, 14% (n = 13) had the same BS and only 5% (n = 4) had a BS lower on the referral than as assessed in the EDS clinic.			The findings call into question the validity of solely applying the BS as a measure of GJH in this population, or alternatively supports the notion that those symptomatic patients with a low BS have a different disorder entirely	N N
Meyer *et al.*, (2017)	**Case Control Study** **JBI LEVEL 3B** Participant characteristics: 112 participants in total 100 females 12 males Age range 17-71 years likely hypermobile group were comprised of elite dancers (working as a dancer full-time) from Western Australian Academy of Performing Arts (Edith Cowan University) and the West Australian Ballet (Perth) Control group: healthy individuals who had not undertaken long term training likely to affect joint mobility such as dance, gymnastics and acrobatics from University of Sydney (Australia)	**LLAS** **Y**	Inter-rater reliability ICC 2,1 = 0.85, 95%CI (0.67 - 0.94) p < 0.001			Beighton score (BS) as assessed by primary care practitioner was found to be higher than assessment by EDS practitioner in 81% (n = 74 of 91) of cases. Generalized joint hypermobility was confirmed in only 46% (n = 51 of 111) of patients who had a previous diagnosis of hEDS	**N** **N**
Mikołajczyk *et al.*, (2012)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** Participant characteristics: 120 healthy asymptomatic children aged 15 years of age 60 girls and 60 boys (Poland)	**Sasche Scale** **N**	N/A	N/A	Cut off ≥ 7/13 30% positive	Not compared to BS limiting usefulness of findings	N N

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
De Moraes *et al*., (2011)	**Cross Sectional Cohort Survey** **JBI LEVEL 2B** Portuguese was applied to 2,523 individuals, of whom one was an elementary school student and 2,522 attended the following three universities: USP, at the Ribeirão Preto Campus; Universidade de Franca (Unifran); and Centro Universitário Barão de Mauá, in the city of Ribeirão Preto. The university students attended the following courses: medicine (1st to 4th year, 609 students); nursing; psychology; physical therapy; occupational therapy; law; chemistry; medical physics; and speech therapy (Brazil)	**5pQ (translated into Porteuguese)** **5pQ** **N**	N/A	N/A	Sensitivity = 70.9% specificity = 77.4% ROC= 0.786		N N
Morris *et al*., (2017)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 1584 participants at 14 years of age taken from pregnancy cohort at King Edwards Menorial Hospital Western Australia (Australia)	**BS** **N**	N/A	N/A	For cut off ≥4 prevalence of GJH= 48.0% (60.6% and 36.7% in girls and boys, respectively) OR, 1.29; 95% CI, 1.12-1.49 For cut off ≥6, GJH = 18.6% (26.1% and 11.5% in girls and boys, respectively) OR, 1.22; 95% CI, 1.01-1.47		N N
Naal *et al*., (2014)	**Cross Sectional Cohort Study** **JBI LEVEL 2A** 55 participants with a diagnosis for femero-acetabular impingement syndrome, 18 females 37 males mean age 28.5±4.1 years (Switzerland)	**BS-self in FAI (FAI is an indicator for GJH)** **N**	N/A	N/A	Cut off ≥4 Prevalence of GJH = 32.7 % (50 % of females and 24.3 % of males) Cut off ≥6 Prevalence of GJH 16.4 % (27.8 % of females and 10.8 % of males)	Correlative relationship between BS self and hip flexion ($r = 0.61$, $p < 0.01$), internal rotation ($r = 0.56$, $p < 0.01$), and external rotation ($r = 0.44$, $p < 0.01$)	N Y
Nicholson and Chan (2018)	**Case Control Study** **JBI LEVEL 3B** Convenience sample of 112 adult participants (mean age 24.3 ± 5.5years) divided into known hypermobile cases, likely hypermobile cases and controls (Australia)	**BS** **Y** **ULHAT**	Inter-rater reliability ICC2,1 = 0.92 N/A	N/A	The cut-point ≥7/12 (sensitivity 0.84, specificity 0.77, +LR 3.7, -LR 0.2)	Beighton score cut-off of ≥4/9 revealed low to moderate levels of agreement with clinical opinion for the control and likely hypermobile groups (48% and 45% respectively), and high levels of agreement only with the known hypermobile group at 91%. Utilizing a Beighton cut-off score of ≥5/9 improved the moderate levels of agreement with clinical opinion for the control and likely hypermobile groups (58% for both), while maintaining high levels of agreement with the known hypermobile group at 88%. Having identified a cut-off score of ≥5/9 for the identification of GJH, the Beighton scoring system identified GJH in 45% control, 60% likely hypermobile and 94% known hypermobile participants. With the more stringent cut-off of ≥5/9 and higher agreement with clinical opinion, the McNemar's test still revealed a significant difference in the agreement between clinical opinion of generalized hypermobility and the Beighton score for the control and likely hypermobile participants ($p < 0.001$), but not in the known hypermobile group ($p = 1.0$)	N N

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Ohman, Westblomm, Henrikkson (2014)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** A total of 128 healthy children 74 females 54 males aged 5-8 years (Sweden)	**BS** **Hospital del Mar** **N**	N/A	N/A	Cut off ≥4 BS Prevalence of GJH =12% Hospital Del Mar Criteria = 34%	The Hospital del Mar criteria has been developed over a period of years and the version used in this study is not exactly the same as in other studies. This makes comparisons more complicated.	N N
Parvenerh and Shiari (2016)	**Case Control study** **JBI LEVEL 3B** 200 paediatric participants aged 3 to 16. 100 children with BJH compared to age and sex matched controls 42 males in each group 64 females in each group (Iran)	**Parvener and Shiari Score (2016)** **N**	N/A	N/A	Sensitivity = 100%, specificity = 98%, positive predictive value= 100%, negative predictive values 98% Accuracy was 99% and balanced accuracy was 99%. The area under the ROC = 0.99	Data analysis revealed significant correlation between Beighton and the new criteria	N N
Patel *et al.*, (2018)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 200 patients 108 females 92 males aged between 3 and 15 years (mean 10.1) (UK)	**BS** **N** **Brighton Critera**	N/A	N/A	Mean standard deviation (SD) BS was 2.06 (2.2), and the range was 0–8. Comparing males versus females, mean BS SD was 1.71 (2.25) versus 2.36 (2.14); p=0.0378, age was 9.75 versus 10.13, and BS range was 0–7 versus 0–8.64 children (32%) complained of pain in at least one joint, though the mean SD BS in these patients was 1.71 (1.86)	The "Brighton" score, when combined with BS using the higher diagnostic score of 5/9, could be the more reliable scoring system	N N
Pearsall *et al.*, (2006)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 57 athletes (29 men; 28 woman; age = 20.9 ± 1.45 yr) without a history of previous injury (USA)	**BS** **KT 2000 arthrometer** **N (Ankle arthrometer)**	N/A	N/A	N/A	Non-significant correlations were observed among the test variables for generalized joint laxity (.21 to .37; p > .05) and instrumented ankle and knee joint laxity (.19 to .21; p > .05). When examined by gender, no statistically significant correlations (.05 to .40; p > .05) were found	N N
Phan *et al.*, (2020)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 57 pre-professional and 29 professional ballet dancers (21±4years, 64% female, mean 13.7years training) were recruited (Australia)	**LLAS** **N**	ICC2,1= .0.85,95% (CI . 0.67 to 0.94, p < 0.001)	N/A	N/A	The right leg was significantly more hypermobile than the left for the whole cohort (44% vs 40% meeting 7/12 for the LLAS; LLAS mean/12(SD): right:5.0(2.4) and 7.6(1.9); left:4.8(2.1) and 6.7(2.0) in pre-professionals and professionals respectively (p ¼ 0.02)). Subtalar pronation (p < 0.001) and hip abduction/external rotation (left:p¼ 0.01; right p < 0.001) were significantly more hypermobile bilaterally in professionals. Three hypermobility profiles on the left and four on the right lower limb were identified.	N N
Romeo *et al.*, (2016)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 284 healthy preschool children 146 boys 138 girls 26 preschool children with genetic disorders 15 boys and 11 girls Mean age was 33.6 (Italy)	**Modified BS** **Y**	For cut off score of 6, kappa = 0.75, Kappa = 0.78	N/A	N/A	**Revised version of the Beighton score can be used to define generalized hypermobility for children up to 5 years of age**	**N** **N**
Riley *et al.*, (2020)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 51 University students 17 male 34 female Average age 23 (USA)	**BS** **Y**	Interrater reliability ICC =0.52 for intertester reliability at visit 1 and 0.86 at Visit 2	intratester reliability ICC = 0.88 for Tester 1 and 0.71 for Tester 2		Intraclass correlation coefficients (ICC2,1) were 0.52 for intertester reliability at visit 1 and 0.86 at visit 2, with intratester reliability of 0.88 for Tester 1 and 0.71 for Tester 2 for the BS. Intertester prevalence-adjusted bias-adjusted kappa (PABAK) values for the Beighton GJH cutoff scores were 0.80 e0.84 and 0.80 to 0.92 for intratester reliability.	**N** **N**
Sauers *et al.*,2001)	**Cross Sectional Cohort Study** **JBI LEVEL 2B** 51 adults 28 females 23 men mean age 22 years (USA)	**Shoulder hypermobility as a marker for GJH using a shoulder arthrometer**	N/A	N/A	N/A	No moderate or stronger correlations between laxity, passive range of motion, and generalized joint laxity were seen The BS scores did not correlate highly with any of the passive range of motion values (range, 0.01- 0.48)	N N

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Schlager *et al.*, (2018)	**Cross Sectional Cohort Study** **LEVEL 2B**	**BS** **Y**	Inter-rater reliability 0.72 (95% CI 0.55-0.83)	Intra-rater reliability 0.76 (95% CI 0.54-0.88)	**N/A**	Good results for all 3 scores	**Y (Refers to (GRRAS) (QAREL) Kottner *et al.*, (2011) Lucas *et al.*, (2010)**
	30 healthy adults aged 18-65 from a convenience sample of a rehabilitation company within primary care Forty-nine adults, 38 women and 11 men, mean (SD) age 39.8 (13.5) years participated in the inter-raterreliability study. Twenty-nine adults, 23 women and 6 men, mean (SD) age 39.9 (12.5) years participated in the intra-raterreliability study (Sweden)	**Hospital Del Mar**	Inter-rater reliability 0.81 (95% CI 0.69-0.89)	Intra-rater reliability 0.86 (95% CI 0.73-0.93)	N/A		
		Contompasis	Inter-rater reliability 0.82 (95% CI 0.69-0.89)	Intra-rater reliability 0.79 (95% CI 0.57-0.90)	N/A		**Y**
Shlager *et al.*, (2020)	**Cross sectional study** **JBI LEVEL 2B**	**BS** **N**			Cut off ≥5, (15.9%) indicating a low post-test probability	There is uncertainty in identifying generalized joint hypermobility in young women using the BS-self reported with a cut-off level of ≥2 when the Beighton score ≥ 5 is used as the reference test. The strength of the BS-self reported is to rule-out women without generalized joint hypermobility	**N**
	301 pregnant women chosen from a larger sample of 8029 (Sweden)	**BS-self reported**			Cut off ≥2 AUC 0.73 (95% CI 0.67–0.79) Sensitivity 84.1% (95% CI 69.9–93.4) specificity, 61.9% (95% CI 55.6–67.8). The false-positive rate was 38%. The PPV and NPV ranged between 13.4 and 42.4% and between 90.3 and 99.2%, respectively. The LR+ increased by 2.2 when using a cut-off level of ≥2 for the self-reported 5PQ		**N**
Singh *et al.*, (2017)	**Cross Sectional Study** **JBI LEVEL 2B**	**BS** **Y**	N/A	N/A	Cut-off of ≥ 4 Sensitivity of 0.8% Specificity of 99.3% (P<0.001) False positive rate of 60%	To lower the risk of a false-positive diagnosis of GJH, further tests of hypermobility need to be utilized A cut-off of54 was only found to be appropriate for females aged 40_59 years and males aged 8_39 years	N
	1000 healthy participants aged 3 to 101 years were recruited via a convenience sample (Australia)						N
Sirajudeen *et al.*, (2020)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **Y**	N/A	N/A	Cut off ≥ 4 Prevalence of GJH =15.2% Cut off ≥ 6 Prevalence of GJH =7.6%	The prevalence reported in this study among school-aged children was comparable with those reported worldwide	**N**
	303 children. 161 girls 142 boys Age range 8-14 (Majmaah region of Saudi Arabia)						**N**
Skiwiot *et al.*, (2018)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **N (Digital Inclinometer)**	N/A	N/A	GJH = 64.9% BS $\chi2(1) = 6.485$; $p = 0.011$		N
	77 dancers 19 female 58 male Aged 18-25 years from the Polish dance theatre (Poland)	**5pQ**			GJH = 74% 5pQ–$\chi2(1) = 11.199$; $p = 0.001$		
		Sachse's criteria			GJH = 59.7% Sachse–$\chi2(1) = 11.206$; $p = 0.001$		N
Smits-Engelman *et al.*, (2018)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **Y**	N/A	N/A	Cut off score >= 5 More than 35% of children scored more than 5/9 on the Beighton score	Authors recommended that 7/9 be the cut-off for the Beighton score.	N
	551 paediatric participants 293 females 258 males aged 6 to 12 years (Holland)						N

