The reliability of physical examination tests for the clinical assessment of scapular dyskinesis in subjects with shoulder complaints : a systematic review

Reference:
Lange Toni, Struyf Filip, Schmitt Jochen, Lützner Jörg, Kopkow Christian.- The reliability of physical examination tests for the clinical assessment of scapular dyskinesis in subjects with shoulder complaints : a systematic review
Physical therapy in sport - ISSN 1466-853X - (2016), p. 1-26
Full text (Publishers DOI): http://dx.doi.org/doi:10.1016/J.PTSP.2016.10.006
To cite this reference: http://hdl.handle.net/10067/1409730151162165141

The reliability of physical examination tests for the clinical assessment of scapular dyskinesis in subjects with shoulder complaints: a systematic review

Toni Lange, BSc PT, MPH[1], Filip Struyf, PT, MSc, PhD[2], Jochen Schmitt, MD, MPH[1], Jörg Lützner, MD[3], Dr. Christian Kopkow, BSc PT, MPH[1]

[1] University Hospital Carl Gustav Carus Dresden, Center for Evidence-Based Healthcare, Fetscherstr. 74, 01307 Dresden, Germany
[2] Department of Rehabilitation Sciences and Physiotherapy, Faculty of Medicine and Health Sciences, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium
[3] University Hospital Carl Gustav Carus Dresden, Department of Orthopaedic and Trauma Surgery, Fetscherstr. 74, 01307 Dresden, Germany

Word count: 6131

Corresponding author
Dr. Christian Kopkow, BSc PT, MPH
Center for Evidence-Based Healthcare, University Hospital Carl Gustav Carus Dresden
Fetscherstr. 74, 01307 Dresden
Germany

Tel. +49 (0)351 4586492

Fax +49 (0)351 4587238

E−Mail: Christian.Kopkow@uniklinikum-dresden.de

Highlights

FINDINGS: This systematic review identified a lack of high-quality studies evaluating intra- as well as interrater reliability of physical examination tests for the clinical assessment of scapular dyskinesis.

IMPLICATIONS: There is a strong need for sound research of high methodological quality to establish the reliability especially of symptom alteration tests for scapula dyskinesis, which are used to develop rehabilitation program, to evaluate the effectiveness of the therapy, and to monitor the patient's status. There is a strong need for further research in this area and future studies evaluating the reliability of scapular dyskinesis tests should be in line with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).

CAUTION: Many of the studies included in this review were prone to bias, and their design is likely to overestimate test characteristics.

Abstract

Study Design: Systematic review

Objectives: The aim of this systematic review was to summarize and evaluate intra- and interrater reliability research of physical examination tests used for the assessment of scapular dyskinesis.

Background: Scapular dyskinesis, defined as alteration of normal scapular kinematics, is described as a non-specific response to different shoulder pathologies.

Methods: A systematic literature search was conducted in MEDLINE, EMBASE, AMED and PEDro until March 20th 2015. Methodological quality was assessed with the Quality Appraisal of Reliability Studies (QAREL) by two independent reviewers

Results: The search strategy revealed 3259 articles, of which 15 met the inclusion criteria. These studies evaluated the reliability of 41 test and test variations used for the assessment of scapular dyskinesis.

Conclusion: This review identified a lack of high-quality studies evaluating intra- as well as interrater reliability of tests used for the assessment of scapular dyskinesis. In addition, reliability measures differed between included studies hindering proper cross-study comparisons. The effect of manual correction of the scapula on shoulder symptoms was evaluated in only one study, which is striking, since symptom alteration tests are used in routine care to guide further treatment. Thus, there is a strong need for further research in this area.

Level of Evidence: Diagnosis, level 3a.

Keywords: Literature search, Diagnosis, QAREL, Symptom alteration tests

INTRODUCTION

An alteration of normal scapular kinematics is defined as scapular dyskinesis (Kibler, Ludewig, McClure, Uhl, & Sciascia, 2009; Kibler et al., 2013; Kibler & Sciascia, 2010), and has been identified in patients with shoulder impingement syndrome (Cools, Witvrouw, Declercq, Danneels, & Cambier, 2003; Ludewig & Cook, 2000; P. W. McClure, Michener, & Karduna, 2006; Struyf, Cagnie, et al., 2014; Timmons et al., 2012), rotator cuff pathology (Ludewig & Reynolds, 2009; Mell et al., 2005), shoulder instability (Ludewig & Reynolds, 2009; Ogston & Ludewig, 2007; Struyf, Cagnie, et al., 2014) labral injuries (Myers, Laudner, Pasquale, Bradley, & Lephart, 2006; Roche, Funk, Sciascia, & Kibler, 2015) and adhesive capsulitis (Fayad et al., 2008; Ludewig & Reynolds, 2009). Scapular dyskinesis is assumed to be an unspecific response to specific shoulder pathology (Kibler et al., 2013; Kibler & Sciascia, 2010). In a cohort of Norwegian elite handball players, Clarsen, Bahr, Andersson, Munk, and Myklebust (2014) identified obvious scapular dyskinesis as a risk factor for shoulder injury. Also in rugby players scapular dyskinesis was associated with shoulder discomfort (ie, pain, apprehension, or fatigue; if any, persisting more than 2 weeks) (Kawasaki, Yamakawa, Kaketa, Kobayashi, & Kaneko, 2012). In contrast, in a prospective study of recreational overhead athletes, Struyf, Nijs, Meeus, et al. (2014) did not identify scapular characteristics predicting shoulder pain. In addition, two other studies did not find an association between scapular dyskinesis and shoulder complaints in baseball players (Myers, Oyama, & Hibberd, 2013; Shitara et al., 2015). Because of conflicting evidence regarding the causal relationship of scapular dyskinesis and shoulder pathologies, it seems comprehensible that this issue is still under debate (Cools et al., 2014). It is suggested that during active elevation of the upper limb the scapula needs to upwardly rotate, posteriorly tilt and externally rotate to avoid pathology (Bourne, Choo, Regan, MacIntyre, & Oxland, 2007; Braman, Engel, LaPrade, & Ludewig, 2009; Braman, Zhao, Lawrence, Harrison, & Ludewig, 2014; Kibler et al., 2013; Kibler & Sciascia, 2010; Lawrence, Braman, Laprade, & Ludewig, 2014; Ludewig & Cook, 2000; Ludewig et al., 2009; Matsuki et al., 2011; P. W. McClure, Michener, Sennet, & Karduna, 2001; Yano et al., 2010). Consistent patterns of alterations of scapular movement have not yet been identified for specific pathologies (Kibler et al., 2013; Lawrence et al., 2014; Ludewig & Reynolds, 2009; Ratcliffe, Pickering, McLean, & Lewis, 2014), and altered positions or motions of the scapula might be part of normal variations or adaptations to sporting demands (Cools et al., 2010; Myers, Laudner, Pasquale, Bradley, & Lephart, 2005; Oyama, Myers, Wassinger, Daniel Ricci, & Lephart, 2008; Ribeiro & Pascoal, 2013; Wright, Wassinger, Frank, Michener, & Hegedus, 2013).

Physical examination tests for the assessment of scapular dyskinesis are being used to identify existing impairments of patients with shoulder pain and based on identified impairments during the examination, adequate rehabilitation programs addressing possible cause for scapular dyskinesis can be developed (Cools et al., 2014; Klintberg et al., 2015; Wright et al., 2013). There is evidence that rehabilitation, including exercises targeting the scapulothoracic muscles, is effective in patients with shoulder impingement syndrome (Baskurt, Baskurt, Gelecek, & Ozkan, 2011; De Mey, Danneels, Cagnie, & Cools, 2012; Struyf et al., 2013) and chronic type III acromioclavicular dislocation (Carbone, Postacchini, & Gumina, 2015). Scapular dyskinesis has been primarily evaluated by clinical observation (Kibler & Sciascia, 2013). Therefore, it is essential to use reliable and valid clinical testing procedures to assess the position of the scapula during rest as well as active movement and to evaluate if scapular position or dynamics contribute to shoulder pain. Wright et al. (2013) conducted a systematic review to assess the validity of scapular physical examination tests and questioned the use of specific tests to diagnose shoulder pain or pathology.