Author	Type of study JBI level of evidence Participant characteristics	Scoring/measurement system Goniometer use Y/N	Summary Statistics Inter-rater Reliability	Intra-rater Reliability.	Validity	Other Statistics, findings from paper	Reference to COSMIN or other similar guidelines Y/N Blinded Y/N
Steinberg *et al*., (2016)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **Y** **Range of Motion including Passive En Pointe, Ankle Plantar Flexion/Ankle Dorsiflexion, Hip ROM**	Intraclass correlation ICC range between 0.74and 0.95 and for body measurements between 0.90 and 0.95. Kappa = 0.82	Intraclass correlation ICC range between 0.87 and 0.96 and for body measurements between 0.95 and 0.7. Kappa = 0.86	prevalence of JHM is significantly higher among dancers compared with the control subjects (P < 0.001). Joints' ROM is higher among dancers with JHM compared with dancers without JHM (P < 0.05).		N
	240 non professional female dancers, aged 8 to 16 years, and 226 female nondancers of similar age (Israel)						N
Van der Giessen *et al*., (2001)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **N** **5pQ**	N/A	N/A	Cut off ≥ 4 GJH present in 26.5% of participants		N
	773 Dutch children 378 girls 395 boys aged 4-12 years old (Holland)						N
Vallis *et al*., (2015)	**Cross Sectional Cohort Study** **JBI LEVEL 2B**	**BS** **N**	Inter-rater reliability ICC range 0.72-0.80	Intra-rater reliability ICC range 0.71-0.82	N/A	BS superior to Contompasis	N
	36 physiotherapy students 9 female 27 males mean age 22.7 years (USA)	**Contompasis**	Inter-rater reliability ICC range 0.58-0.62	Intra-rater reliability ICC 0.73-0.82			Y

Table 3.2 Summary of Systematic Review and Narrative Review Findings	
Systematic Reviews	**Conclusions/Summaries of Findings (directly from the literature)**
Juul-Kristensen *et al*., (2017) JBI LEVEL 3A	*'Studies demonstrated limited positive to conflicting evidence with regards to the BS. There is a lack of quality evidence with regards to the other scoring systems assessed, so no reliable conclusions were able to be drawn. More research is required to further evaluate validity and reliability of the BS as well as the other scoring systems, however the authors still recommend: For adults BS should be used with a cut off score of 5, but recommend including historical information. For paediatric patients a BS with cut-point of 6 is recommended.'* *'More research is required before evidence-based recommendations can be developed about which assessment method is the best for identifying GJH.'*
Bokhorn *et al*., (2022) JBI LEVEL 3A	*'The Beighton score is a highly reliable clinical tool that shows substantial to excellent inter- and intrarater reliability when used by raters of variable backgrounds and experience levels. While individual components of risk of bias among studies demonstrated large discrepancy, most of the items were adequate to very good, despite the range in study designs, participant characteristics and different cut off levels used.'*
Palmer *et al*., (2020) JBI LEVEL 2A	*'Studies reviewed were of high standard in terms of both methodology and reporting. Three of the four studies included reported at least one measure of tissue mechanics that distinguished between adults with hEDS/HSD and healthy participants. These included assessment of skin, muscle (brachioradialis and biceps brachii) and tendon (patellar and Achilles). Only 4 studies have been conducted on this subject and further validation of methodologies reviewed is required including tissue stiffness meters and ultrasound and other techniques of standardisation.'*
Narrative reviews	**Conclusions Summaries of Findings (directly from the literature**
Malek *et al*., (2021) JBI LEVEL 5A	*'A consistent finding in the literature reviewed by these demonstrates that the BS is unable to identify GJH in joints not included in the scoring system, making it an unviable scoring system for identifying GJH and it should not be used as the primary tool in differentiating between L-JH and GJH.'*
Desfor (2003) JBI LEVEL 5A	*'Criteria for assessment of JH and BJHS are needed, further research is required. Dance educators and clinicians need more information on the relationship between hypermobility and dance training, strength, proprioception, injury and osteoarthritis.'*
Czaprowski *et al*., (2012) JBI LEVEL 5A	*'Out of the many methods in use, the most common are the Marshall scale, the Beighton scale and the Carter and Wilkinson scale. What makes the value of the Beighton and Carter & Wilkinson scales higher is that they assess mobility in multiple joints, which reduces the risk of false positive or negative results. The Hakim and Grahame questionnaire seems to be a potential alternative, in accordance with the hands-off principle (no need to conduct a direct clinical examination) promoted by the authors.'*
Remvig *et al*., (2007) JBI LEVEL 5A	*'Using a 6 cut level, intra- and inter-observer kappa scores were 0.75 and 0.78, respectively. Beighton scoring recommendations have been correlated with a global joint mobility index as well as with 2 other scoring systems, the Carter and Wilkinson, and the Rotès-Quérol. All illustrate high concurrent validity with one another.'*

Table 1 Intra- and inter-examiner reliability for positive hypermobility findings for individual manoeuvres, GJH classification, and agreements with overall score

	Intra-examiner reliability—adults	Inter-examiner reliability	
		Adults	Children
Individual manoeuvres			
Hyperextension of the fifth finger—Kappa	0.78 [47] 1.00 [47]	0.40 [48] 0.61 [48] 0.70 [47] 0.73 [47] 0.79 [6]	0.43 [49] 0.48 [49] 0.49 [49] 0.69 [49]
Apposition of the thumb to the forearm—Kappa	0.52 [47] 0.71 [47]	0.87 [6] 0.91 [47] 0.94 [48] 1.00 [47] 1.00 [48]	0.82 [50] 0.82 [49] 0.84 [49] 0.89 [49] 0.94 [49]
Hyperextension of the elbows—Kappa	0.61 [47] 0.76 [47]	0.21 [47] 0.34 [48] 0.50 [47] 0.64 [48] 0.92 [6]	0.68 [50] 0.54 [49] 0.60 [49] 0.63 [49] 0.73 [49]
Hyperextension of the knees—Kappa	0.19 [47] 0.84 [47]	0.31 [47] 0.40 [47] 0.85 [48] 0.86 [6] 0.90 [48]	0.30 [49] 0.43 [49] 0.44 [50] 0.62 [49] 0.62 [49]
Forward flexion of the trunk—Kappa	1.00 [47]	0.88 [48] 0.93 [6] 0.94 [47]	0.64 [49] 0.82 [50] 0.84 [49]
Classification of GJH (BS ≥ 4)—Kappa	0.78 [47]	0.54 [47] 0.78 [48]	–
Classification of GJH (BS ≥ 5)—Kappa	0.47 [47]	0.66 [47] 0.66 [48]	0.59 [49] 0.64 [49]
Agreement with overall score—Spearman's γ	0.86 [51]	0.87 [51]	
Agreement with overall score—Intraclass Correlation Coefficient (ICC)	0.76 [47] 0.84 [52] 0.92 [52]	0.72 [47] 0.89 [52] 0.94 [52]	

'In one article (Bird, 1979), Beighton tests using 0–9 scoring correlated well with a Global Index: r = 0.81 (p < 0.001). A cut off level for GJH was not discussed.'

Drabik, Bys and Gawda (2020)

JBI LEVEL 5A

'In order to diagnose EDS-HT, one should be guided not only by the established standard, which is the Beighton score, but also by additional tests that would confirm the diagnostic decision and reduce the risk of error. Extension of diagnostic tests with additional criteria, which are presented in the article, would make diagnoses more accurate and reduce the possibility of false diagnoses.'

Day, Koutedakis and Wyon (2011)

JBI LEVEL 5A

Reference	Number of subjects	Population	Age range (yrs)	Diagnostic test	Cutoff Point	Prevalence %
Seckin *et al.* [73]	861	Turkish high school students	15.36 ± 1.11	Beighton score	≥4/9	11.7%
Al-Rawi *et al.* [5]	1 774	Iraqi university students	20–24	Beighton score	≥4/9	30.0%
Beighton *et al.* [9]	502	Tswana (South African) villagers	20–65+	Beighton score	≥3/9	15.7%
Birrell *et al.* [10]	204	Yoruba (Nigeria) villagers	6–66	Beighton score	≥4/9	43.0%
Pountain [63]	309	Cross-section of Omani people	16–25	Beighton score	≥4/9	20.0%
Didia *et al.* [16]	550	Nigerian undergraduate university students	17–30	Beighton score	≥5/9	13.0%
Klemp *et al.* [49]	Maori = 438, European = 354	Cross-section of Maori and European New Zealanders	30–47	Beighton score	≥4/9	Maori = 6.2% European = 4%
Lamari *et al.* [54]	1 120	Brazilian pre-school children	4–7	Beighton score	≥4/9	64.6%
Subramanyam *et al.* [76]	1 000	South Indian school children	6–15	Beighton score	≥4/9	17.2%
Gedalia *et al.* [24]	260	North American school children	5–17	Beighton score	≥3/9	12.0%
Larsson *et al.* [55]	660	North American music school students and faculty members	14–68	Carter & Wilkinson test	≥3/5	9.7%
Vougioka *et al.* [80]	2 432	Greek school children	5–14	Beighton score	≥3/9	8.8%
Forleo *et al.* [20]	1 005	Brazilian school children	5–17	Beighton score	≥5/9	36.3%
Carter and Wilkinson [13]	285	English primary school children	6–11	Carter & Wilkinson test	≥4/5	7.0%
Rikken-Bultman *et al.* [68]	1 568	Dutch secondary school children	Beighton for 4–17 Biro for 12 = 17	Biro & Beighton score	Biro = ≥3/5, Beighton = ≥4/9	Beighton = 14.2%, Biro = 12.3%
Kirk *et al.* [45]	64	English rheumatology patients and their relatives and healthy people	5–57	Carter & Wilkinson test	≥3/5	25%
Kirk *et al.* [45]	64	English rheumatology patients and their relatives and healthy people	5–57	Measurement in excess of normal ROM in 8 joints	≥8/17	41.0%
Seow *et al.* [74]	306	Singaporean Chinese, Malay and Indian adults	15–39	Beighton score	≥5/9	17.0%

'The modest amount of research on injury rates in hypermobile dancers confirm that they have substantially more tendon injuries and longer healing times than normal dancers. Further research is required into this vulnerable group of dancers in order to ensure that there is more understanding of the attrition rate from ballet schools to professional companies and health issues of the hypermobile dancer.'

Table 3.3 Grey Literature Summaries		
Author, date, JBI Level	**Type of paper**	**Summary of information (directly from the literature)**
(Castori *et al*., 2017) JBI LEVEL 3B	Framework	*'Many EDS patients, however, still remain without a laboratory confirmation and this lack of knowledge contributes to the patients' burden. This is mostly the case for individuals affected by the two largely overlapping conditions previously termed "Ehlers–Danlos syndrome, hypermobility type (EDS-HT)" and "joint hypermobility syndrome (JHS)".'* *'The attribution of GJH as a feature remains partly influenced by the examiner's professional experience and recognition of the need to look at all the joints (certainly at least those in the context of the clinical presentation) and not simply those assessed in these tools'*
(Grahame, 2007) JBI LEVEL 5C	Editorial	'*The true diagnosis in a majority of patients with JHS is overlooked in up to 95% of cases.'* Refers to the Remvig (2007) review acknowledging a lack of statistical sophistication in the majority of the papers published at the time and the need for further epidemiological studies to establish standardised joint ROM
(Grahame, 2004) JBI LEVEL 5C	Medscape article	*'Useful as an initial screen, the BS can no longer be considered the gold standard for recognizing hypermobility syndrome in clinical or epidemiological practice. It is an arbitrary all-or-none test that does not take into account the degree of laxity and in some individuals the score diminishes with advancing age, perhaps even reaching zero. Moreover, it samples only five sites, although studies have now shown that pauci-articular hypermobility is more common than polyarticular.'*
Kacheria, *et al.,* (2006) JBI LEVEL 2B	Research Poster, International Association for Dental Research, AADR, AACR Annual Meeting	*'Prevalence of hypermobility using the 9 point Beighton was 12.4%, while it was 18.4% using the 4 point scale. The sensitivity was 85% and the specificity was 91%. Stratified by gender, the sensitivity and specificity were 76% and 97%, respectively, for males, and 90% and 87%, respectively, for females. The 4 point method can be accomplished in just over a minute, versus 3 minutes for the full Beighton, and the subject does not need to remove clothing or get out of a chair. Conclusions: The shortened Beighton demonstrated good sensitivity and specificity and can be done easily and quickly. It is more convenient and hence may be used as a substitute for the full Beighton scoring examination in research studies.'*
Cincinnati Children's Hospital Medical Center Joint Hypermobility Team (2014) JBI LEVEL 5C	Clinical Practice Guidelines Comment: Comprehensive and holistic guidelines. Incorrectly defines "joint hypermobility" as a primary diagnosis, however written prior to 2017 nosology changes recommended by Castori (2017)	*'A Beighton Scale score of 5/9 or greater is used to classify JH (Scheper 2013 [1b], Junge 2013 [2b], Beighton 1998 [5a]). However, 4/9 was also frequently used to classify JH in the literature reviewed for this guideline.'*

Table 3.4 Summary Statistics – Ranges of Clinimetric Values for GJH Scores (Taken from Table 3.1 and 3.2)

Interpretation of Agreement Scores:

ICC (Koo and Li, 2016) less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively.
Kappa (McHugh, 2012) Minimally acceptable kappa score 0.8 or greater.
Spearman Ro (Akoglu, 2018) 1 = complete or perfect correlation, 0=no correlation
Crohnbach alpha (Taber, 2018) >0.7 is acceptable
Chi squared If the value is greater than the significance level, then the null hypothesis is rejected

Scoring System	Intra-rater agreement range	Inter-rater agreement range	Validity Statistics range
BS/BHJS	ICC= range between 0.74 (Steinberg *et al*., 2016) – 0.992 (Armstrong and Grieg, 2018), kappa = 0.82 (Steinberg *et al*., 2016)	ICC= 0.72 (Shlager *et al*., 2018) - 0.96-0.98 (Evans Rome and Peet, 2012) Spearman rho = 0.86 (p<0.0001) Boyle *et a.,* 2003)	The Cronbach's a for defining GJH, EDS-HT and JHS was 0.61, 0.79, and 0.44, respectively (Remvig *et al*., 2014)
Modified BS	No data	Interrater reliability of kappa = 0.49-0.94 Method A, Kappa= 0.30-0.84 Method B (Junge *et al*., 2013)	No data
Carter and Wilkinson Scoring System	No data	No data	Cut off score ≥ 3 / 5 GJH=9.2% - 25 % Day, Koutedakis and Wyon (2011)
Sasche Score	No data	No data	GJH = 56.5% Sachse–$\chi 2(1)$ = 11.206; p = 0.001 Skwiot *et al*., (2018)
Brighton Critera	No data	Inter-rater reliability ICC = 0.93 kappa = 0.84 (Juul-Kristensen *et al*., (2007)	For paediatric patients: GJH present in (28.3%) Of these 58.8% were female. Of the 34, 28 (82.4%) met both major criteria one of which is BS >4. (Bale, n.d.)
5pQ	No data	ICC = 0.92 (95% CI 0.85, 0.96) (Glans, 2020)	Sensitivity = 70.9% specificity = 77.4% ROC= 0.786 (de Moraes *et al*., 2011) Sensitivity =84%Specificity 89% HSD: Sensitivity=80%Specificity = 84%

			(Hakim and Grahame, 2003) presence of GJH (rho = 0.857; p<0.001) Bulbena *et al*., (2014)
Line Drawings	No data	k= 1.00; (95% CI 1.00-1.00) (Cooper *et al*., 2018)	Self-reported participant-repeatability k =0.91 (95% CI 0.74-1.00) Sensitivity of 0.87 (95% CI 0.81, 0.91) Specificity of 0.99 (95% CI 0.98, 1.00) (Cooper *et al*., 2018)
Screening Questionaire to Detect Joint Hypermobility	No data	ICC= 0.961 CI (0.922-0.980) (Bulbena *et al*., 2014)	GJH 44.3% PPV 92.79% NPV 46.22% (Bulbena, *et al*., 2014)
BS – self reported	ICC = 0.92 (95% CI 0.85, 0.96) (Glans *et al.*, 2020)	No data	Patients with FAI (Naal *et al*., 2014): Correlative relationship between BS-self and hip flexion (r = 0.61, p < 0.01), internal rotation (r = 0.56, p < 0.01), and external rotation (r = 0.44, p < 0.01) In pregnant women (Schlager *et al*., 2020): There is uncertainty in identifying generalized joint hypermobility in young women using the BS-self reported with a cut-off level of ≥2 when the Beighton score ≥ 5 is used as the reference test. The strength of the BS-self reported is to rule-out women without generalized joint hypermobility Cut off ≥2 AUC 0.73 (95% CI 0.67–0.79) Sensitivity 84.1% (95% CI 69.9–93.4) Specificity 61.9% (95% CI 55.6–67.8) The false-positive rate was 38% The PPV and NPV ranged between 13.4 and 42.4% and

			between 90.3 and 99.2%, respectively The LR+ increased by 2.2 when using a cut-off level of ≥2 for the self-reported 5PQ Paediatric population (Glans *et al*., 2020): Sensitivity = 91% (95% CI 72, 99%) Specificity = 75% (95% CI 69,72) 80%)ROC = 0.87 (95% CI 0.79, 0.95)
Marshal Test	No data	No data	PPV = 34% NPV = 99% Sensitivity = 99% Specificity = 67% Positive likelihood ratio = 3.3 (Cahil *et al*., 2019)
KRSP	No data	ICC 0.846- 0.986 (Ahn *et al*., 2019)	No data
Upper Limb Hypermobility Assessment Tool	No data	ICC2,1 = 0.92 (Chan and Nicholas, 2018)	The cut-point ≥7/12 (sensitivity 0.84, specificity 0.77, +LR 3.7, -LR 0.2) (Chan and Nicholas, 2018)
Lower limb Hypermobility Assessment Score	Spearman rho = 0.86 (p<0.0001) (Boyle *et al*., 2003)	Spearman rho = 0.87 (p<0.0001) (Boyle *et al.*, 2003)	Sensitivity = 67% Specificity =94% Positive Predictive Value (50%) Negative Predictive Value (97%) (Boyle *et al.*, 2003)
Hospital Del Mar	ICC=0.86 (95% CI 0.73-0.93) (Schlager *et al*., 2018)	Kappa = 0.81 Bevilaqua et al (2019) ICC = 0.81 (95% CI 0.69-0.89 (Schlager *et al*., 2018)	For presence of GJH (rho = 0.745; p<0.001) (Schlager *et al*., 2018)
Contompasis	ICC = 0.73-0.82 (Vallis *et al*., 2016) ICC = 0.79 (95% CI 0.57-0.90) (Schlager *et al*., 2018)	ICC = 0.58-0.62 (Vallis *et a*l., 2016) ICC = 0.82 (95% CI 0.69-0.89) (Schlager *et al*., 2018)	No data
Rotes-Querol	No data	Kappa = 0.31 – 0.80 (Juul-Kristensen *et al*., 2007) Kappa > 0.6 (Remvig *et al*., 2007)	No data
Parvaneh-Shiari Criteria	No data	No data	Sensitivity = 100%, specificity = 98%, positive predictive value= 100%, negative predictive values 98% Accuracy was 99% and balanced accuracy was 99%. The area under the ROC = 0.9 (Parvaneh and Shiari, 2016)

4.0 Discussion:

Systematic reviews, mainly included cohort studies, some included reviews of case control studies and were therefore given a rating of 3A instead of 2A. Narrative reviews rated 5A as per recommendations from the University of Canberra (no date) suggesting narrative reviews are to be considered as expert opinion.

The PRISMA (2020) checklist was used to evaluate both systematic reviews and narrative literature reviews to demonstrate any methodological flaws and group the reviews together.

Certain studies were designed to be association studies, rather than studies focusing specifically on clinimetric statistics and these were excluded from the review.

A statistical synthesis of results was not conducted due to heterogeneity.

Interpretation of reliability statistics can be problematic and non-straightforward (Koo and Li, 2016). Wide variation is reported for clinimetric statistics with no standardised reporting making direct comparisons difficult. This includes interpretation of effect sizes. Few papers addressed strength of association.

There was inconsistency of reporting inter-rater reliability, but not intra-rater reliability. Only some papers assessed validity and no papers assessed responsiveness.

Compared to the BS there is sparce literature on clinimetric properties of alternative scoring systems.

Intra-rater reliability scores for the BS range from ICC 0.74-0.99, K=0.82. Inter-rater reliability score ranged from ICC 0.72-0.98, spearman rho =0.86. For the other scoring systems assessed the score ranges are summarised in table 4.

Johnson, Wand and Simmonds (2019) found a strong correlation between LLAS and BS scores (rho =0.732; $p < 0.001$). The ULHAS had an intra-rater reliability ICC of 0.92 (Nicholson and Chan, 2018). No data is available on inter-rater reliability, or validity and so a direct comparison between ULHAS and the BS cannot be made.

Vallis, Wray and Smith (2016) Found both inter-rater reliability and intra-rater reliability of Contompasis inferior to those of BS (Inter-rater reliability ICC range 0.58-0.62 and intra-rater reliability ICC 0.73-0.82 versus Inter-rater reliability ICC range 0.72-0.80, intra-rater reliability ICC range 0.71-0.82), However Schlager et al., (2018) found both inter-rater reliability and intra-rater reliability of BS inferior to that of Contompasis (Inter-rater reliability 0.82 (95% CI 0.69-0.89 and Intra-rater reliability 0.79 (95% CI 0.57-0.90 vs Inter-rater reliability 0.72 (95% CI 0.55-0.83 and Intra-rater reliability 0.76 (95% CI 0.54-0.88, respectively). Both these studies were performed in healthy adult populations, with Vallis Wray and Smith (2016) selecting the sample from physiotherapy students at a university.

There was only a single paper for reliability statistics for the Rotes-Querol Score by Juul-Kristensen et al., (2007) reporting a Kappa range of 0.31-0.80. This range is lower than the BS.

Parvaneh and Shiari (2016) report a perfect sensitivity of 100% for their score which is essentially a modified BS. There are no reliability statistics available for this scoring system.

Bulbena *et al*., (2014) found the self-reported questionnaire positively correlates with both the Hospital del Mar criteria (rho = 0. 745; $p<0.001$) and the self-reporting questionnaire of Hakim and Grahame (2003) (rho = 0.857; $p<0.001$).

Shlager *et al*., (2020) found the Hospital del Mar Criteria superior to the BS with regards to both inter-rater and intra-rater reliability statistics (Inter-rater reliability 0.81 (95% CI 0.69-0.89) and Intra-rater reliability 0.86 (95% CI 0.73-0.93) versus Intra-rater reliability 0.76 (95% CI 0.54-0.88) and Intra-rater reliability 0.76 (95% CI 0.54-0.88) respectively. Using a cut off score of ≥4 Ohman Westblom and Hendrikkson (2014) found the Hospital Del Mar Criteria superior in its ability to detect GJH at 34% versus 12% for the BS.

No reliability data was available for the Sasche Scale. Skwiot *et al*., (2018) reported an ability to detect GJH in

56.5% of participants with a $\chi2(1) = 11.206$; $p = 0.001$ (no cut off score mentioned), however Mikołajczyk *et al*., (2012) reported the Sasche Scale's ability to detect GJH as significantly lower at 30% using a cut off score of ≥ 7/13. This large variance is likely to be due to differences in patient cohorts.

The Rotes-Querol scoring system is the only test that provides non dichotomous scores for hypermobility.

Calhil *et al*., (2021) found the Marshall test a good rule out test with specificity of 99%, but low sensitivity of 67%.

The BS is brought into question for identifying GJH in the H-EDS patient group by McGillis *et al*., (2021). This is the principle method for screening H-EDS when combined with the 5pQ as per the EDS (2017) diagnostic criteria. Joint hypermobility is regarded as the hallmark feature of H-EDS and other EDS subtypes, however authors McGillis et al., (2016), Hamonet and Brock (2015) and Malek et al., (2021) report many patients with EDS do not always score highly on the BS.

A reason not identified in current literature is the fact EDS is a disorder of collagen fragility (hence it's systemic nature), however hypermobility has traditionally been utilised as a proxy measure of tissue fragility. As discussed in the introduction, joint hypermobility, whilst related to tissue fragility, is not one in the same phenomenon.

This raises a question regarding nosology of HSD and whether this is truly a condition of hypermobility, or whether it is a disorder of tissue fragility and whether "tissue fragility spectrum disorders" is worth considering as alternative nomenclature, or whether this represents an entirely different subtype of collagen disorder.

An additional question arises about whether it is possible for some patients to have tissue fragility in the absence of GJH. Currently it is not possible to identify such patients who might be currently misclassified as HSD, or not classified at all.

Summary of Information by Patient Cohort (Study information in supplementary material E)		
Patient Cohort	**Scoring Systems Assessed**	**Summary**
Pregnant Women	**BS, 5PQ, BS-Self Reported, Extensometer**	• BS-Self Reported good rule out test for GJH Cut off ≥2 AUC 0.73 (95% CI 0.67–0.79) Sensitivity 84.1% (95% CI 69.9–93.4) specificity, 61.9% (95% CI 55.6–67.8) The false-positive rate was 38%. The PPV and NPV ranged between 13.4 and 42.4% and between 90.3 and 99.2%, respectively. The LR+ increased by 2.2 when using a cut-off level of ≥2 for the self-reported 5PQ • Relationship between increased score in 5PQ and pelvic girdle pain during pregnancy • BS not good for identifying BS later in pregnancy due to inability to conduct hand to floor test
Healthy Adults	**BS, Modified BS, KSRP, SQCH, 5PQ, Hospital Del Mar, Self Reported Line Drawings, CSES, ULHAT, Shoulder Hypermobility, Hospital Del Mar, Contompasis, Rote`s-Que´ rol, Brighton Criteria, FAI**	• The mean Corrected Extensibility Score was 23.84%/mm in participants found to be hypermobile versus 13.55%/mm in the normal mobility group • ULHAT has The cut-point ≥7/12 (sensitivity 0.84, specificity 0.77, +LR 3.7, -LR 0.2), Inter-rater reliability ICC2,1 = 0.92 • No validity statistics for Rotes Querol, Kappa for interrater reliability varied between 0.31 and 0.81 • Brighton Criteria Inter-rater reliability ICC = 0.93, kappa = 0.84 • Patients with FAI have a prevalence of GJH = 32.7 % (50 % of females and 24.3 % of males) if using BS cut off Cut off ≥4 • BS Inter-rater reliability >0.7 in all studies except Riley *et al*., (2020) who report Interrater reliability ICC =0.52 for intertester reliability at visit 1 and 0.86 at Visit 2. • BS Intra-rater reliability >0.7 in all studies