Not only validity, but also reliability are of concern regarding the application of clinical tests (Drummond, 2008; Karanicolas et al., 2009; Scholtes, Terwee, & Poolman, 2011), since a test will not be valid if not reliable (Atkinson & Nevill, 1998; H. C. T. de Vet, C. B.Mokkink, L. B.Knol, D. L., 2011). Reliability is generally defined as “the degree to which a measurement is free from measurement error” (Mokkink et al., 2010b). Measurements by the same persons on different occasions are referred to as intrarater reliability, whereas measurements by different persons on the same occasion as interrater reliability (Mokkink et al., 2010b). Intra- as well as interrater reliability both provide valuable information for clinicians. Intrarater reliability is relevant for repeated measurements by the same rater (H. C. de Vet, Mokkink, Terwee, Hoekstra, & Knol, 2013) and therefore a prerequisite for documentation of changes in the patient’s status (e.g. position of the scapula or pain before and after treatment) over time. In contrast, if clinicians are interested in knowing if their diagnosis made is in agreement with the diagnosis made by colleagues, this would be reflected by interrater reliability (H. C. de Vet et al., 2013). Variability among raters influences diagnostic accuracy, as empirical research has shown (P. Whiting et al., 2004; P. F. Whiting, Rutjes, Westwood, Mallett, & Group, 2013).

Scapular physical examination tests are likely to be performed in routine care very frequently to identify patients in need of specific rehabilitation programs, to evaluate the effectiveness of the therapy and to monitor the patient’s status. Thus not only the validity, but also the reliability of physical examination tests for scapular dyskinesis should be summarized systematically and assessed critically.

Therefore, the objective of this systematic review was to summarize and evaluate intra- and interrater reliability research on physical examination tests used for the assessment of scapular dyskinesis in symptomatic subjects.

METHODS

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used during the whole process of this review (Liberati et al., 2009). The PRISMA statement aims to improve the reporting of systematic reviews and meta-analyses. This systematic review was registered a priori within the International prospective register of systematic reviews (PROSPERO; CRD42014009018).

Inclusion/exclusion criteria

Studies assessing the intra- and/or interrater reliability of a physical examination test for the assessment of scapular dyskinesis applied as a single test or in combination with other tests were included if written in English or German. Studies on patients of every age and setting were considered eligible. Studies evaluating physical examination tests were excluded from this systematic review if they did not named or described the evaluated physical tests or did not reference a source that did so. Studies were excluded if the overall reliability of a group of tests was reported, but individual tests were not specified/named and therefore individual tests could not clearly be identified. Studies were also excluded if the authors made use of generic terms, such as physical examination, to denote an unspecified combination of physical tests. In addition, studies were excluded if only asymptomatic subjects were evaluated or if the physical examination test was performed under anesthesia or immediately postoperative. Animal studies and cadaveric studies and studies which used device supported testing procedures (defined as devices which are deemed too expensive or time-consuming for daily clinical practice) were also excluded.

Search strategy

A comprehensive systematic literature search in the following databases via the Ovid interface from inception until March 18th, 2014 was performed, accessed via the Saxon State and University Library Dresden (SLUB): MEDLINE from 1946, EMBASE from 1974, and the Allied and

Complementary Medicine Database (AMED) from 1985. A search update using the same search strategy and electronic databases was conducted on March 20th 2015 to identify recently published articles. The search strategy included terms about diagnostic tests, the conditions of interest, structures at risk, and reliability (Appendix). Additionally, PEDro was searched with a modified search strategy using the body part filter (upper arm, shoulder or shoulder girdle) in combination with the terms for reliability as used for the search in MEDLINE, EMBASE and AMED. Furthermore, reference lists of all eligible articles were screened for further relevant studies. The original search strategy was designed to identify studies on the reliability of specific physical examination tests evaluating specific structures (e.g. rotator cuff tear) and general physical examinations tests (e.g. strength or range of motion testing for the shoulder as well as shoulder girdle). However, in this review only the results for clinical tests for the assessment of scapular dyskinesis are reported.

Study selection and data abstraction

Identified titles and abstracts were screened by two independent reviewers (T.L. and C.K.), according to the described inclusion criteria. Subsequently, full texts were checked independently for eligibility by the same two reviewers. Any disagreements were resolved in discussion between the two reviewers, and if necessary with the help from a third reviewer (J.S.). Before titles and abstracts screening initiation, two subsamples consisting of randomly selected 50 titles and abstracts from all identified articles were performed. Afterwards the two reviewers (T.L. and C.K.) discussed their procedure to avoid following inequalities and started with the titles and abstract screening after an almost perfect agreement (according to classification system proposed by Landis and Koch (1977)) was reached in the second pretest subsample (subsample 1: Cohen´s Kappa [k] = 0.22, percentage agreement = 88.00 %; subsample 2: Cohen´s k = 1.00, percentage agreement = 100.00 %) Data extraction was done by one reviewer (T.L.) and all abstracted data were subsequently checked by the second reviewer (C.K.). For standardized data extraction, Mircosoft® Access forms were created and used according to the data extraction form of the Quality Appraisal of Reliability Studies (QAREL) checklist (N. P. Lucas, Macaskill, Irwig, & Bogduk, 2010). Data about the objectives, subjects, raters, physical examination tests, outcome variables and results were extracted. Authors of primary studies were contacted if inappropriate reliability measures were used or if reliability measures were incompletely reported. In case the authors provided the requested information, the appropriate reliability measures were calculated if possible.

Quality assessment

Quality assessment of included studies was done independently by two reviewers (T.L. and C.K.) using the QAREL checklist (N. P. Lucas et al., 2010). QAREL is especially designed for the quality assessment of reliability studies and is considered to be reliable for use (N. Lucas et al., 2013). The checklist consists of 11 items evaluating 7 methodological domains of reliability studies (spectrum of subjects and of examiners, examiner blinding, time interval between repeated measures, test application and interpretation, order of examination and statistical analysis of the data). Items can be answered with "yes", "no" or "unclear" (and in addition if necessary with "not applicable"). Fulfilled quality aspects of studies are indicated with a "yes", whereas not fulfilled aspects with a "no". If insufficient information are provided to properly judge the quality aspect of studies, this is indicated with an "unclear". As recommended (N. Lucas et al., 2013), criteria by which judgments were made for each item of QAREL were a priori defined and tested. If in one study both intra- and interrater reliability were evaluated, two separate assessments of methodological quality using QAREL were performed to account for specific risk of bias of reliability study types.

Data synthesis

Reliability measures are presented as reported by the authors of primary studies. Kappa values less than 0.00 indicate poor, 0.00 - 0.20 slight, 0.21 - 0.40 fair, 0.41 - 0.60 moderate, 0.61 - 0.80 substantial and greater than 0.81 almost perfect agreement (Landis & Koch, 1977). Intraclass correlation coefficient (ICC) values less than 0.40 represent poor, values between 0.40 and 0.75 represent fair to good, and values above 0.75 represent excellent reliability (Fleiss, 1999).

The agreement among reviewers of title and abstract screening as well as assessment of methodological quality using the QAREL tool was measured with percentage agreement and Cohen`s Kappa statistic (95 % confidence interval [CI]) (Cohen, 1960). In addition, prevalence-adjusted bias-adjusted kappa (PABAK), positive and negative percentage agreement as well as bias- (BI) and prevalence index (PI) were calculated to evaluate agreement of the title and abstract screening process (Byrt, Bishop, & Carlin, 1993; H. C. de Vet et al., 2013; Looney & Hagan, 2007). PI helps to determine prevalence effects with values ranging from -1 to +1 and "is equal to zero when 'yes' and 'no' are equally probable" (Byrt et al., 1993). BI indicates bias with absolute values ranging from 0 to 1 and is zero if the marginal proportions, representing the assessment of the frequency of occurrence of a condition, are equal (Byrt et al., 1993). PABAK adjusts the K for imbalances due to differences in the prevalence and bias (Byrt et al., 1993; Looney & Hagan, 2007).

All statistical analyses were performed using R Version 3.2.0 (The R Project for Statistical Computing, Vienna, Austria) and RStudio (RStudio, Inc, Boston, MA).

RESULTS

Study selection

Results of the literature search and study selection are shown in the PRISMA flow chart (**FIGURE 1**). In total, 3259 records were screened after duplicates were removed, of which 15 publications met the predefined inclusion/exclusion criteria (Baertschi, Swanenburg, Brunner, & Kool, 2013; Curtis & Roush, 2006; Juul-Kristensen, Hilt, Enoch, Remvig, & Sjogaard, 2011; Kibler et al., 2002; J. S. Lewis & Valentine, 2007, 2008; P. McClure, Tate, Kareha, Irwin, & Zlupko, 2009; Nijs, Roussel, Vermeulen, & Souvereyns, 2005; Odom, Taylor, Hurd, & Denegar, 2001; Rabin, Irrgang, Fitzgerald, & Eubanks, 2006; Shadmehr, Bagheri, Ansari, & Sarafraz, 2010; Struyf, Meeus, et al., 2014; Terwee et al., 2005; Uhl, Kibler, Gecewich, & Tripp, 2009; Watson, Balster, Finch, & Dalziel, 2005). Agreement among reviewers regarding screening of titles and abstracts yielded a Cohen's k of 0.76 (CI: 0.70, 0.82) and of full-text articles 0.68 (CI: 0.54, 0.81); additional reliability statistics are presented in the **FIGURE 1**. After finalization of the full text screening, fifteen publications were included in this review, in which 41 different physical examination tests and test modifications for scapular dyskinesis were investigated. Characteristics of included studies are summarized in **TABLE 1**.