Athletes and Dancers	**BS, LLAS, Contompasis + BS + Brighton Criteria, KT arthrometer, Sasche's Criteria, Villefranche,**	• BS Inter-rater reliability ICC>0.7 in all studies, Intra-rater reliability>087, but only 2 studies looked at this statistic • GJH>60% using BS cut off ≥4 in all studies assessing prevalence • LLAS demonstrated GJH= 38% with cut off score ≥ 7/12 • The Villefranche Criteria identified more dancers with JHS/EDS-HT than the Brighton (84% vs 31%, p < 0.001)
Paediatrics	**BS, Modified BS, Hospital Del Mar, Marshall Test, LLAS, BS Self Reported in Swedish, Sasche Scale, 5PQ, Brighton Criteria, Parveneer and Shiari Score**	• Hospital Del Mar GJH ranges between 38% and 70%, Inter-rater reliability Kappa = 0.81PPV = 34% NPV = 99% Sensitivity = 99% Specificity = 67% Positive likelihood ratio = 3.3 • LLAS Inter-rater reliability ICC = 0.78 , Intra-rater reliability ICC = 0.90-0.98 • BS self reported sensitivity = 91% (95% CI 72, 99%) Specificity = 75% (95% CI 69,72) 80%)ROC = 0.87 (95% CI 0.79, 0.95) • Sasche Scale Cut off ≥ 7/13 (30% positive) • BS cut off Cut off ≥ 4 Prevalence of GJH ranged from 12% to 64.6% • BS Cut off ≥ 6 GJH ranged from 7.6% to 18.6% • Parveneh and Shiari Score Sensitivity = 100%, specificity = 98%, positive predictive value= 100%, negative predictive values 98% Accuracy was 99% and balanced accuracy was 99%. The area under the ROC = 0.99 • Brighton Score Prevalence of GJH in sample = 0.98% (in healthy sample) vs 28.3% in secondary paediatric care sample • BS Self-Reported (in Swedish) Intra-rater reliability ICC = 0.92 (95% CI 0.85, 0.96) , sensitivity = 91% (95% CI 72, 99%) Specificity = 75% (95% CI 69,72-80%)ROC = 0.87 (95% CI 0.79, 0.95) • Marshall Test PPV = 34% NPV = 99% Sensitivity = 99% Specificity = 67% Positive likelihood ratio = 3.3
EDS/HSD Patients	**Glenohumeral abduction + BS, Retraction of tendons, BS**	• 87.5% of sample had retraction of hamstrings • Retractions of the triceps surae were found in 90.9% of patients, and retractions of the soles of feet were observed in 95.9% of patients • The impact of retraction on the Beighton palms-on-the-floor test is very great indeed 97.8% of patients who present a retraction of the hamstrings of over 45° cannot perform this manoeuvre, which will affect the BS and potentially identification of the condition in this cohort • Skin hypermobility statistically significant for classical EDS patients • Only 46% HSD/LJH patients scored a BS of 4/9 as assessed by an EDS practitioner (SD = 2.12). In the H-EDS group the average BS was 6/9 (SD = 1.45).

Degrees of consensus around reliability and validity statistics were reported in systematic and narrative reviews with conflicting conclusions about the usefulness of the BS in detecting GJH.

The majority of reviews including Remvig *et al*., (2007), Juul-Kristensen *et al*., (2017), Malek *et al*., (2020) and Bokhorn *et al*., (2020) included a high number of association studies. The author determined many of these are not strictly related to clinimetric properties of the scoring systems analysed and therefore should not be included in a review primarily focused on reliability and validity statistics.

Remvig *et al*., (2007) conclude all scores are comparable, however report the Rotes-Querol has a lower Kappa value compared with other scoring systems assessed.

Malek *et al*., (2020) conclude use of the BS as a clinical diagnostic tool, particularly within the 2017 International Classification of EDS for the diagnosis of hypermobile EDS (hEDS), remains controversial.

Kristensen *et al*., (2017) conclude evidence supports the BS as a reliable clinical tool, however are concerned with a lack of research relating to validity. There is insufficient data to draw conclusions about the other scoring methods assessed in their review.

(Drabik *et al*., 2020 report GJH and H-EDS might be more accurately identified when combining the BS with other assessments.

Day, Koustekadis and Wyon (2011) reviewed hypermobility in dance including papers relating to dancers and non-dancers with a difference in detection rates between the Carter and Wilkinson method vs the Beighton Score, however these studies included different types of participants it is difficult to ascertain whether detection rates differ as a result of participant characteristics, or difference in the ability of scores to detect GJH.

Bokhorn *et al*., (2020) report the BS is excellent in terms of reliability. This review states publication bias and the lack of standardised reporting with use of either composite scores, individual measurements, or variable cut off scores could lead researchers to be influenced to choose the value with the highest level of agreement.

Lack of research to establish validity of the BS raised by Remvig *et al*., (2007), Malek *et al*., (2020) and Juul-Kristensen *et al*., (2017) should be addressed as a priority area in future research.

Most reviews agreed there is lack of high-quality studies, with the exception of the paper by Palmer *et al*., (2020) a different patient cohort from most papers (known EDS/HSD), hence only four high quality studies were included in their review.

Few papers are high quality. Methodological quality issues include failure to use, or report use of gold standard goniometry devices for assessment of ROM, non-representative patient samples, convenience sampling, non-blinding of participants and assessors, lack of appropriate statistical analysis and, or reporting and failure to compare diagnostic studies to the recommended best practice guidelines. Small sample size is potentially an issue in the studies. Many papers used sample sizes of less than 100, the average sample size ranged from 100-300. 6 studies included a sample size greater than 1000. A lack of blinding of participants and researchers is an additional area of concern. Less than a third of the studies reviewed included some sort of blinding in their methodology. Some studies did not discuss what cut-off score was used, whilst others did not explain the rationale of cut-off score choice.

In establishing diagnoses, clinicians should rely on evidence-based diagnostic methodology that is standardised according to the COSMIN 2010 criteria, satisfying 3 domains of validity, reliability and responsiveness (the test's ability to detect change over time) (Scholtes, Terwee and Poolman, 2011). None of the studies reviewed looks at all three of these domains.

Few papers referenced appropriate reporting, or test design standards such as CONSORT, STARD or QUADAS-2 (2011). Some studies were conducted prior to authorship of the COSMIN (2010) guidelines, however there is a general lack of referencing to other appropriate standards. Riley *et al*., (2020) reference the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).

As no standardised reporting of clinimetric statistics exists, comparison is challenging.

Only one randomised control trial (RCT) was identified: Ahn *et al*., (2019), however this only partially adhered to standards for conducting a RCT.

Specific patient characteristics possibly impacted results of the BS ability to detect GJH. For example, in athlete and dance cohorts. The results and analysis of a study by Riley *et al*., (2020) noted prevalence of GJH within a sample has an impact on the kappa values of this tool. As reported by Lijmer *et al*., (1999) diagnostic studies with methodological flaws, possibly over-estimate the accuracy of a test and falsely increase the pre-test probability.

A challenge in assessment of validity and reliability studies with regards to the BS, is the fact that most prevalence studies use the BS score itself to establish the presence of GJH. This introduces an inherent level of bias as the scoring method is essentially tested against itself.

This bias is likely to overestimate the sensitivity and specificity values, because these have been based on prevalence calculated using the same test. This is a source of major bias and flaw in study methodology in validity assessment of the BS. This source of bias in validity studies of diagnostic test accuracy is discussed by Worster and Carpenter (2008) and Key, Hall and Wang (2019). Incorporation bias would also impact the cut-off values for tests in different cohorts and is an area that requires serious review.

Paucity of literature exists on whether a combination of tests would improve the validity of scoring systems. It is possible the BS in conjunction with other methods yields a higher sensitivity compared with a single scoring system (Drabik Bys and Gawda 2020), (Nicholson and Chan, 2018) (Lamari, Chueire and Cordeirdo (2005).

Heterogeneity of Results

High heterogeneity as well as conflicting conclusions drawn by researchers is present.

The complexity of subject material and lack of standardisation of procedures and protocols in clinical practice and current literature are some reasons for heterogeneity.

Other causes of heterogeneity include:

- High variability in study design methodology (case-control, vs cohort study, vs trial vs review)
- Comparison of different range of scores (not all studies compared the same scores, not all studies compared the scores with the gold standard, or included a full clinical assessment)
- Use of different cut off score (most papers use between 4-5 in adult populations and 6-7 in paediatric papers)
- Failure to use goniometer
- Lack of standardised reporting methods
- Lack of standardisd statistical parameters for validity and reliability (kapha vs cronbach alpha, vs ICC)
- No standardised protocol for performing BS
- Wide differences in participant characteristics (children, vs pregnant women, vs adults, vs dancers, vs elite athletes)
- Wide differences in study size

Limitations and Bias in this Review

The single reviewer created possible selection bias and lack of use of systematic review software are the two most important limitations of this review. Other limitations include the search strategy, that did not use meta search terms and did not use the term “psychometric properties” when conducting the search.

The variety of non-standardised terminology potentially poses an issue, however no critical studies have been missed from the review that would alter findings due to a sufficiently broad search criteria and thorough manual search.

Other sources of bias and limitations of this paper include:

- Manual download of papers creates possible reproducibility issues
- Methodological quality of papers used in the review
- Did not use recommended software for conducting a systematic review
- Limitations of inclusion and exclusion criteria
- Broad inclusion criteria
- Variability of reported results
- Limited review of grey literature
- Limited review clinical practice guidelines
- Inclusion of some low-quality papers, grey literature and narrative reports
- English only publications
- Limited to 4 databases
- Only free papers were accessed

Areas for future research should focus on the following areas:

- Standardisation of protocols
- Evaluation between tip of thumb vs more of thumb approximation to forearm
- Consensus on cut off scores
- Lack of individual studies assessing BS specifically for COSMIN (2010) criteria for validation
- Explore alternative scores for patients with tight hamstrings -> hand to floor measurement (Cypel, 2019) (Hamonet and Brock, 2015)
- Explore alternative scores for patients with ligament laxity as opposed to joint hypermobility
- Whether combining BS with other scoring systems provides increased reliability and validity of findings
- The need to establish a consensus-based gold standard test (or combination of tests) for recognition of GJH
- Investigation of a standardised "global index of joint hypermobility"
- Appropriate study designs to further investigate and compare validity, reproducibility and reliability properties of all widely used scoring systems including questionnaires as these potentially have numerous benefits in a clinical context for the assessment of GJH, including reduced time spent assessing a patient without an examination and the ability to be conducted prior to appointments
- More studies on the Rotes Querol scoring system that measures degrees of hypermobility, rather than the currently dichotomous scoring systems
- A study to determine appropriate statistically derived cut-off scores for the general population vs subgroups such as professional dancers and athletes where a high pre-test probability is likely to affect the scoring systems utilised and affect interpretation of validity and reliability
- A study designed to assess joint ROM in patients with comorbid arthritis with adjusted scoring systems
- Consensus on ROM measurements for hypermobility
- Research into combined assessment methods for establishing a diagnosis of GJH vs single scoring systems
- Systematic review of instruments, software and devices for measurement of joint ROM
- Systematic review of novel associations between various clinical signs and GJH Standardised reporting methods in accordance with the many established guidelines on assessing diagnostic tests, such as COSMIN (2010)

Table 4 Summary of Findings
• The Beighton Score has demonstrated good intra-rater and inter-rater reliability • A lack of rigorous study design and methodological flaws prevents an evidence-based conclusion regarding the Beighton Score's validity in assessment of GJH • The literature is inadequate for determining whether the Beighton Score is Superior to other commonly used scoring systems in clinical use in detecting GJH • As there is a lack of literature on the validity of the BS, it should not be used to rule out a diagnosis of GJH. In patients where there is a high level of clinical suspicion who score <4 on the BS, additional referral +/- investigations for HSD should be carried out • There is need for improvement of methodological quality in studies designs to assess clinimetric properties of scoring systems • Future study designs eliminating incorporation and verification bias for the BS are required • Novel scoring systems require further evaluation to assess their clinimetric properties and usefulness in clinical practice • Cut off scores should be chosen with care according to the characteristics of the patient cohort being assessed • Patients who achieve a low value on any of the scoring systems discussed in this review, should still be assessed for clinical signs of collagen disorders and if these exist, referred or investigated in greater detail to avoid missing the diagnosis of EDS and other collagenopathies • Further research is urgently required in this important area of musculoskeletal medicine with areas of focus recommended in Table 4.9

5.0 Conclusion

The Beighton Score is the most commonly used score for identify GJH. This review compared reliability and validity statistics of BS to other scores for identifying BS.

The findings of this systematised review support those of previous similar reviews conducted by Juul-Kristensen *et al*., (2017), Bokhorn *et al*., (2020) and Malek *et al*., (2021) and the systematic and narrative reviews included in this study.

Lack of consensus around terminology and lack of precision of terms creates nosological and corresponding clinical dilemmas for researchers, clinicians and patients. In medical literature it creates difficulty identifying relevant literature. As per Castori *et al*., (2017) the author recommends future research use the term Generalised Joint Hypermobility (GJH) and Hypermobility Spectrum Disorders (HSD) and abandon other approximating terms.

Additionally, imprecise use of terms creates the risk that hypermobility is used as a proxy term for tissue fragility, (the true marker of Ehlers Danlos Syndrome and disorders of defective collagen) will potentially limit diagnosis. This is an area that requires exploration in future research.

The Beighton Score has acceptable inter-rater reliability and intra-rater reliability ranges reported in the literature with the lowest ICC reported as 0.72 and 0.74 (respectively). Questions still remain regarding its validity.

Incorporation bias is a major issue of concern not previously addressed in literature to date. As such, The BS should not be used to rule out a diagnosis of GJH. In patients where there is a high level of clinical suspicion who score <4 on the BS, further clinical evaluation is warranted to ensure diagnoses are not missed as delays, or missed diagnoses can have serious negative health outcomes for patients.