All included studies were prospective and intrarater reliability in four (Juul-Kristensen et al., 2011; J. S. Lewis & Valentine, 2007, 2008; Watson et al., 2005), interrater reliability was assessed in seven (Baertschi et al., 2013; Curtis & Roush, 2006; P. McClure et al., 2009; Nijs et al., 2005; Rabin et al., 2006; Terwee et al., 2005; Uhl et al., 2009), and both types of reliability in four studies (Kibler et al., 2002; Odom et al., 2001; Shadmehr et al., 2010; Struyf, Meeus, et al., 2014). In twelve out of fifteen publications studies were conducted in primary (Baertschi et al., 2013; Curtis & Roush, 2006; J. S. Lewis & Valentine, 2008; P. McClure et al., 2009; Nijs et al., 2005; Odom et al., 2001; Rabin et al., 2006; Shadmehr et al., 2010;

Struyf, Meeus, et al., 2014; Terwee et al., 2005; Uhl et al., 2009; Watson et al., 2005), one in secondary (J. S. Lewis & Valentine, 2007) and two in undefined care settings (Juul-Kristensen et al., 2011; Kibler et al., 2002). Primary care was defined as generally in the community setting performed clinical care (low prevalence or pretest probability of target disorder), and secondary care as referral following preliminary screening or referral to a specialist center (higher prevalence or pretest probability of target disorder) (Christian Kopkow et al., 2015; Sackett & Haynes, 2002).

Methodological quality

Results of methodological assessment using QAREL for included studies are summarized in **TABLE 2**. Methodological quality ranged from 4/11 rating (J. S. Lewis & Valentine, 2007; P. McClure et al., 2009; Watson et al., 2005) to 7/11 total positive ratings (Baertschi et al., 2013; Curtis & Roush, 2006; Kibler et al., 2002; Odom et al., 2001; Struyf, Meeus, et al., 2014) and from 2/11 ratings (Curtis & Roush, 2006; J. S. Lewis & Valentine, 2007) to 5/11 total unclear ratings (Kibler et al., 2002; J. S. Lewis & Valentine, 2008; P. McClure et al., 2009; Rabin et al., 2006; Uhl et al., 2009; Watson et al., 2005). Recruitment of raters was not specified in any of the included studies, which limits identification of possible bias on reliability values. In two included studies, subjects were recruited consecutively (Juul-Kristensen et al., 2011; Terwee et al., 2005), in eight studies through convenience sampling (Curtis & Roush, 2006; J. S. Lewis & Valentine, 2008; P. McClure et al., 2009; Nijs et al., 2005; Odom et al., 2001; Struyf, Meeus, et al., 2014; Uhl et al., 2009; Watson et al., 2005). In three studies subjects were referred (J. S. Lewis & Valentine, 2007; Rabin et al., 2006; Shadmehr et al., 2010) and in two studies the recruitment protocol was unclear (Baertschi et al., 2013; Kibler et al., 2002). In five out of eight intrarater reliability studies, the blinding of raters to their own prior findings was judged as unclear due to insufficient information (Juul-Kristensen et al., 2011; Kibler et al., 2002; J. S. Lewis & Valentine, 2008; Shadmehr et al., 2010; Watson et al., 2005). Blinding of raters to the findings of other raters was unclear in two out of eleven interrater reliability studies (Baertschi et al., 2013; Shadmehr et al., 2010). Blinding to further clinical information was stated in three out of fifteen included studies (Baertschi et al., 2013; Curtis & Roush, 2006; Struyf, Meeus, et al., 2014). None of the included studies described blinding to additional cues (e.g. scars, tattoos) of investigated subjects.

Percentage agreement among reviewers regarding the rating of methodological quality of included studies using QAREL ranged for the different QAREL items from 60 up to 100 %. The overall agreement between raters of the methodological assessment using QAREL yielded a Cohen's k of 0.83 (CI: 0.77, 0.89).

Physical examination tests

Physical examination tests for the clinical assessment of scapular dyskinesis were categorized as follows: observation of scapular positioning / motion (N= 10), measurement of scapular positioning (N= 29) and symptom alterations tests (N= 2). Altogether forty-one different physical examination tests were evaluated in the studies included in this systematic review.

Observation of scapular positioning / motion

Reliability of visual observation of static scapular positioning was evaluated less frequently than visual observation of dynamic scapular positioning. Intrarater reliability of visual observation of static scapular positioning was evaluated in one study (Juul-Kristensen et al., 2011) and no studies were identified evaluating interrater reliability. Intrarater reliability of dynamic scapular observation was evaluated in two studies (Juul-Kristensen et al., 2011; Kibler et al., 2002) and interrater reliability in four studies (Baertschi et al., 2013; Kibler et al., 2002; P. McClure et al., 2009; Uhl et al., 2009). It should be noted that all studies evaluating interrater reliability of dynamic scapular observation used different assessment methods. Intrarater reliability of visual observation of scapular positioning / motion ranged from k= 0.59 to k= 1.00 (**TABLE 3**), and interrater reliability from k= 0.31 to k= 0.84 (**TABLE 4**).

Measurement of scapular positioning

Intrarater reliability of clinical assessments for measurement of scapular positioning during resting position ranged from ICC= 0.52 to ICC= 0.98 and was evaluated more frequently than interrater reliability, which ranged from ICC= 0.47 to ICC= 0.97 (**TABLE 3**). Intrarater reliability of clinical assessments for measurement of scapular positioning during non-resting position ranged from ICC= 0.42 to ICC= 0.97, whereas interrater reliability from ICC= 0.15 to ICC= 0.96 (**TABLE 4**).

Symptom alterations test

One included study evaluated the interrater reliability of the modified scapular assistance test (mSAT), showing k= 0.62 for test performance in the sagittal plane and k= 0.53 for test performance in the scapular plane (**TABLE 3**) (Rabin et al., 2006).

DISCUSSION

Main findings

This systematic review identified fifteen articles, which examined the reliability of forty-one physical examination tests for the clinical assessment of scapular dyskinesis with varying intrarater as well as interrater reliability. The included studies were partly of low methodological quality according to the methodological assessment using the QAREL tool (N. P. Lucas et al., 2010). In addition, it is notable that for only five out of forty-one (12.20 %) physical examination tests evaluated in the studies included in this review both intra- as well as interrater reliability measures were available. Because of methodological quality flaws, different and partially inadequate reliability measures of included studies, meta-analysis of results was not possible.

McQuade, Borstad, and de Oliveira (2016) stated that the `connection between non-normal movement and stability has not been established and many individuals with scapular dyskinesis maintain healthy functional use of the extremity` (McQuade et al., 2016). Tests to steer patient management have been advocated (J. Lewis, McCreesh, Roy, & Ginn, 2015) and rehabilitation should not solely be based on pathoanatomic diagnosis, but also on tissue irritability and identified impairments (P. W. McClure & Michener, 2015). Obviously, scapular dyskinesis cannot be seen as a pathoanatomic diagnosis, and observed abnormal motion may represent normal variability, questioniong the clinical use of observation of scapular positioning/motion and measurement of scapular positioning (McQuade et al., 2016). However, some studies (Clarsen et al., 2014;

Kawasaki et al., 2012) have demonstrated that scapular dyskinesis might be predictive for the development of shoulder discomfort, the causal relationship of scapular dyskinesis and shoulder pathologies is still under debate (Cools et al., 2014).

Observation of scapular positioning / motion, measurement of scapular positioning and furthermore symptom alteration tests have been recommended in different publications including consensus statements (Kibler et al., 2013; Kibler & Sciascia, 2010; Klintberg et al., 2015; P. McClure, Greenberg, & Kareha, 2012; Pluim, 2013; Seitz, McClure, Lynch, Ketchum, & Michener, 2012; Struyf, Nijs, Mottram, et al., 2014; Tate, McClure, Kareha, & Irwin, 2008), representing face validity and acceptance in clinical practice of those tests. However, the evaluation of the validity of the aforementioned tests is limited by the `nebulous concept´ (Wright et al., 2013) and subjective and qualitative character of scapular dyskinesis (McQuade et al., 2016) which hinders research efforts within this field.

Observation of scapular positioning / motion

Based on the results of this systematic review, k as a reliability measure for the intrarater reliability for observation of scapular positioning and motion is ranges from 0.59 to 1.00 (**TABLE 3**), and the intrarater reliability from 0.31 to 0.84 (**TABLE 4**). In a recently published study, physical therapists were not able to reliably and validly determine if abnormal or asymmetrical scapular movements are related to injury (Wassinger, Williams, Milosavljevic, & Hegedus, 2015). Additionally, Hickey, Milosavljevic, Bell, and Milburn (2007) concluded that movement analysis alone is not sufficient for determining shoulder complaints. Thus, visual scapular evaluation might be of limited clinical value since the technique is not able to diagnose the presence or absence of shoulder pain due to malposition of the scapula (Hickey et al., 2007; Wright et al., 2013). Because of the aforementioned reasons, the validity of visual scapular evaluation remains unclear. The clinical usefulness of visual observation should be questioned, especially since this systematic review did not identified homogenous evidence for the reliability of visual observation from studies of high methodological quality.

Measurement of scapular positioning

Measurement of scapular positioning, either performed during resting or non-resting position, are performed in a static manner and therefore probably not able to reproduce complex 3-dimensional scapular kinematics during upper limb movement. Thus, the results from such measurements are of limited clinical relevance even though the intrarater and interrater reliability of measurements of scapular positioning during resting and non-resting position seems to be reliable based on the results of this systematic review (**TABLE 3** and **4**). Scapular asymmetry may be normal in overhead athletes (Cools, Johansson, Borms, & Maenhout, 2015; Hosseinimehr, Anbarian, Norasteh, Fardmal, & Khosravi, 2015; Kibler et al., 2013), furthermore questioning the usefulness of scapular positioning measurement. In another systematic review from Larsen, Juul-Kristensen, Lund, and Sogaard (2014), the authors concluded that visual observation and inclinometer measurement of the scapular position and function is both reliable and valid, but results do not provide information on shoulder pain due to scapular alterations. Wright et al. (2013) concluded in their systematic review that scapular physical examination tests lack diagnostic accuracy, and discussed lacking reliability, validity or both and dyskinesis as being a poor reference standard as possible explanations for their findings. However, if altered scapular position or motion impairments are observed, further tests (e.g. for scapular muscle strength, length of soft tissue) might help to select specific treatment intervention (P. McClure et al., 2012; Wright et al., 2013).

Symptom alterations test

The scapular assistance test (SAT), (Kibler, 1998; Kibler & Sciascia, 2010) (or the modified version of this test; the mSAT (Rabin et al., 2006)) a test where the examiner assists the scapula during elevation of the upper limb, is a symptom alteration test recommended for clinical practice (Kibler & Sciascia, 2010; Pluim, 2013; Struyf, Nijs, Mottram, et al., 2014). The mSAT is assumed to identify if scapular dyskinesis is a contributing factor to shoulder pain, and is therefore helpful to identify patients in need of specific treatment programs. However, only one study included in this systematic review evaluated the interrater reliability of the mSAT reporting k= 0.62 (for test performance in the sagittal plane) and k= 0.53 (for test performance in the scapular plane) (**TABLE 3**). A more recently published study published after finalization of the database searching process reported k= 0.68 for test performance in the scapular plane (C. Kopkow, Lange, Schmitt, & Kasten, 2015). However, both studies did not evaluate intrarater reliability (C. Kopkow et al., 2015; Rabin et al., 2006). Studies evaluating another symptom alteration test, the scapular retraction/repositioning test (SRT) (Struyf, Nijs, Mottram, et al., 2014), were not identified. This is notable, since both tests are mentioned explicitly during a recently published assessment algorithm for patients with shoulder pain (Klintberg et al., 2015).

Therefore, the mSAT should be considered reliable and helpful to identify whether the patient will benefit from a scapulothoracic approach. However, prospective studies are needed to provide empirical evidence if following treatment will improve patient-relevant outcomes, ideally measured using patient-reported outcomes and not only kinematic variables without having an effect on e.g. pain or function.

Methodological considerations and generalizability of results

The overall generalizability of this review results is limited due to the low quality of included studies (**TABLE 2**). Reliability measures reported in included studies might be inflated due to insufficient study methodology and statistical analysis. Altogether 35.45 % of the QAREL items were judged as “unclear” during critical appraisal of methodological quality of included studies, representing insufficient reporting of methodological aspects within primary studies. Furthermore, generalizability is limited due to differences in performance as well as interpretation of physical examination tests.

Rater experience and training status can have a major impact on reliability results (Carlsson & Rasmussen-Barr, 2013; Karanicolas et al., 2009; Kottner et al., 2011; Rousson, Gasser, & Seifert, 2002), but was not reported in six out of fifteen studies included in this review (Baertschi et al., 2013; Kibler et al., 2002; J. S. Lewis & Valentine, 2007; Nijs et al., 2005; Shadmehr et al., 2010; Uhl et al., 2009). Blinding of raters to the reference standard, clinical information, and additional cues in most included studies was not reported sufficiently, which illustrates a considerable quality flaw.

Due to the lack of blinding, reliability will most likely be overestimated. E.g. if raters are not blinded to additional cues, raters might recognize former tested patients and/or remember previous ratings in intrarater reliability studies. If raters are aware of results from the reference standard, raters might tend to rate the test under evaluation in accordance with the reference standard results (relevant for intra- as well interrater reliability studies). However, for scapular dyskinesis a generally accepted reference standard seems to be lacking, but the knowledge of shoulder pathologies for which a reference standard is available (e.g. shoulder impingement syndrome, rotator cuff pathology, shoulder instability, labral injuries or adhesive capsulitits) might influence rater judgements due to anticipated coexistence of pathologies and scapular dyskinesis (Cools et al., 2003; Fayad et al., 2008; Ludewig & Cook, 2000; Ludewig & Reynolds, 2009; P. W. McClure et al., 2006; Mell et al., 2005; Myers et al., 2006; Ogston & Ludewig, 2007; Roche et al., 2015; Struyf, Cagnie, et al., 2014; Timmons et al., 2012).

Prospective studies using consecutive or randomly sampled patients were considered to be of highest methodological quality (Fritz & Wainner, 2001; Lange et al., 2015). However, only two included studies described consecutive subject recruitment (Juul-Kristensen et al., 2011; Terwee et al., 2005).

For the reporting of reliability study results, the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were published (Kottner et al., 2011). GRRAS intends to improve the quality of reporting, similar to the STARD initiative for studies of diagnostic accuracy (Bossuyt, Reitsma, Bruns, et al., 2003; Bossuyt, Reitsma, & Standards for Reporting of Diagnostic, 2003). Only two of the studies included in this systematic review (Baertschi et al., 2013; Struyf, Meeus, et al., 2014) were published after publication of GRRAS, but only the study by Struyf, Meeus, et al. (2014) used GRRAS for the reporting of study results. It must be acknowledged that the review process and publication of accepted manuscripts can take months, and it may have been likely that the article which did not used GRRAS was submitted and already in the review process prior to the publication of the GRRAS.

Larsen et al. (2014) conducted a systematic review of measurement properties of existing clinical assessment methods to evaluate scapular position and function, but used the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist (Mokkink et al., 2010a), instead of the QAREL tool which was used in this review (N. P. Lucas et al., 2010), for the assessment of methodological quality of included studies. The COSMIN Checklist was initially designed to evaluate the methodological quality of studies on measurement properties of health-related patient reported outcomes (HR-PROs), the QAREL tool in contrast was especially designed for the assessment of key aspects of the quality and applicability of reliability studies (N. P. Lucas et al., 2010) and has been used in other systematic reviews on the reliability of physical examination tests (Barrett, McCreesh, & Lewis, 2014; Carlsson & Rasmussen-Barr, 2013; Lange et al., 2015) or measurement methods (McCreesh, Crotty, & Lewis, 2015). In addition, in this review studies on asymptomatic subjects were included which limits the generalizability of the review results, since reliability should be evaluated in the population in which the tests are intended to use (Karanicolas et al., 2009; Kottner et al., 2011; N. P. Lucas et al., 2010).

Statistical considerations

A priori sample size calculation is recommended for reliability studies (Donner & Rotondi, 2010; Karanicolas et al., 2009; Sim & Wright, 2005), but none of the included studies performed such a calculation or reported that a priori sample size calculation / post hoc power analysis was performed, potentially limiting the value and generalizability according to statistical considerations. Studies with insufficient sample sizes are likely to produce unprecise estimates of agreement; therefore sample size calculations are needed and must be reported to enable readers to determine if the number of recruited study participants is sufficient or not.