There is limited high quality research available hence it is not possible to draw a conclusion about whether other tools are more or less reliable compared to BS, or valid in comparison to the BS.

No large multicentre, blinded, randomised control trial has been conducted comparing the BS to other commonly used methods, or the alternative of combination scores, or global joint ROM index.

Further research is required to quantify validity and reliability of the BS and other scoring systems. Although recommended by systematic and narrative reviews in the past, there remains paucity of literature. This should be a priority as it creates significant challenges for patients who rely on scoring systems to achieve a diagnosis of HSD/EDS and other HDCT involving GJH.

6.0 Registration and protocol

This paper was not registered as it formed part of a thesis for the university of South Wales, UK.

7.0 Acknowledgment

I would like to thank my supervisor Dr Rizwan Rajak from the University of South Wales for his continued support while completing this project.

8.0 Conflict of Interest

The author received no financial, or other incentives in undertaking this review. The author has no conflict of interest to declare.

9.0 Availability of data and other materials

Appendices, additional data and other material can be obtained by contacting the authors at the email addresses provided.

10. References

Ahlqvist, K., Bjelland, E., Pingel, R., Schlager, A., Nilsson-Wikmar, L., Kristiansson, P., (2020) 'The Association of Self-Reported Generalized Joint Hypermobility with pelvic girdle pain during pregnancy: a retrospective cohort study', *BMC Musculoskeletal Disorders*, 21(1), pp. 474-484. Available at: https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-020-03486-w (Accessed: 11 December 2021)

Ahn, S. Y., Ko, H., Yoon, J. O., Cho, S. U., Park, J. H., & Cho, K. H. (2019) 'Determining the Reliability of a New Method for Measuring Joint Range of Motion Through a Randomized Controlled Trial', *Annals of Rehabilitation Medicine*, 43(6), pp. 707–719. Available at: https://doi.org/10.5535/arm.2019.43.6.707 (Accessed: 12 December 2021)

Akoglu, H., (2018) 'User's guide to correlation coefficients', *Turkish Journal of Emergency Medicine*, 18, pp. 91–93. Available at: https://doi.org/10.1016/j.tjem.2018.08.001 (Accessed: 16 December 2021)

Anchan, S., (2019) *Chidambaram, Thilai Nataraja Temple*. Available at: https://backpackersnovel.in/2019/10/28/chidambaram-thillai-natraja-temple/ (Accessed: 20 December 2021)

Antonio, D., Magalhaes, C., (2018) 'Survey on joint hypermobility in university students aged 18–25 years old', *Advances in Rheumatology*, 58(3), pp. 1-7. Available at: https://advancesinrheumatology.biomedcentral.com/articles/10.1186/s42358-018-0008-x#citeas (Accessed: 12 December 2021)

Armstrong, R., (2018) 'Relative Joint Contribution to Joint Hypermobility in Rugby Players, Netballers and Dancers: The Need for Careful Consideration of Lumbar Flexion'. *International Journal of Sports and Physical. Therapy*, 13, pp. 676–686. Available at: https://doi.org/10.26603/ijspt20180676 (Accessed: 11 December 2021)

Armstrong, R., Greig, M., (2018) 'The Beighton score as a predictor of Brighton criteria in sport and dance', *Physical Therapy in Sport: Official Journal of the Association of Chartered Physiotherapists in Sports Medicine*, 32, pp. 145-154. Available at: https://www.semanticscholar.org/paper/The-Beighton-score-as-a-predictor-of-Brighton-in-Armstrong-Greig/ec002f63523ca92712259066f44e4484d3a898b3 (Accessed: 12 December 2021)

Arneja, S., Leith, J., (2009) 'Review article: Validity of the KT-1000 knee ligament arthrometer', *Journal of Orthopaedic Surgery*, 17, pp. 77–79. Available at: https://doi.org/10.1177/230949900901700117 (Accessed: 16 December 2021)

Aslan, U., Çelik, E., Cavlak, U., Akdağ, B., (2006) 'Evaluation of interrater and intrarater reliability of Beighton and Horan Joint Mobility Index', *Fizyoterapi Rehabilitasyon*, 17(3), pp. 113-119. Available at: https://www.researchgate.net/publication/287350205_Evaluation_of_interrater_and_intrarater_reliability_of_Beighton_and_Horan_Joint_Mobility_Index (Accessed: 19 December 2021)

Aslam, S., Emmanuel, P. (2010) 'Formulating a researchable question: A critical step for facilitating good clinical research' *Indian journal of sexually transmitted diseases and AIDS*, *31*(1), pp. 47–50. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3140151/ (Accessed: 21 December 2021)

Bale, P., MacGregor, A., Armon, K., Easton, V., Bacon, H., Jerman, E., (2014) 'The Brighton Criteria Fail to Capture the Clinical Characteristics of Benign Joint Hypermobility Syndrome in Children: Data from the Bendy Study', *Rheumatology*, 53(1), pp. 167. Available at: https://doi.org/10.1093/rheumatology/keu125.010 (Accessed: 12 December 2021)

Ballenger, L., Moore-Clingenpeel, M., Oberle, E., (2021) 'Musculoskeletal ultrasound assessment in pediatric knee hypermobility: a case control study', *Paediatric Rheumatology Online Journal*. 19(1), pp. 58-66. Available at: https://ped-rheum.biomedcentral.com/articles/10.1186/s12969-021-00546-w (Accessed: 12 December 2021)

Balshem, H., Stevens, A., Ansari, M., Norris, S., Kansagara, D., Shamliyan, T., Chou, R., Chung, M., Moher, D., Dickersin, K., (2013) 'Finding Grey Literature Evidence and Assessing for Outcome and Analysis Reporting Biases When Comparing Medical Interventions: AHRQ and the Effective Health Care Program, Methods Guide for Effectiveness and Comparative Effectiveness Reviews', *Agency for Healthcare Research and Quality (US).* Available at: https://www.ncbi.nlm.nih.gov/books/NBK174882/ (Accessed: 16 December 2021)

Beighton, P., Solomon, L., Soskolne, C.L., (1973) 'Articular Mobility in an African Population', *Annals of Rheumatic Disease*, 32, pp. 413–418. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1006136/ (Accessed: 11 December 2021)

Bennink., L., Deekshithar, K., Deekshithar, J., Deekshithar, S., (2013) *Shiva's Karanas in the Temples of Tamil Nadu: the Natya Shastra in Stone*. Available at: https://www.asianart.com/articles/karanas/index.html (Accessed: 19 December 2021)

Bevilacqua, D., Maillard, S., Ferrari, J., (2019) 'Measuring Joint Hypermobility Using the Hospital Del Mar Criteria - A Reliability Analysis Using Secondary Data Analysis'. *Archives of Rheumatology & Arthritis Research*, 1(1), pp. 1-6. Available at: https://irispublishers.com/arar/fulltext/measuring-joint-hypermobility-using-the-hospital.ID.000502.php (Accessed: 12 December 2021)

Bockhorn, L., Vera, A., Dong, D., Delgado, D., Varner, K., Harris, J. (2021) 'Interrater and Intrarater Reliability of the Beighton Score: A Systematic Review', *Orthopaedic Journal of Sports Medicine*, 38, pp.2245-2250. Available at: https://www.semanticscholar.org/paper/Validation-of-a-self-reported-Beighton-score-to-in-Naal-Hatzung/c98b6340218344392426b6121e35ad8aa5677597 (Accessed: 14 December 2021)

Boyle, K.L., Witt, P., Riegger-Krugh, C., (2003) 'Intrarater and Interrater Reliability of the Beighton and Horan Joint Mobility Index', *Journal of Athletic Training*, 38, pp. 281–285. Available at: https://pubmed.ncbi.nlm.nih.gov/14737208/ (Accessed: 12 December 2021)

Brazzaventre, C., Celletti, C., Gobattoni, P., Santilli, V., Camerota, F., (2013) 'The Ehlers-Danlos syndrome: hystory of a clinical hendiadys', *Medicina nei Secoli*, 25(2), pp. 491-501. Available at: https://pubmed.ncbi.nlm.nih.gov/25807780/ (Accessed: 19 December 2021)

Bridges, A.J., Smith, E., Reid, J., (1992) 'Joint hypermobility in adults referred to rheumatology clinics', *Annals of Rheumatic Disease*, 51, pp793–796. Available at: https://doi.org/10.1136/ard.51.6.793 (Accessed: 11 December 2021)

Bulbena, A., Mallorquí-Bagué, N., Pailhez, G., Rosado, S., González, I., Blanch-Rubió, J., MD, Carbonell, J., (2014) 'Self-reported screening questionnaire for the assessment of Joint Hypermobility Syndrome (SQ-CH), a collagen condition, in a Spanish population', *European Journal of Psychiatry*, 28(1), Pp. 17-26. Available at: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-61632014000100002 (Accessed: 12 December 2021)

Cattalini, M., Khubchandani, R., Cimaz, R., (2015) 'When flexibility is not necessarily a virtue: a review of hypermobility syndromes and chronic or recurrent musculoskeletal pain in children', *BMC Paediatric Rheumatology*, Vol 13:4, Pp 1-15. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4596461/#__ffn_sectitle (Accessed: 11 December 2021)

Castori, M., Tinkle, B., Levy, H., Grahame, R., Malfait, F., Hakim, A., (2017) 'A framework for the classification of joint hypermobility and related conditions', *American Journal of Medical Genetics Part C Seminars in Medical Genetics*, 175, Pp. 148–157. Available at: https://doi.org/10.1002/ajmg.c.31539 (Accessed: 11 December 2021)

Cahill, S., Sharkey, M., Carter, C., (2021) 'Clinical assessment of generalized ligamentous laxity using a single test: is thumb-to-forearm apposition enough?', *Journal of Paediatric Orthopaedics B*, 1:30(3), pp. 296-300. Available at: https://pubmed.ncbi.nlm.nih.gov/32301823/ (Accessed: 12 December 2021)

Chan, C., Hopper, L., Zhang, F., Pacey, V., Nicholson, L., (2018) 'The prevalence of generalized and syndromic hypermobility in elite Australian dancers', *Physical Therapy in Sports*, 32, pp. 15-21. Available at: doi: 10.1016/j.ptsp.2018.02.001 (Accessed: 12 December 2021)

Cherni, Y., Desseauve, D., Decatoire, A., Veit-Rubinc, N., Begon, M., Pierre, F., Fradet, L., (2019) 'Evaluation of ligament laxity during pregnancy', *Journal of Gynecology Obstetetrics and Hum Reproduction*, 48(5), pp. 351-357. Available at: https://pubmed.ncbi.nlm.nih.gov/30794956/ (Accessed: 12 December 2021)

Cincinnati Children's Hospital Medical Center Joint Hypermobility Team (2014) *Evidence-based clinical care guideline for Identification and Management of Pediatric Joint Hypermobility, In children and adolescents aged*

4 to 21 years old, *CCHMC EBDM Website Guideline 43*. Available at: https://www.cincinnatichildrens.org/-/media/cincinnati%20childrens/home/service/j/anderson-center/evidence-based-care/recommendations/type/joint%20hypermobility%20guideline%2043 (Accessed: 15 December 2021)

Cooper, D.J., Scammell, B.E., Batt, M.E., Palmer, D., (2018) 'Development and validation of self-reported line drawings of the modified Beighton score for the assessment of generalised joint hypermobility', *BMC Medical Research Methodology*, 18(11), pp. 1-8. Available at: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-017-0464-8#citeas (Accessed: 12 December 2021)

Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN), (2010) *Checklists for Assessing Study Qualities*. Available at: https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/ (Accessed: 12 December 2021)

Critical Appraisal Skills Programme (2019) *CASP Diagnostic Study Checklis*t. Available at: https://casp-uk.net/casp-tools-checklists/ (Accessed: 19 December 2021)

Czaprowski, D., Kotwicki, T., Stoliński, L., (2012) 'Assessment of joint laxity in children and adolescents: a review of methods', *Journal of Orthopaedics Traumatology and Rehabilitation*, 14, pp. 407–420. Available at: https://doi.org/10.5604/15093492.1016368 (Accessed: 12 December 2021)

Czaprowski D, Kędra A, Pawłowska P, Kolwicz-Gańko A, Leszczewska J, Tyrakowski M. (2015), 'The examination of the musculoskeletal system based only on the evaluation of pelvic-hip complex muscle and trunk flexibility may lead to failure to screen children for generalized joint hypermobility', *PLoS One*,10(3), pp.1-12. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121360 (Accessed: 14 December 2021)

Corten, L., Ferguson, G., Smits-Engelsman, B., (2020) 'Does the Item 'Hands on Floor' Add Value to the Beighton Score in Identifying Joint Hypermobility?', *European Journal of Rheumatology,* 7, pp. 79–83. Available at: https://doi.org/10.5152/eurjrheum.2020.19185 (Accessed: 11 December 2021)

Cypel., D., (2019) 'Gleno-humeral abduction measurement in patients with Ehlers-Danlos syndrome', *Orthopaedics & Traumatology: Surgery & Research*, 105, pp. 287-290. Available at: https://www.researchgate.net/publication/331808395_Gleno-humeral_abduction_measurement_in_patients_with_Ehlers-Danlos_syndrome (Accessed: 11 December 2021)

Day, H., Koutedaki,s Y., Wyon, M., (2011) 'Hypermobility and dance: a review', *International Journal of Sports Medicine*, 32(7), pp. 485-489. Available at: https://pubmed.ncbi.nlm.nih.gov/21695667/ (Accessed: 14 December 2021)