Since prevalence rates of certain disorders or impairments in routine care are presumably not equally distributed, agreement of categorical judged physical examination tests might occur purely by chance (Gilchrist, 2009). Relative reliability measures which take the agreement occurring by chance into account (e.g. Cohens' k) are therefore necessary (Sim & Wright, 2005). However, Cohen's k as a relative reliability measure which is often used, has been criticized by several authors, since Cohen’s k is also affected by prevalence of test result categories (Byrt et al., 1993; Cicchetti & Feinstein, 1990; H. C. de Vet et al., 2013; Feinstein & Cicchetti, 1990; Sim & Wright, 2005). If the prevalence (of the condition in the population under evaluation) differs from 50%, this will maximize the divergence between absolute (proportion of observed agreement) and relative reliability measures (e.g. Cohen's k) (Cicchetti & Feinstein, 1990; Feinstein & Cicchetti, 1990). Resolving this, Byrt et al. (1993) introduced the prevalence-adjusted bias-adjusted kappa (PABAK), prevalence and bias index. However, PABAK as a reliability measure relates to a hypothetical situation without any prevalence as well as bias effects (Sim & Wright, 2005). Interestingly, only one (Baertschi et al., 2013) of the six studies (Baertschi et al., 2013; Juul-Kristensen et al., 2011; Kibler et al., 2002; P. McClure et al., 2009; Rabin et al., 2006; Uhl et al., 2009) which reported reliability measures for categorical data provided alongside Cohen’s k values PABAK, prevalence and bias index values. However, which reliability measures are the most preferable and should be calculated and reported for categorical data, is to date unclear. Therefore authors of primary studies should provide absolute and relative reliability measures as well as summarized original data in the form of 2x2 (contingency) tables to enable recalculation of additional measures like positive and negative agreement for absolute reliability and PABAK, prevalence and bias indexes for relative reliability, respectively. Since all named reliability measures are point estimates, confidence intervals must be reported in order to interpret the values approximate values of reported reliability measures.

Included studies using continuous data (e.g. measuring distance in cm) calculated the recommended ICC (H. C. de Vet, Terwee, Knol, & Bouter, 2006; Scholtes et al., 2011) as a reliability measure, suggesting appropriate statistical analysis. Measures of uncertainty (e.g. standard error of measurement [SEM], minimal detectable change [MDC]) were not reported from four (Curtis & Roush, 2006; Juul-Kristensen et al., 2011; Nijs et al., 2005; Terwee et al., 2005) and confidence intervals from two (Curtis & Roush, 2006; Nijs et al., 2005) out of ten studies using continuous data. However, measures of uncertainty and confidence intervals are needed to enable clinicians to interpret a given measure of continuous data.

It should be acknowledged that generally accepted classification systems for reliability measures are currently lacking, but the classification systems by Landis and Koch (1977) for categorical data and Fleiss (1999) for continuous data are widely used. Therefore, within this review the aforementioned classification systems were used to categorize the strength of agreement for individual physical examination tests. In addition, minimal requirements regarding clinical acceptable values of reliability measures are currently not available neither for categorical (e.g. Cohen´s k) nor continuous data (e.g. ICC) (Atkinson & Nevill, 1998; Bruton, Conway, & Holgate, 2000; Scholtes et al., 2011), but would be of great help for clinicians to decide which physical examination tests should be considered reliable for clinical use.

Implications for further research

Kibler et al. (2013) recommended that scapular dyskinesis should be primarily evaluated through 1) visual observation, 2) effect of manual correction on shoulder symptoms (symptom alteration tests), and 3) examination of surrounding anatomic structures. Based on the results of this review, most research has been performed within the area of clinical assessments for measurement of scapular positioning (29 out of 41 tests from included studies). However, to identify if scapular dyskinesis contributes to the patients symptoms (e.g. pain and functional impairments), symptom alteration tests are used. Visual observation of scapular positioning / motion fails to identify if existing alterations contribute to the patients shoulder symptoms, thus more research is warranted for symptom alteration tests. In detail, the intrarater reliability of the mSAT needs to be evaluated and in addition the reliability of the Scapular Reposition/Retraction test, for which neither interrater nor intrarater reliability has been evaluated. Future reliability studies evaluating physical examination tests for scapular dyskinesis should calculate and report for dichotomous outcome data the 2x2 table, absolute (e.g. proportion of positive as well as negative agreement) and relative reliability measures (e.g. k, maximum k, prevalence-adjusted bias-adjusted kappa [all with 95% CI]), prevalence and bias index as recommended by several authors (Byrt et al., 1993; Cicchetti & Feinstein, 1990; H. C. de Vet et al., 2013; Feinstein & Cicchetti, 1990; Kottner et al., 2011; Sim & Wright, 2005). For continuous data, ICC values (with 95% CI) and standard error of measurement should be calculated and reported (Weir, 2005). To ensure that readers are able to properly interpret and adopt the reliability measures accordingly into clinical practice as well as research, the aforementioned reliability measures should be calculated and reported. Reliability studies should be registered prospectively in trial registers (e.g. International Clinical Trials Registry Platform (ICTRP) or ClinicalTrials.gov) to ensure transparency. In addition, prospective study designs with consecutive or randomly sampled patient samples based on a priori sample size calculation should be used in reliability studies.

Furthermore, an international consensus is needed regarding minimal standards for reliability studies conduction and the consistent application of GRRAS (Kottner et al., 2011) to improve the report quality.

Limitations

Even though we performed electronic literature searches in three databases (MEDLINE, EMBASE, and AMED) to ensure identification of all relevant articles, our search might have missed relevant studies since we did not searched grey literature and were not able to search within SCOPUS or SPORTDiscus due to license reasons. In addition, due to language restrictions the possibility of language bias might exist. In this review, four studies were excluded according to language restrictions (Patel, Hossain, Colaco, El-Husseiny, & Lee, 2011; Perez-Santonja, Bellot, Claramonte, Ismail, & Alio, 1997; T'Jonck, 2001; Yang et al., 2015). Initially, we planned a meta-analysis, but due to the already discussed reasons for heterogeneity and the small number of included studies this was not possible. For a successful meta-analysis the test(s) of interest must have been assessed frequently and the same reliability measures must have been reported (Sun, 2011), _ENREF_92 which was not the case in this systematic review.

Despite contacting all authors if incomplete reliability statistics were reported in primary studies, data could not be obtain from all contacted authors. Thus a recall bias might exist.

CONCLUSION

Numerous physical examinations tests for the assessment of scapular dyskinesis are described in the literature. Overall, there is a lack of high-quality studies evaluating intra- as well as interrater reliability. In addition, reliability measures differed between included studies which hinder proper comparison of results. The effect of manual correction of the scapula on shoulder symptoms was evaluated in only one study, which is striking, since symptom alteration tests are used in routine care to guide further treatment. Thus, there is a strong need for further research in this area. Future studies evaluating the reliability of scapular dyskinesis tests should be in line with GRRAS (Kottner et al., 2011). Furthermore, an international consensus regarding minimal requirements for the design of reliability studies as well as test conduction of physical examination tests is needed to ensure comparison of studies results.

Contributors

TL made a substantial contribution to the design of the study; performed the literature search; reviewed the literature; methodologically appraised the articles; extracted, analyzed and interpreted the data; produced the figures and graphs; critically revised and wrote the manuscript.

FS assisted with analysis and interpretation of data; critically revised the article and wrote the manuscript.

JS critically commented on the design of the study; critically revised the article.

JL critically commented on the design of the study; critically revised the article.

CK made a substantial contribution to the design of the study; reviewed the literature; methodologically appraised the articles; extracted the data in duplicate; analyzed and interpreted the data; critically revised and wrote the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*, *26*, 217-238.

Baertschi, E., Swanenburg, J., Brunner, F., & Kool, J. (2013). Interrater reliability of clinical tests to evaluate scapulothoracic motion. *BMC Musculoskelet Disord*, *14*, 315.

Barrett, E., McCreesh, K., & Lewis, J. (2014). Reliability and validity of non-radiographic methods of thoracic kyphosis measurement: a systematic review. *Man Ther*, *19*, 10-17.

Baskurt, Z., Baskurt, F., Gelecek, N., & Ozkan, M. H. (2011). The effectiveness of scapular stabilization exercise in the patients with subacromial impingement syndrome. *J Back Musculoskelet Rehabil*, *24*, 173-179.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D., de Vet, H. C., & Standards for Reporting of Diagnostic, A. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ*, *326*, 41-44.

Bossuyt, P. M., Reitsma, J. B., & Standards for Reporting of Diagnostic, A. (2003). The STARD initiative. *Lancet*, *361*, 71.

Bourne, D. A., Choo, A. M. T., Regan, W. D., MacIntyre, D. L., & Oxland, T. R. (2007). Three-dimensional rotation of the scapula during functional movements: An in vivo study in healthy volunteers. *Journal of Shoulder and Elbow Surgery*, *16*, 150-162.

Braman, J. P., Engel, S. C., LaPrade, R. F., & Ludewig, P. M. (2009). In vivo assessment of scapulohumeral rhythm during unconstrained overhead reaching in asymptomatic subjects. *Journal of Shoulder and Elbow Surgery*, *18*, 960-967.

Braman, J. P., Zhao, K. D., Lawrence, R. L., Harrison, A. K., & Ludewig, P. M. (2014). Shoulder impingement revisited: evolution of diagnostic understanding in orthopedic surgery and physical therapy. *Medical & Biological Engineering & Computing*, *52*, 211-219.

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy*, *86*, 94-99.

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *J Clin Epidemiol*, *46*, 423-429.

Carbone, S., Postacchini, R., & Gumina, S. (2015). Scapular dyskinesis and SICK syndrome in patients with a chronic type III acromioclavicular dislocation. Results of rehabilitation. *Knee Surgery, Sports Traumatology, Arthroscopy*, *23*, 1473-1480.

Carlsson, H., & Rasmussen-Barr, E. (2013). Clinical screening tests for assessing movement control in non-specific low-back pain. A systematic review of intra- and inter-observer reliability studies. *Man Ther, 18*, 103-110.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol, 43*, 551-558.

Clarsen, B., Bahr, R., Andersson, S. H., Munk, R., & Myklebust, G. (2014). Reduced glenohumeral rotation, external rotation weakness and scapular dyskinesis are risk factors for shoulder injuries among elite male handball players: a prospective cohort study. *Br J Sports Med, 48*, 1327-1333.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*, 37-46.

Cools, A. M., Johansson, F. R., Borms, D., & Maenhout, A. (2015). Prevention of shoulder injuries in overhead athletes: a science-based approach. *Braz J Phys Ther*, *19*, 331-339.

Cools, A. M., Johansson, F. R., Cambier, D. C., Velde, A. V., Palmans, T., & Witvrouw, E. E. (2010). Descriptive profile of scapulothoracic position, strength and flexibility variables in adolescent elite tennis players. *Br J Sports Med, 44*, 678-684.

Cools, A. M., Struyf, F., De Mey, K., Maenhout, A., Castelein, B., & Cagnie, B. (2014). Rehabilitation of scapular dyskinesis: from the office worker to the elite overhead athlete. *Br J Sports Med, 48*, 692-697.

Cools, A. M., Witvrouw, E. E., Declercq, G. A., Danneels, L. A., & Cambier, D. C. (2003). Scapular muscle recruitment patterns: Trapezius muscle latency with and without impingement symptoms. *American Journal of Sports Medicine, 31*, 542-549.

Curtis, T., & Roush, J. R. (2006). The Lateral Scapular Slide Test: A Reliability Study of Males with and without Shoulder Pathology. *North American Journal of Sports Physical Therapy, 1*, 140-146.

De Mey, K., Danneels, L., Cagnie, B., & Cools, A. M. (2012). Scapular muscle rehabilitation exercises in overhead athletes with impingement symptoms: effect of a 6-week training program on muscle recruitment and functional outcome. *Am J Sports Med, 40*, 1906-1915.

de Vet, H. C., Mokkink, L. B., Terwee, C. B., Hoekstra, O. S., & Knol, D. L. (2013). Clinicians are right not to like Cohen's kappa. *BMJ, 346*, f2125.

de Vet, H. C., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *J Clin Epidemiol, 59*, 1033-1039.

de Vet, H. C. T., C. B.Mokkink, L. B.Knol, D. L. (2011). Measurement in medicine. In H. C. T. de Vet, C. B.Mokkink, L. B.Knol, D. L. (Ed.), *Measurement in medicine / a practical guide Measurement in medicine : a practical guide* (pp. 150 - 201). Cambridge: Cambridge University Press.

Donner, A., & Rotondi, M. A. (2010). Sample size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. *Int J Biostat, 6*, Article 31.

Drummond, R. S., D. L. (2008). The Rational Clinical Examination: Evidence-Based Clinical Diagnosis. In. United States: McGraw-Hill.

Fayad, F., Roby-Brami, A., Yazbeck, C., Hanneton, S., Lefevre-Colau, M. M., Gautheron, V., Poiraudeau, S., & Revel, M. (2008). Three-dimensional scapular kinematics and scapulohumeral rhythm in patients with glenohumeral osteoarthritis or frozen shoulder. *Journal of Biomechanics, 41*, 326-332.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol, 43*, 543-549.

Fleiss, J. L. (1999). Reliability of Measurement. In *The Design and Analysis of Clinical Experiments* (pp. 1-32): John Wiley & Sons, Inc.

Fritz, J. M., & Wainner, R. S. (2001). Examining diagnostic tests: an evidence-based perspective. *Physical Therapy, 81*, 1546-1564.

Gilchrist, J. M. (2009). Weighted 2 x 2 kappa coefficients: recommended indices of diagnostic accuracy for evidence-based practice. *J Clin Epidemiol, 62*, 1045-1053.

Hickey, B. W., Milosavljevic, S., Bell, M. L., & Milburn, P. D. (2007). Accuracy and reliability of observational motion analysis in identifying shoulder symptoms. *Man Ther, 12*, 263-270.

Hosseinimehr, S. H., Anbarian, M., Norasteh, A. A., Fardmal, J., & Khosravi, M. T. (2015). The comparison of scapular upward rotation and scapulohumeral rhythm between dominant and non-dominant shoulder in male overhead athletes and non-athletes. *Man Ther*.

Juul-Kristensen, B., Hilt, K., Enoch, F., Remvig, L., & Sjogaard, G. (2011). Scapular dyskinesis in trapezius myalgia and intraexaminer reproducibility of clinical tests. *Physiother Theory Pract, 27*, 492-502.

Karanicolas, P. J., Bhandari, M., Kreder, H., Moroni, A., Richardson, M., Walter, S. D., Norman, G. R., Guyatt, G. H., & Collaboration for Outcome Assessment in Surgical Trials Musculoskeletal, G. (2009). Evaluating agreement: conducting a reliability study. *Journal of Bone and Joint Surgery-American Volume, 91 Suppl 3*, 99-106.

Kawasaki, T., Yamakawa, J., Kaketa, T., Kobayashi, H., & Kaneko, K. (2012). Does scapular dyskinesis affect top rugby players during a game season? *J Shoulder Elbow Surg, 21*, 709-714.

Kibler, W. B. (1998). The role of the scapula in athletic shoulder function. *Am J Sports Med, 26*, 325-337.

Kibler, W. B., Ludewig, P. M., McClure, P., Uhl, T. L., & Sciascia, A. (2009). Scapular Summit 2009: introduction. July 16, 2009, Lexington, Kentucky. *J Orthop Sports Phys Ther, 39*, A1-A13.

Kibler, W. B., Ludewig, P. M., McClure, P. W., Michener, L. A., Bak, K., & Sciascia, A. D. (2013). Clinical implications of scapular dyskinesis in shoulder injury: the 2013 consensus statement from the 'Scapular Summit'. *Br J Sports Med, 47*, 877-885.

Kibler, W. B., & Sciascia, A. (2010). Current concepts: scapular dyskinesis. *Br J Sports Med*, *44*, 300-305.

Kibler, W. B., & Sciascia, A. D. (2013). Introduction to the Second International Conference on Scapular Dyskinesis in Shoulder Injury--the 'Scapular Summit' Report of 2013. *Br J Sports Med*, *47*, 874.

Kibler, W. B., Uhl, T. L., Maddux, J. W., Brooks, P. V., Zeller, B., & McMullen, J. (2002). Qualitative clinical evaluation of scapular dysfunction: a reliability study. *J Shoulder Elbow Surg*, *11*, 550-556.

Klintberg, I. H., Cools, A. M., Holmgren, T. M., Holzhausen, A. C., Johansson, K., Maenhout, A. G., Moser, J. S., Spunton, V., & Ginn, K. (2015). Consensus for physiotherapy for shoulder pain. *Int Orthop*, *39*, 715-720.