De Groot, J., (2011) 'Verification problems in diagnostic accuracy studies: consequences and solutions', *BMJ*, 343, pp. 1-9. Available at: https://www.bmj.com/content/343/bmj.d4770.long (Accessed: 23 December 2021)

Deeks J., Bossuyt, P., Leeflang, M., Takwoingi, Y., Flemyng, E., Mellor, L., (2021a) *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*, Version 2, Cochrane Training. Available at: https://training.cochrane.org/handbook-diagnostic-test-accuracy/PDF/v2 (Accessed 26 June 2021)

Deeks, J., Higgins, J., Altman, D., (editors), (2021b) *Chapter 10: Analysing data and undertaking meta-analyses, In: Cochrane Handbook for Systematic Reviews of Interventions*, Version 6.2. Cochrane Training. Available at: https://training.cochrane.org/handbook/current/chapter-10 (Accessed: 12 December 2021)

Drabik, M., Bys, A., Gawda, P., (2020) 'Difficulties in diagnosis and selection of optimal diagnostic methods for hypermobile Ehlers Danlos Syndrome –literature review', *Medycyna Ogólna i Nauki o Zdrowiu*, 26(2), pp. 79-86. Available at: http://www.monz.pl/Difficulties-in-diagnosis-and-selection-of-the-optimal-diagnostic-methods-for-hypermobile,122300,0,2.html (Accessed: 15 December 2021)

Desfor, F., (2003) 'Assessment of Joint Hypermobility in Dancers', *Journal of Dance Science & Medicine*, 7(1), pp. 17-23. Available at: https://www.ingentaconnect.com/content/jmrp/jdms/2003/00000007/00000001/art00003 (Accessed: 11 December 2021)

The Ehlers Danlos Society, (2017*) EDS Diagnostics 2017*. Available at: https://www.ehlers-danlos.com/eds-diagnostics/ (Accessed: 11 December 2021)

Evans, A., Rome, K., Peet, L., (2012) 'The foot posture index, ankle lunge test, Beighton scale and the lower limb assessment score in healthy children: a reliability study', *Journal of Foot and Ankle Research*, 5(1), pp.1757-1146. Available at: https://www.semanticscholar.org/paper/The-foot-posture-index%2C-ankle-lunge-test%2C-Beighton-Evans-Rome/ec70e08fed83e2414acd0fb6002f7860b61855be (Accessed: 12 December 2021)

Farmer, A., Douthwaite, H., Gardiner, S., Aziz, Q., Grahame R. (2010) 'A novel in vivo skin extensibility test for joint hypermobility', *Journal of Rheumatology*, 37(7) pp. 1513-8.

Ferrari, J., Parslow, C., Lim, E.J., Hayward, A., (2005) 'Joint hypermobility: The use of a new assessment tool to measure lower limb hypermobility', Clinical and Experimental Rheumatology, 23, pp. 413–420. Available at: https://www.clinexprheumatol.org/article.asp?a=2612 (Accessed: 19 December 2021)

Foley, C., Killeen, O., (2019) 'Musculoskeletal anomalies in children with Down syndrome: an observational study', *Archives of Disease in Childhood*, 104(5), pp 482-487. Available at: https://pubmed.ncbi.nlm.nih.gov/30472668/ (Accessed: 11 December 2021)

Fraeulin, L., Holzgreve, F., Brinkbäumer, M., Dziuba, A., Friebe, D., Klemz, S., Schmitt, M., Tenberg, S., Mark, A. van, Maurer-Grubinger, C., Ohlendorf, D., (2020) 'Intra- and inter-rater reliability of joint range of motion tests using tape measure, digital inclinometer and inertial motion capturing', *PLOS ONE*, 15, pp. 1-15. Available at: https://doi.org/10.1371/journal.pone.0243646 (Accessed: 16 December 2021)

Glans, M., Humble, M.B., Elwin, M., Bejerot, M., (2020) 'Self-rated joint hypermobility: the five-part questionnaire evaluated in a Swedish non-clinical adult population', *BMC Musculoskeletal Disorders*, 21(174), pp. 1-8. Available at: https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-020-3067-1#citeas (Accessed: 12 December 2021)

Grahame, (2004), *Joint hypermobility syndrome: a complex constellation of symptoms*. Available at: https://www.medscape.com/viewarticle/537938 (Accessed: 11 December 2021)

Grahame, R., (2007) 'The need to take a fresh look at criteria for hypermobility', *Journal of Rheumatology*, 34, pp. 664–665. Available at: https://pubmed.ncbi.nlm.nih.gov/17407226/ (Accessed: 11 December 2021)

Hakim., A., Grahame, R., (2001) 'A simple questionnaire to detect hypermobility: An adjunct to the assessment of patients with diffuse musculoskeletal pain', *International Journal of Clinical Practice*, 57(3), pp.163-166. Available at: https://www.researchgate.net/publication/10779714_A_simple_questionnaire_to_detect_hypermobility_An_adjunct_to_the_assessment_of_patients_with_diffuse_musculoskeletal_pain (Accessed: 11 December 2021)

Hamonet, C., Brock, I., (2015) 'Joint mobility and Ehlers-Danlos syndrome (EDS), new data based on 232 cases', *Journal of Arthritis*, 4(148), pp. 1-5. Available at: http://claude.hamonet.free.fr/eng/eds_joint-mobility.htm (Accessed: 12 December 2021)

Hansen, A., Damsgaard, R., Kristensen, J., Bagger J., Remvig, L., (2002) 'Inter-examiner Reliability of Selected Tests for Hypermobility', *Journal of Orthopaedic Medicine*, 24(2), pp. 48-51. Available at: https://www-tandfonline-com.ergo.southwales.ac.uk/action/showCitFormats?doi=10.1080%2F1355297X.2002.11736164 (Accessed: 12 December 2021)

Heidbreder, A., Ringelstein, E., Dittrich, R., Nabavi, D., Metze, D., Kuhlenbäumer, G., (2008) 'Assessment of skin extensibility and joint hypermobility in patients with spontaneous cervical artery dissection and Ehlers-Danlos syndrome', *Journal of Clinical Neuroscience*, 15(6), pp. 650-653. Available at: https://pubmed.ncbi.nlm.nih.gov/18400501/ (Accessed: 12 December 2021)

Hornberger, B., Sneha, R., Eds: Gooneratne, N., Yerrabolu, G., Chen, K., (2020) *Designing Inclusion and Exclusion Criteria in Writing Biomedical Research Proposals: A Concise Guide,* Version 1. Available at: https://repository.upenn.edu/crp/1 (Accessed: 16 December 2021)

Huddad., C., Rosno, J., Prasad, V., (2013) 'When Research Evidence is Misleading', *AMA Journal of Medical Ethics*, 15(1), pp. 29-33. Available at: https://journalofethics.ama-assn.org/article/when-research-evidence-misleading/2013-01 (Accessed: 21 December 2021)

Ionnidis, J., (2005), 'Why Most Published Research Findings Are False', *PLOS Medicine*, 294(2), pp. 218-228. Available at: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124 (Accessed: 21 December 2021)

Izatt, M.T., Bateman, G.R., Adam, C.J., (2012) 'Evaluation of the iPhone with an acrylic sleeve versus the Scoliometer for rib hump measurement in scoliosis', *Scoliosis*, 7, pp. 14-22. Available at: https://doi.org/10.1186/1748-7161-7-14 (Accessed: 16 December 2021)

Jansson, A., Saartok, T., Werner, S., Renström, P., (2004) 'General joint laxity in 1845 Swedish school children of different ages: age- and gender-specific distributions', *Acta Paediatrica*, 93(9) pp. 1202-6. Available at: https://www.google.com/search?client=firefox-b-d&q=Acta+Paediatrics (Accessed: 12 December 2021)

Jha, B., Ross, M., Reeves, S., Couzens, G., Peters, S., (2015) 'Measuring thumb range of motion in first carpometacarpal joint arthritis: The inter-rater reliability of the Kapandji Index versus goniometry, *Hand Therapy*, 2(21), pp 45-53. Available at: https://www.researchgate.net/publication/284175959_Measuring_thumb_range_of_motion_in_first_carpometacarpal_joint_arthritis_The_inter-rater_reliability_of_the_Kapandji_Index_versus_goniometry (Accessed: 22 December 21)

Joanna Briggs Institute., (2013), *JBI Levels of Evidence*. Available at: https://jbi.global. (Accessed: 11 December 2021)

Jobling, R., D'Souza, R., Baker, N., Lara-Corrales, I., Mendoza-Londono, R., Dupuis, L., Savarirayan, R., Ala-Kokko, L., Kannu, P., (2014) 'The collagenopathies: review of clinical phenotypes and molecular correlations', *Current Rheumatology Reports*, 16(1) pp. 394-407. Available at: https://www.researchgate.net/publication/259336349_The_Collagenopathies_Review_of_Clinical_Phenotypes_and_Molecular_Correlations (Accessed: 11 December 2021)

Johnson, P., Ward, S., Simmonds, J., (2019) 'The Lower Limb Assessment Score: A valid measure of hypermobility in elite football?', Physical Therapy in Sports, 37, pp. 86-90. Available at: https://pubmed.ncbi.nlm.nih.gov/30904749/ (Accessed: 12 December 2021)

Junge, T., Jespersen, E., Wedderkopp, N. Juul-Kristensen, B., (2013) 'Inter-tester reproducibility and inter-method agreement of two variations of the Beighton test for determining Generalised Joint Hypermobility in primary school children', *BMC Pediatrics*, 13(214), pp. 1-8. Available at: https://bmcpediatr.biomedcentral.com/articles/10.1186/1471-2431-13-214#citeas (Accessed: 12 December 2021)

Juul-Kristensen, B., Schmedling, K., Rombaut, L., Lund, H., Engelbert, R., (2017) 'Measurement properties of clinical assessment methods for classifying generalized joint hypermobility—A systematic review', *American Journal of Medical Genetics Part C Seminars in Medical Genetics*, 175, pp 116–147. Available at: https://doi.org/10.1002/ajmg.c.31540 (Accessed: 12 December 2021)

Juul-Kristensen, B., Røgind, H., Jensen, D., Remvig, L., (2007) 'Inter-examiner reproducibility of tests and criteria for generalized joint hypermobility and benign joint hypermobility syndrome', *Rheumatology (Oxford)*, 46(12) pp. 1835-1841. Available at: https://pubmed.ncbi.nlm.nih.gov/18006569/ (Accessed: 12 December 2021)

Juul-Kristensen, B., Schmedling, K., Rombaut, L., Lund, H., Engelbert, R.H.H., (2017) 'Measurement properties of clinical assessment methods for classifying generalized joint hypermobility—A systematic review', *American Journal of Medical Genetics Part C Seminars in Medical*, 175, pp. 116–147. Available at: https://doi.org/10.1002/ajmg.c.31540. (Accessed: 11 December 2021)

Kacheria, T., Drangsholt, M., Leresche, L., Winocur, E., (2006) 'Accuracy of a Shortened Beighton Examination Method for Hypermobility', *International Association for Dental Research Poster session, 2006 AADR/CADR Annual Meeting*, March 2006, Orlando, Florida, AADR/CADR. Available at: https://iadr.abstractarchives.com/abstract/2006Orld-74750/accuracy-of-a-shortened-beighton-examination-method-for-hypermobility (Accessed: 15 December 2021)

Kansas City Joint Experts, (2020) *The Importance of Joint Stability*. Available at: https://fitmjc.com/2020/01/27/the-importance-of-joint-stability-kansas-city-joint-experts/ (Accessed: 16 December 2020)

Kea, B., Hall, M. K., Wang, R. (2019) 'Recognising bias in studies of diagnostic tests part 2: interpreting and verifying the index test' *Emergency Medicine Journal*, 36(8), 501–505. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6693499/ (Accessed: 21 December 2021)

Keller, M.C., Hurschler, C., Schwarze, M., (2021) 'Experimental evaluation of precision and accuracy of RSA in the lumbar spine', *European Spine Journal*, 30, pp. 2060–2068. Available at: https://doi.org/10.1007/s00586-020-06672-5 (Accessed: 16 December 2021)

Knight, I., McCormack, M., Bird, H., (2012), *Managing Joint Hypermobility – A Guide for Dance Teachers*. Available at: https://www.onedanceuk.org/wp-content/uploads/2018/05/Managing-Hypermobility-for-the-Dance-Teacher-FINAL.pdf (Accessed: 11 December 2021)

Koo, T.K., Li, M.Y., (2016), 'A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability', *Journal of Chiropractic Medicine*, 15, pp. 155–163. https://doi.org/10.1016/j.jcm.2016.02.012 (Accessed: 16 December 2021)

Korovessis, P., Petsinis, G., Papazisis, Z., Baikousis, A., (2001) 'Prediction of Thoracic Kyphosis Using the Debrunner Kyphometer', *Journal of Spinal Disorders*, 14, pp. 67–72. Available at: https://doi.org/10.1097/00002517-200102000-00010 (Accessed: 16 December 2021)

Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B., Hrobjartsson, A., Roberts, C., Shoukri M, Streiner DL., (2011) 'Guidelines for reporting reliability and agreement studies (GRRAS) were Proposed', International Journal of Nursing Studies 48, pp. 661–71. Available at: https://pubmed.ncbi.nlm.nih.gov/21514934/ (Accessed: 20 December 2021)