Kopkow, C., Lange, T., Hoyer, A., Lützner, J., & Schmitt, J. (2015). Physical tests for diagnosing anterior cruciate ligament rupture. In *Cochrane Database of Systematic Reviews*: John Wiley & Sons, Ltd.

Kopkow, C., Lange, T., Schmitt, J., & Kasten, P. (2015). Interrater reliability of the modified scapular assistance test with and without handheld weights. *Man Ther*.

Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B. J., Hrobjartsson, A., Roberts, C., Shoukri, M., & Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol*, *64*, 96-106.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Lange, T., Freiberg, A., Droge, P., Lutzner, J., Schmitt, J., & Kopkow, C. (2015). The reliability of physical examination tests for the diagnosis of anterior cruciate ligament rupture--A systematic review. *Man Ther*, *20*, 402-411.

Larsen, C. M., Juul-Kristensen, B., Lund, H., & Sogaard, K. (2014). Measurement properties of existing clinical assessment methods evaluating scapular positioning and function. A systematic review. *Physiother Theory Pract*, *30*, 453-482.

Lawrence, R. L., Braman, J. P., Laprade, R. F., & Ludewig, P. M. (2014). Comparison of 3-dimensional shoulder complex kinematics in individuals with and without shoulder pain, part 1: sternoclavicular, acromioclavicular, and scapulothoracic joints. *J Orthop Sports Phys Ther*, *44*, 636-645, A631-638.

Lewis, J., McCreesh, K., Roy, J. S., & Ginn, K. (2015). Rotator Cuff Tendinopathy: Navigating the Diagnosis-Management Conundrum. *J Orthop Sports Phys Ther*, 1-43.

Lewis, J. S., & Valentine, R. E. (2007). The pectoralis minor length test: a study of the intra-rater reliability and diagnostic accuracy in subjects with and without shoulder symptoms. *BMC Musculoskelet Disord*, *8*, 64.

Lewis, J. S., & Valentine, R. E. (2008). Intraobserver reliability of angular and linear measurements of scapular position in subjects with and without symptoms. *Arch Phys Med Rehabil, 89*, 1795-1802.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Plos Medicine, 6*, e1000100.

Looney, S. W., & Hagan, J. L. (2007). 4 Statistical Methods for Assessing Biomarkers and Analyzing Biomarker Data. In J. P. M. C. R. C.R. Rao (Ed.), *Handbook of Statistics* (Vol. Volume 27, pp. 109-147): Elsevier.

Lucas, N., Macaskill, P., Irwig, L., Moran, R., Rickards, L., Turner, R., & Bogduk, N. (2013). The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol, 13*, 111.

Lucas, N. P., Macaskill, P., Irwig, L., & Bogduk, N. (2010). The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol, 63*, 854-861.

Ludewig, P. M., & Cook, T. M. (2000). Alterations in shoulder kinematics and associated muscle activity in people with symptoms of shoulder impingement. *Physical Therapy, 80*, 276-291.

Ludewig, P. M., Phadke, V., Braman, J. P., Hassett, D. R., Cieminski, C. J., & LaPrade, R. F. (2009). Motion of the Shoulder Complex During Multiplanar Humeral Elevation. *Journal of Bone and Joint Surgery-American Volume, 91a*, 378-389.

Ludewig, P. M., & Reynolds, J. F. (2009). The association of scapular kinematics and glenohumeral joint pathologies. *J Orthop Sports Phys Ther, 39*, 90-104.

Matsuki, K., Matsuki, K. O., Mu, S., Yamaguchi, S., Ochiai, N., Sasho, T., Sugaya, H., Toyone, T., Wada, Y., Takahashi, K., & Banks, S. A. (2011). In vivo 3-dimensional analysis of scapular kinematics: comparison of dominant and nondominant shoulders. *Journal of Shoulder and Elbow Surgery, 20*, 659-665.

McClure, P., Greenberg, E., & Kareha, S. (2012). Evaluation and management of scapular dysfunction. *Sports Med Arthrosc, 20*, 39-48.

McClure, P., Tate, A. R., Kareha, S., Irwin, D., & Zlupko, E. (2009). A clinical method for identifying scapular dyskinesis, part 1: reliability. *J Athl Train, 44*, 160-164.

McClure, P. W., & Michener, L. A. (2015). Staged Approach for Rehabilitation Classification: Shoulder Disorders (STAR-Shoulder). *Physical Therapy, 95*, 791-800.

McClure, P. W., Michener, L. A., & Karduna, A. R. (2006). Shoulder function and 3-dimensional scapular kinematics in people with and without shoulder impingement syndrome. *Physical Therapy, 86*, 1075-+.

McClure, P. W., Michener, L. A., Sennet, B. J., & Karduna, A. R. (2001). Direct 3-dimensional measurement of scapular kinematics during dynamic movements in vivo. *Journal of Shoulder and Elbow Surgery, 10*, 269-277.

McCreesh, K. M., Crotty, J. M., & Lewis, J. S. (2015). Acromiohumeral distance measurement in rotator cuff tendinopathy: is there a reliable, clinically applicable method? A systematic review. *Br J Sports Med, 49*, 298-305.

McQuade, K. J., Borstad, J., & de Oliveira, A. S. (2016). Critical and Theoretical Perspective on Scapular Stabilization: What Does It Really Mean, and Are We on the Right Track? *Physical Therapy*.

Mell, A. G., LaScalza, S., Guffey, P., Ray, J., Maciejewski, M., Carpenter, J. E., & Hughes, R. E. (2005). Effect of rotator cuff pathology on shoulder rhythm. *J Shoulder Elbow Surg, 14,* 58S-64S.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res, 19*, 539-549.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol, 63*, 737-745.

Myers, J. B., Laudner, K. G., Pasquale, M. R., Bradley, J. P., & Lephart, S. M. (2005). Scapular position and orientation in throwing athletes. *Am J Sports Med, 33*, 263-271.

Myers, J. B., Laudner, K. G., Pasquale, M. R., Bradley, J. P., & Lephart, S. M. (2006). Glenohumeral range of motion deficits and posterior shoulder tightness in throwers with pathologic internal impingement. *American Journal of Sports Medicine, 34*, 385-391.

Myers, J. B., Oyama, S., & Hibberd, E. E. (2013). Scapular dysfunction in high school baseball players sustaining throwing-related upper extremity injury: a prospective study. *J Shoulder Elbow Surg, 22*, 1154-1159.

Nijs, J., Roussel, N., Vermeulen, K., & Souvereyns, G. (2005). Scapular positioning in patients with shoulder pain: a study examining the reliability and clinical importance of 3 clinical tests. *Arch Phys Med Rehabil, 86*, 1349-1355.

Odom, C. J., Taylor, A. B., Hurd, C. E., & Denegar, C. R. (2001). Measurement of scapular asymetry and assessment of shoulder dysfunction using the Lateral Scapular Slide Test: a reliability and validity study. *Physical Therapy, 81*, 799-809.

Ogston, J. B., & Ludewig, P. M. (2007). Differences in 3-dimensional shoulder kinematics between persons with multidirectional instability and asymptomatic controls. *Am J Sports Med, 35*, 1361-1370.

Oyama, S., Myers, J. B., Wassinger, C. A., Daniel Ricci, R., & Lephart, S. M. (2008). Asymmetric resting scapular posture in healthy overhead athletes. *J Athl Train, 43*, 565-570.

Patel, S., Hossain, F. S., Colaco, H. B., El-Husseiny, M., & Lee, M. H. (2011). The accuracy of primary care teams in diagnosing disorders of the shoulder. *Journal of Evaluation in Clinical Practice, 17*, 118-122.

Perez-Santonja, J. J., Bellot, J., Claramonte, P., Ismail, M. M., & Alio, J. L. (1997). Laser in situ keratomileusis to correct high myopia. *J Cataract Refract Surg, 23*, 372-385.

Pluim, B. M. (2013). Scapular dyskinesis: practical applications. *Br J Sports Med, 47*, 875-876.

Rabin, A., Irrgang, J. J., Fitzgerald, G. K., & Eubanks, A. (2006). The intertester reliability of the Scapular Assistance Test. *J Orthop Sports Phys Ther, 36*, 653-660.

Ratcliffe, E., Pickering, S., McLean, S., & Lewis, J. (2014). Is there a relationship between subacromial impingement syndrome and scapular orientation? A systematic review. *Br J Sports Med, 48*, 1251-1256.

Ribeiro, A., & Pascoal, A. G. (2013). Resting scapular posture in healthy overhead throwing athletes. *Man Ther, 18*, 547-550.