Kulik., O., Gebska., M., (2018) 'Comparison of Beighton score and Brighton Criterion in order to diagnose joint hypermobility in children', *Journal of Education, Health and Spor*t, 8(5), pp. 133-148. Available at: http://ojs.ukw.edu.pl/index.php/johs/article/view/5472 (Accessed: 12 December 2021)

Kumar, B., Lenert, P., (2017) 'Joint Hypermobility Syndrome: Recognizing a Commonly Overlooked Cause of Chronic Pain', *American Journal of Pain*, 130(6), pp. 640-647. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0002-9343(17)30220-6 (Accessed: 11 December 2021)

Kwon, J. W., Lee, W. J., Park, S. B., Kim, M. J., Jang, S. H., Choi, C. K., (2013) 'Generalized joint hypermobility in healthy female koreans: prevalence and age-related differences', *Annals of Rehabilitation Medicine*, 37(6), pp. 832–838. Available at: https://doi.org/10.5535/arm.2013.37.6.832 (Accessed: 12 December 2021)

Lamari, N., Chueire, A., Cordiero, J., (2005) 'Analysis of joint mobility patterns among preschool children', *Sau Paulo Medical Journal*, 123(3), pp. 119-123. Available at: https://www.scielo.br/j/spmj/a/hZPMtBm9H4dJsM7chJbbNnH/?lang=en (Accessed: 12 December 2021)

Libre Texts, (2020), *8.4D Stability and Range of Motion at Synovial Joints*. Available at: https://med.libretexts.org/@go/page/7520 (Accessed: 16 December 2021)

Lidova, N., (2014), *Natyashastra*, Oxford Bibliographies. Available at: https://www.oxfordbibliographies.com/view/document/obo-9780195399318/obo-9780195399318-0071.xml (Accesssed: 19 December 2021)

Lijmer, J.G., Mol, B.W., Heisterkamp, S., Bonsel, G.J., Prins, M.H., van der Meulen, J.H.P., Bossuyt, P.M.M., (1999) 'Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests', *Journal of the American Medical Association*, 282, pp. 1061–1066. Available at: https://doi.org/10.1001/jama.282.11. (Accessed: 16 December 2021)

Lucas, NP., Macaskill, P., Irwig, L., Bogduk N., (2010) 'The development of a quality appraisal tool for studies of diagnostic reliability (QAREL)', *Journal of Clin Epidemiology,* 63, pp. 854–61. Available at: https://pubmed.ncbi.nlm.nih.gov/20056381/ (Accessed: 20 December 2021)

Lubberts, B., Guss, D., Vopat, B.G., Johnson, A.H., van Dijk, C.N., Lee, H., DiGiovanni, C.W., (2020) 'The arthroscopic syndesmotic assessment tool can differentiate between stable and unstable ankle syndesmoses' *Knee Surgery, Sports Traumatology, Arthroscopy,* 28, pp. 193–201. Available at: https://doi.org/10.1007/s00167-018-5229-3 (Accessed: 15 December 2021)

Macdonell, A., (1917), *A Vedic Reader for Students, Containing Thirty Hymns of the Rigveda in the original Samhita and Pada Texts, with Transliteration, Translation, Explanatory Notes, Introduction, Vocabulary*, Oxford: Clarendon Press, UK. Available at: https://oll.libertyfund.org/title/macdonell-a-vedic-reader-for-students (Accessed: 19 December 2021)

Malek, S., Reinhold, E.J., Pearce, G.S., (2021) 'The Beighton Score as a measure of generalised joint hypermobility', *Rheumatology International*, Vol 41, pp. 1707–1716. Available at: https://doi.org/10.1007/s00296-021-04832-4 (Accessed: 11 December 2021)

Malfait, F., Francomano, C., Byers, P., Belmont, J., Berglund, B., Black, J., Bloom, L., Bowen, J.M., Brady, A.F., Burrows, N.P., Castori, M., Cohen, H., Colombi, M., Demirdas, S., Backer, J.D., Paepe, A.D., Fournel-Gigleux, S., Frank, M., Ghali, N., Giunta, C., Grahame, R., Hakim, A., Jeunemaitre, X., Johnson, D., Juul-Kristensen, B., Kapferer-Seebacher, I., Kazkaz, H., Kosho, T., Lavallee, M.E., Levy, H., Mendoza-Londono, R., Pepin, M., Pope, F.M., Reinstein, E., Robert, L., Rohrbach, M., Sanders, L., Sobey, G.J., Damme, T.V., Vandersteen, A., Mourik, C. Voermans, N., Wheeldon, N., Zschocke, J., Tinkle, B., (2017) 'The 2017 international classification of the Ehlers–Danlos syndromes', *American Journal of Medical Genetics Section C Seminars in Medical Genetics*, 175, pp. 8–26. Available at: https://doi.org/10.1002/ajmg.c.31552 (Accessed: 11 December 2021)

McCormack, M., Briggs, J., Hakim, A., Grahame, R., (2004) 'Joint laxity and the benign joint hypermobility syndrome in student and professional ballet dancers', *Journal of Rheumatology*, 31(1), pp. 173-178. Available at: https://pubmed.ncbi.nlm.nih.gov/14705238/ (Accessed: 12 December 2021)

McGillis, L., Mittal, N., Santa Mina, D., So, J., Soowamber, M., Weinrib, A., Soever, L., Rozenberg, D., Liu, L., Tse, Y., Katz, J., Charames, G., Murphy, K., Vadas, P., Slepian, M., Walsh, S., Wilson, L., Adler, A., Franzese, A., Hussey, L., Nevay, D., Guzman, J., Clarke, H., (2019) 'Utilization of the 2017 diagnostic criteria for hEDS by the Toronto GoodHope Ehlers-Danlos syndrome clinic: A retrospective review', *American Journal of Medical Genetics Part A*, 182(3), pp. 484-492. Available at: https://www.researchgate.net/publication/337955567_Utilization_of_the_2017_diagnostic_criteria_for_hEDS_by_the_Toronto_GoodHope_Ehlers-Danlos_syndrome_clinic_A_retrospective_review (Accessed: 12 December 2021)

McHugh, M.L., (2012) 'Interrater reliability: the kappa statistic', *Biochemica Medica*, 22, pp. 276–282. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/ (Accessed on 16 December 2021)

Meyer, K. J., Chan, C., Hopper, L., Nicholson, L., (2017) 'Identifying lower limb specific and generalised joint hypermobility in adults: validation of the Lower Limb Assessment Score', *BMC Musculoskeletal Disorders*, 18(1), Pp. 514-523. Available at: https://doi.org/10.1186/s12891-017-1875-8 (Accessed: 12 December 2021)

Mejia-Hernandez, K., Chang, A., Eardley-Harris, N., Jaarsma, R., Gill, T.K., McLean, J.M., (2018) 'Smartphone applications for the evaluation of pathologic shoulder range of motion and shoulder scores—a comparative study', *Journal of Shoulder and Elbow Surgery Open Access*, 2, pp. 109–114. Available at: https://doi.org/10.1016/j.jses.2017.10.001 (Accessed: 11 December 2021)

Mikołajczyk, E., Szczygieł, A., Jankowicz-Szymańska, A., Jankowicz-Szymańska, A., Kopf, P., (2012) 'The occurrence of join hypermobility syndrome (JHS) in 15-year-old girls and boys in the context of diagnostic, therapeutic and prophylactic problems', *Fizjoterapia Polska*, 12, pp. 229-240. Available at: https://www.researchgate.net/publication/272893729_The_occurrence_of_join_hypermobility_syndrome_JHS_in_15-year_old_girls_and_boys_in_the_context_of_diagnostic_therapeutic_and_prophylactic_problems (Accessed: 12 December 2021)

Moraes, D., Baptista, C., Crippa, J., Louzada-Junior, P., (2011) 'Translation into Brazilian Portuguese and validation of the five-part questionnaire for identifying hypermobility, *Revista Brasileira de Reumatologia*, (1), pp. 53-69. Available at: https://pubmed.ncbi.nlm.nih.gov/21412606/. (Accessed: 12 December 2021)

Morishita, K., Petty, R., (2011) 'Musculoskeletal Manifestations of Mucopolysaccharidoses', Rheumatology, 50, 5, pp. 19–25. Available at: https://doi.org/10.1093/rheumatology/ker397 (Accessed: 11 December 2021)

Moromizato, K., Kimura, R., Fukase, H., Yamaguchi, K., Ishida, H., (2016) 'Whole-body patterns of the range of joint motion in young adults: masculine type and feminine type', *Journal of Physiological Anthropology*, 35 (23), pp 1-12. Available at: https://doi.org/10.1186/s40101-016-0112-8 (Accessed: 11 December 2021)

Morris, S., O'Sullivan, P., Murray, K., Bear, N., Hands, B., Smith, A., (2017) 'Hypermobility and Musculoskeletal Pain in Adolescents', *Journal of Paediatrics*, 181, pp. 213-221.e1. Available at: https://pubmed.ncbi.nlm.nih.gov/27863848/ (Accessed: 12 December 2021)

Mourcou, Q., Fleury, A., Diot, B., Franco, C., Vuillerme, N., (2015) 'Mobile Phone-Based Joint Angle Measurement for Functional Assessment and Rehabilitation of Proprioception', *BioMed Research International*, 2015, pp. 1-15. Available at: https://doi.org/10.1155/2015/3 (Accessed: 16 December 2021)

Naal, F., Hatzung, G., Müller, A., Impellizzeri, F., Leunig, M., (2014) 'Validation of a self-reported Beighton score to assess hypermobility in patients with femoroacetabular impingement', *International Orthopaedics*, 38(11), pp. 2245-2250. Available at: https://pubmed.ncbi.nlm.nih.gov/24993650/ (Accessed: 12 December 2021)

Nassery, Segal and Nassery (2016), *PCORI Methodology Standards: Academic Curriculum*. Available at: https://www.pcori.org/sites/default/files/PCORI-Methodology-Standards-Curriculum-Diagnostic-Tests-3.pdf (Accessed: 16 December 2021)

Nijs, J., Aerts, A., De Meirleir, K., (2006) 'Generalized joint hypermobility is more common in chronic fatigue syndrome than in healthy control subjects', *Journal of Manipulative Physical Therapy*, 29(1), pp. 32-9. Available at: doi: 10.1016/j.jmpt.2005.11.004 (Accessed: 16 December 2021)

Öhman, A., Westblom, C., Henriksson, M., (2014) 'Hypermobility among school children aged five to eight years: the Hospital del Mar criteria gives higher prevalence for hypermobility than the Beighton score', *Clinical and Experimental Rheumatology*, 32, pp. 285-290. Available at: https://www.scirp.org/%28S%28lz5mqp453edsnp55rrgjct55%29%29/reference/referencespapers.aspx?referenceid=1467436 (Accessed: 12 December 2021)

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, 371 (71), pp. 1-9. Available at: https://doi.org/10.1136/bmj.n71 (Accessed: 11 December 2021)

Palmer, S., Denner, E., Riglar, M., Scannell, H., Webb, S., Young, G., (2021) 'Quantitative measures of tissue mechanics to detect hypermobile Ehlers-Danlos syndrome and hypermobility syndrome disorders: a systematic review', *Clinical Rheumatology* (3), pp. 715-725. Available at: https://doi.org/10.1007/s10067-020-04939-2 (Accessed: 14 December 2021)

Palmer, S., Terry, R., Rimes, K., Clark, C., Simmonds, J., Horwood, J., (2016) 'Physiotherapy management of joint hypermobility syndrome--a focus group study of patient and health professional perspectives' , *Physiotherapy,* pp. 93-102. Available at: https://onlinelibrary.wiley.com/doi/10.1002/msc.1091 (Accessed: 11 December 2021)

Parveneh, V., Shiari, R., (2016) 'Proposed modifications to Beighton criteria for the diagnosis of joint hypermobility in children', *Indian Journal of Rheumatology*, 11(2), pp. 97-100. Available at: https://www.indianjrheumatol.com/article.asp?issn=0973-3698;year=2016;volume=11;issue=2;spage=97;epage=100;aulast=Parvaneh;type=0 (Accessed: 11 December 2021)

Parvaneh, V., Shahvaladi, H., Rahmani, K., Yekta, S., Gorji, F., Shiari, R., Abdollahimajd, F., (2020) 'Correlation between benign joint hypermobility syndrome and primary focal hyperhidrosis in children: a novel concept', *BMC Musculoskeletal Disorders*, 21(1), pp. 268-274. Available at: https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-020-03264-8 (Accessed: 11 December 2021)

Patel, N., Riddlestone, P., Gunalan, R., Gough, A., Metcalfe, J., Holroyd, B., (2017) 'Reliability of Beighton's Score in a Pediatric Population', *International Journal of Paediatric Orthopaedics*, 3(1), pp. 9-11. Available at: http://ijpoonline.com/reliability-of-beightons-score-in-a-pediatric-population/ (Accessed: 22 June 2021)

Phan, K., Nicholson, L., Hiller, C., Chan, C., (2020) 'Prevalence and unique patterns of lower limb hypermobility in elite ballet dancers', *Physical Therapy in Sport*, 41, pp. 55-63. Available at: https://pubmed.ncbi.nlm.nih.gov/31759239/ (Accessed: 22 June 2021)

Pearsall, A., Kovaleski, J., Heitman, R., Gurchiek, L., Hollis, J., (2006) "The relationships between instrumented measurements of ankle and knee ligamentous laxity and generalized joint laxity', *Journal of Sports Medicine and Physical Fitness*, 46(1), pp. 104 110. https://pubmed.ncbi.nlm.nih.gov/16596107/ (Accessed: 12 December 2021)

Pratt, A., Ball, C., (2016) 'What are we measuring? A critique of range of motion methods currently in use for Dupuytren's disease and recommendations for practice', *BMC Musculoskelealt Disordorders,* vol 17, pp. 20-31. Available at: https://doi.org/10.1186/s12891-016-0884-3 (Accessed: 11 December 2021)

Raff, M., Byers, P., (1996) 'Joint Hypermobility Syndromes', *Current Opinions in Rheumatology*. 1996 Sep;8(5), pp. 459-66. Available at: https://pubmed.ncbi.nlm.nih.gov/8941450/ (Accessed: 11 December 2021)

Rethlefsen, M., Kirtely S., Waffenschmidt, S., Ayala A., Moher, D., Page, M., Koffel, J., 2021, *PRISMA S: An Extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews*, Rethlefsen ML, Kirtley S, Available at: http://prisma-statement.org/ (Accessed: 19 December 2021)

Remvig, L., Jensen, D., Ward, R., (2007) 'Epidemiology of general joint hypermobility and basis for the proposed criteria for benign joint hypermobility syndrome: review of the literature', *Journal of Rheumatology*, 34(4), Pp.804-9. Available at: https://pubmed.ncbi.nlm.nih.gov/17407233/ (Accessed: 14 December 2021)

Remvig, L., Flycht, L., Christensen, K., Juul-Kristensen, B., (2014) 'Lack of Consensus on Tests and Criteria for Generalized Joint Hypermobility, Ehlers-Danlos Syndrome: Hypermobile Type and Joint Hypermobility Syndrome', *American Journal of Medical Genetics Section A*, 164, pp. 591-596. Available at: https://pubmed.ncbi.nlm.nih.gov/24464988/ (Accessed: 11 December 2021)

Rodgers, K., Gui, J., Dinulos, M.., Chou, R., (2017) 'Ehlers-Danlos syndrome hypermobility type is associated with rheumatic diseases', *Scientific. Reports*, 7, pp. 1-7. Available at: https://doi.org/10.1038/srep39636 (Accessed: 11 December 2021)

Rombaut, L., Deane, J., Simmonds, J., De Wandele, I., De Paepe, A., Malfait, F., Calders, P., (2015) 'Knowledge, assessment, and management of adults with joint hypermobility syndrome/Ehlers-Danlos syndrome hypermobility type among Flemish physiotherapists', *Americal Journal of Medical Genetics Part C Seminars in Medical Genetics*, 169., pp. 76-83. Available at: doi: 10.1002/ajmg.c.31434. (Accessed: 11 December 2021)

Romeo, D., Lucibello, S., Musto, E., Brogna, C., Ferrantini, G., Velli, C., Cota, F., Ricci, D., Mercuri, E., (2016) 'Assessing Joint Hypermobility in Preschool-Aged Children', *Journal of Pediatrics*, 176, pp. 162-166. Available at: https://pubmed.ncbi.nlm.nih.gov/27339250/ (Accessed: 12 December 2021)

Riley, S., Grimes, J., Calandra, K., Foster, K., Peet, M., Walsh, M., (2020) 'Reliability of the beighton score and impact of generalized joint mobility and resting scapular position on Median Neurodynamic Test 1', *Journal of Bodywork and Movement Therapy,* 24(4), pp. 131-137. https://pubmed.ncbi.nlm.nih.gov/33218501/ (Accessed: 12 December 2021)

Russek, L., Stott, P., Simmonds, J., (2019) 'Recognizing and Effectively Managing Hypermobility-Related Conditions', *Physical Therapy*, 99(9), pp. 1189-1200. Available at: https://pubmed.ncbi.nlm.nih.gov/31158283/ (Accessed: December 2021)

Russo, R.R., Burn, M.B., Ismaily, S.K., Gerrie, B.J., Han, S., Alexander, J., Lenherr, C., Noble, P.C., Harris, J.D., McCulloch, P.C., (2017) 'Is digital photography an accurate and precise method for measuring range of motion of the hip and knee?', *Journal of Experimental. Orthopaedics*, 4, pp. 29-37. Available at: https://doi.org/10.1186/s40634-017-0103-7 (Accessed: 16 December 2021)

Sanches, S.H.B., Osório, F. de L., Udina, M., Martín-Santos, R., Crippa, J.A.S., (2012) 'Anxiety and joint hypermobility association: a systematic review', *Brazilian Journal of Psychiatry*, 34, pp. 53–60. Available at: https://doi.org/10.1590/S1516-44462012000500005 (Accessed: 16 December 2021)

Sanders, S., Doust, J., Glasziou, P., (2015) 'A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment', *PLoS One*, 10(6), pp. 1-25. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4454557/ (Accessed: 23 December 2021)

Santos, C.M. dos, Ferreira, G., Malacco, P.L., Sabino, G.S., Moraes, G.F. de S., Felício, D.C., (2012) 'Intra and inter examiner reliability and measurement error of goniometer and digital inclinometer use', *Revista Brasileira de Medicina do Esporte*, 18, pp. 38–41. Available at: https://doi.org/10.1590/S1517-86922012000100008 (Accessed: 16 December 2021)

Sauers, E., Borsa P., Herling, D., Stanley R., (2001) 'Instrumented measurement of glenohumeral joint laxity and its relationship to passive range of motion and generalized joint laxity', *American Journal of Sports Medicine*, 29(2) pp. 143-50. Available at: https://pubmed.ncbi.nlm.nih.gov/11292038/ (Accessed: December 2021)

Schlager, A., Ahlqvist, K., Pingel, R., Nilsson-Wikmar, L., Olsson, C., Kristiansson, P., (2020) 'Validity of the self-reported five-part questionnaire as an assessment of generalized joint hypermobility in early pregnancy', *BMC Musculoskeletal Disorders*, 3;21(1), pp. 514-515. Available at: https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/s12891-020-03524-7 (Accessed: 11 December 2021)

Schlager, A., Ahlqvist, K., Rasmussen-Barr, E., Bjelland, E., Pingel, R., Olsson, C., Nilsson-Wikmar, L., Kristiansson, P., (2018) "Inter- and intra-rater reliability for measurement of range of motion in joints included in three hypermobility assessment methods", BMC Musculoskelet Disorders. 19(1), pp. 376-386. Available at: https://pubmed.ncbi.nlm.nih.gov/30326873/ (Accessed: 12 December 2021)

Scholtes V., Terwee, C., Poolman, R., (2011) 'What makes a measurement instrument valid and reliable?', *Injury*, 42(3), pp 236-40. Available at: https://pubmed.ncbi.nlm.nih.gov/21145544/ (Accessed: 11 December 2021)

Sein, M.L., Appleyard, R.C., Walton, J.R., Bradley, T., Murrell, G.A., (2008) 'Reliability of a new shoulder laxometer to assess inferior glenohumeral joint translation', *British Journal of Sports Medicine*, 42, pp. 178-182. http://dx.doi.org/10.1136/bjsm.2006.033472 (Accessed: 15 December 2021)

Shamsi, M., Mirzaei, M., Khabiri, S.S., (2019) 'Universal goniometer and electro-goniometer intra-examiner reliability in measuring the knee range of motion during active knee extension test in patients with chronic low back pain with short hamstring muscle', *BMC Sports Science and Medical Rehabiltation*, 11, 4, pp. 1-5. https://doi.org/10.1186/s13102-019-0116-x (Accessed: 11 December 2021)

Singh, H., McKay, M., Baldwin, J., Nicholson, L., Chan, C., Burns, J., Hiller, C.E., (2017) 'Beighton scores and cut-offs across the lifespan: cross-sectional study of an Australian population', *Rheumatology*, Oxford, 56, pp. 1857–1864. Available at: https://doi.org/10.1093/rheumatology/kex043 (Accessed: 11 December 2021)

Sirajudeen, M., Waly, M., Alqahtani, M., Alzhrani, M., Aldhafiri, F., Muthusamy, H., Unnikrishnan, R., Saibannavar, R., Alrubaia, W., Nambi, G., (2020) 'Generalized joint hypermobility among school-aged children in Majmaah region', 11;8, pp1-13. Available at: https://pubmed.ncbi.nlm.nih.gov/32864212/ (Accessed: 11 December 2021)

Skwiot, M., Śliwiński, G., Milanese, S., Śliwiński, Z., (2019) 'Hypermobility of joints in dancers', *Plos One*, 14(2), Pp 1-13. Available at: https://pubmed.ncbi.nlm.nih.gov/30794600/ (Accessed: 11 December 2021)

Smits-Engelsman, B., Klerks., M., Kirby, A., (2011) 'Beighton score: a valid measure for generalized hypermobility in children', *Journal of Pediatrics,* 158(1), pp. 119-23. Available at: https://pubmed.ncbi.nlm.nih.gov/20850761/ (Accessed: 11 December 2021)

Sobhani-Eragi, A., Motalebi, M., Sarreshtedari, S., Molazem-Senandaji, B., Hasanlu, Z., (2021) 'Prevalence of joint hypermobility in children and adolescents: A systematic review and meta-analysis', *Journal of Research in Medical Sciences*, 25(104), pp.1-8 . Available at: https://www.jmsjournal.net/article.asp?issn=1735-1995;year=2020;volume=25;issue=1;spage=104;epage=104;aulast=Sobhani%2DEraghi (Accessed: 21 December 2021)

Soucie, J.M., Wang, C., Forsyth, A., Funk, S., Denny, M., Roach, K.E., Boone, D., (2020) 'Range of motion measurements: reference values and a database for comparison studies', *Official Journal of the World Federation of Hemophilia*, 17, pp. 500–507. Available at: https://doi.org/10.1111/j.1365-2516.2010.02399.x (Accessed: 11 December 2021)

Steinberg, N., Hershkovitz, I., Zeev, A., Rothschild, B., Siev-Ner, I., (2016) 'Joint Hypermobility and Joint Range of Motion in Young Dancers', *Journal of Clinical Rheumatology*, 22(4), pp. 171-8. Available at: https://pubmed.ncbi.nlm.nih.gov/27219302/ (Accessed: 12 December 2021)

Taber, K.S., (2018) 'The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education', *Research in Scientific Education*, 48, pp. 1273–1296. Available at: https://doi.org/10.1007/s11165-016-9602-2 (Accessed: 16 December 2021)

Thomas, N., (2016), *Kung fu master keen to bend the trend on suogugong*, Reuters. Available at: https://www.reuters.com/article/us-china-kungfu-idUSKCN11M1IQ (Accessed: 19 December 2021)

Tinkle, B.T., (2020) 'Symptomatic joint hypermobility', *Best Practice Research in Clinical Rheumatology*, 34, pp. 1021-1033. Available at: https://doi.org/10.1016/j.berh.2020.101508 (Accessed: 11 Decmeber 2021)

Tofts, L.J., Elliott, E.J., Munns, C., Pacey, V., Sillence, D., (2009) 'The differential diagnosis of children with joint hypermobility: a review of the literature', *Pediatric Rheumatology,* 7, pp. 2-10. Available at: https://doi.org/10.1186/1546-0096-7-1 (Accessed: 11 December 2021)

Vallis, A., Wray, A., Smith, T., (2015) 'Inter- and intra-rater reliability of the Beighton Score compared to the Contompasis Score to assess generalised joint hypermobility', *A Journal of Myofascial Pain and Fibromyalgia*, 23(1-2), pp. 21-27. Available at: https://www.magonlinelibrary.com/doi/abs/10.12968/ijtr.2014.21.Sup7.S8a (Accessed: 12 December 2021)

Van der Giessen L., Liekens, D., Rutgers, K., Hartman, A., Mulder, P., Oranje, A., (2001) 'Validation of beighton score and prevalence of connective tissue signs in 773 Dutch children', *Journal of Rheumatology*, 28(12), pp. 2726-2730. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6744937/ (Accessed: 12 December 2021)

Vetter, T. R., Schober, P., Mascha, E. J., (2018) 'Diagnostic Testing and Decision-Making: Beauty Is Not Just in the Eye of the Beholder', *Anesthesia and Analgesia*, 127(4), pp. 1085–1091. Available at: https://doi.org/10.1213/ANE.0000000000003698 (Accessed: 23 December 2021)

Voermans N., Bonnemann C., Hamel B., Jungbluth H., Van Engelen B., (2009) 'Joint Hypermobility as a Distinctive Feature in the Differential Diagnosis of Myopathies', *Journal of Neurology*, 256(1), pp. 13-27. Available at: doi: 10.1007/s00415-009-0105-1. (Accessed: 11 December 2021)

Werner, K., (1977) 'Yoga and the Ṛg Veda: An Interpretation of the Keśin Hymn (RV 10, 136)', *Religious Studies*, 13(3), pp. 289-302. Available at: https://www.academia.edu/19561860/Yoga_and_the_Rg_Veda_An_Interpretation_of_the_Ke%C5%9Bin_Hymn_RV_10_136_?auto=download (Accessed: 19 December 2021)

Worster A., Carpenter, C., (2014) 'Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias', *Canadian Journal of Emergency Medicine*, 10 (2), pp. 174 -175. Available at: https://www.researchgate.net/publication/5481885_Incorporation_bias_in_studies_of_diagnostic_tests_How_to_avoid_being_biased_about_bias (Accessed: 21 December 2021)

Zhang, M., McDaid, A., Zhang, S., Zhang, Y., Xie, S.Q., (2019) 'Automated robot-assisted assessment for wrist active ranges of motion', *Medical Engineering and Physics*, 71, pp. 98–101. Available at: https://doi.org/10.1016/j.medengphy.2019.07.003 (Accessed: 15 December 2021)