Roche, S. J., Funk, L., Sciascia, A., & Kibler, W. B. (2015). Scapular dyskinesis: the surgeon’s perspective. *Shoulder & Elbow, 7*, 289-297.

Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med, 21*, 3431-3446.

Sackett, D. L., & Haynes, R. B. (2002). The architecture of diagnostic research. *BMJ, 324*, 539-541.

Scholtes, V. A., Terwee, C. B., & Poolman, R. W. (2011). What makes a measurement instrument valid and reliable? *Injury, 42*, 236-240.

Seitz, A. L., McClure, P. W., Lynch, S. S., Ketchum, J. M., & Michener, L. A. (2012). Effects of scapular dyskinesis and scapular assistance test on subacromial space during static arm elevation. *J Shoulder Elbow Surg, 21*, 631-640.

Shadmehr, A., Bagheri, H., Ansari, N. N., & Sarafraz, H. (2010). The reliability measurements of lateral scapular slide test at three different degrees of shoulder joint abduction. *Br J Sports Med, 44*, 289-293.

Shitara, H., Kobayashi, T., Yamamoto, A., Shimoyama, D., Ichinose, T., Tajika, T., Osawa, T., Iizuka, H., & Takagishi, K. (2015). Prospective multifactorial analysis of preseason risk factors for shoulder and elbow injuries in high school baseball pitchers. *Knee Surgery, Sports Traumatology, Arthroscopy*.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy, 85*, 257-268.

Struyf, F., Cagnie, B., Cools, A., Baert, I., Brempt, J. V., Struyf, P., & Meeus, M. (2014). Scapulothoracic muscle activity and recruitment timing in patients with shoulder impingement symptoms and glenohumeral instability. *J Electromyogr Kinesiol, 24*, 277-284.

Struyf, F., Meeus, M., Fransen, E., Roussel, N., Jansen, N., Truijen, S., & Nijs, J. (2014). Interrater and intrarater reliability of the pectoralis minor muscle length measurement in subjects with and without shoulder impingement symptoms. *Man Ther, 19*, 294-298.

Struyf, F., Nijs, J., Meeus, M., Roussel, N. A., Mottram, S., Truijen, S., & Meeusen, R. (2014). Does scapular positioning predict shoulder pain in recreational overhead athletes? *Int J Sports Med, 35*, 75-82.

Struyf, F., Nijs, J., Mollekens, S., Jeurissen, I., Truijen, S., Mottram, S., & Meeusen, R. (2013). Scapular-focused treatment in patients with shoulder impingement syndrome: a randomized clinical trial. *Clin Rheumatol, 32*, 73-85.

Struyf, F., Nijs, J., Mottram, S., Roussel, N. A., Cools, A. M., & Meeusen, R. (2014). Clinical assessment of the scapula: a review of the literature. *Br J Sports Med, 48*, 883-890.

Sun, S. (2011). Meta-analysis of Cohen's kappa. *Health Services and Outcomes Research Methodology, 11*, 145-163.

T'Jonck, L. S., F.De Smet, L.Lysens, R. (2001). The relationship between clinical shoulder tests and the findings in arthroscopic examination. [Dutch] De relatie tussen klinische schoudertests en de bevindingen bij artroscopisch onderzoek. *Geneeskunde en Sport, 34*, 15-24.

Tate, A. R., McClure, P., Kareha, S., & Irwin, D. (2008). Effect of the scapula reposition test on shoulder impingement symptoms and elevation strength in overhead athletes. *Journal of Orthopaedic & Sports Physical Therapy, 38*, 4-11.

Terwee, C. B., de Winter, A. F., Scholten, R. J., Jans, M. P., Deville, W., van Schaardenburg, D., & Bouter, L. M. (2005). Interobserver reproducibility of the visual estimation of range of motion of the shoulder. *Arch Phys Med Rehabil, 86*, 1356-1361.

Timmons, M. K., Thigpen, C. A., Seitz, A. L., Karduna, A. R., Arnold, B. L., & Michener, L. A. (2012). Scapular kinematics and subacromial-impingement syndrome: a meta-analysis. *Journal of Sport Rehabilitation, 21*, 354-370.

Uhl, T. L., Kibler, W. B., Gecewich, B., & Tripp, B. L. (2009). Evaluation of clinical assessment methods for scapular dyskinesis. *Arthroscopy, 25*, 1240-1248.

Wassinger, C. A., Williams, D. A., Milosavljevic, S., & Hegedus, E. J. (2015). Clinical Reliability and Diagnostic Accuracy of Visual Scapulohumeral Movement Evaluation in Detecting Patients with Shoulder Impairment. *International Journal of Sports Physical Therapy*, *10*, 456-463.

Watson, L., Balster, S. M., Finch, C., & Dalziel, R. (2005). Measurement of scapula upward rotation: a reliable clinical procedure. *Br J Sports Med, 39*, 599-603.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19*, 231-240.

Whiting, P., Rutjes, A. W., Reitsma, J. B., Glas, A. S., Bossuyt, P. M., & Kleijnen, J. (2004). Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*, *140*, 189-202.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., & Group, Q.-S. (2013). A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol, 66*, 1093-1104.

Wright, A. A., Wassinger, C. A., Frank, M., Michener, L. A., & Hegedus, E. J. (2013). Diagnostic accuracy of scapular physical examination tests for shoulder disorders: a systematic review. *Br J Sports Med*, *47*, 886-892.

Yang, J. S., Keener, J. D., Yamaguchi, K., Chen, J., Stobbs-Cucchi, G., Patton, R., & Galatz, L. M. (2015). Reliability of patient self-assessment of shoulder range of motion and strength after shoulder arthroplasty. *J Shoulder Elbow Surg*, *24*, 1-10.

Yano, Y., Hamada, J., Tamai, K., Yoshizaki, K., Sahara, R., Fujiwara, T., & Nohara, Y. (2010). Different scapular kinematics in healthy subjects during arm elevation and lowering: Glenohumeral and scapulothoracic patterns. *Journal of Shoulder and Elbow Surgery, 19*, 209-215.

Table 1: Characteristics and results of included studies.

Study	Design	Objective	Subjects	Raters	Indextest	Method	Outcome
Baertschi, et al. (2013)	Interrater	The aim of this study was to evaluate the interrater reliability of four tests in the assessment of scapular motion during shoulder flexion; setting: Primary Care	PP: N= 39, sym sub= 19, asym sub= 20; sex: sym sub= M:F (10:9), asym sub= M:F (5:15); age: sym sub= (years: 46 ± 15, 21 - 66), asym sub= (years: 32 ± 9, 20 - 55); symptom duration: sym sub= (month: 26 ± 29, 1-120); pain: (VAS(0 - 100): 17 ± 12, 4 - 48); inclusion criteria: shoulder complaints and passive shoulder flexion at least 130° in standing position; exclusion criteria: predominantly glenohumeral restriction; prevalence: mixture; recruitment: unclear	N= 2 (physiotherapists); experience: not reported; qualification: further training from J.D. Stenvers at the training centre for neck, shoulder and arm disorders (NSA); training: additional training conducted, not specified; recruitment: not specified	Observation of scapular positioning / motion (dynamic): Scapular axillary hair test at the end of flexion; Clavicular movement during the first 60 degrees of flexion; Scapular posterior tilting during the last phase of flexion; Movement of the cervicothoracic junction during the last phase of flexion	Interrater reliability; trial: prospective; order of subjects: not relevant; order of raters: combination; blinding to other raters: unclear; blinding to own results: not relevant; blinding to reference standard: yes; blinding to other clinical information: yes; blinding to additional cues: unclear	Observation of scapular positioning / motion (dynamic): Cohen's kappa (k) and prevalence-adjusted-bias-adjusted kappa (PABAK) of dichotomous outcomes (data: positive/negative); measures of uncertainty: confidence intervals (CI), bias index (BI), prevalence index (PI); results stratified: asymptomatic and symptomatic subjects (additionally)
Curtis and Roush (2006)	Interrater	The aim of study was to determine the reliability of the LSST and its error between raters using a scoliometer; setting: Primary Care	PP: N= 33; sex: M:F (33:0); age: (years: 25.5 ± 5.69, 18-25); inclusion criteria: not reported; exclusion criteria: systemic disease that affects neuromuscular function, the inability to maintain at least 90 degrees of bilateral coronal plane shoulder abduction existence of any observed postural or bony deformities regardless of physician's diagnosis, or any existing medical	N= 3 (physical therapists); experience: 22.67 ± 2.52; qualification: not reported; training: additional training conducted, individually trained in the measurement procedure by the primary investigator; recruitment: not specified	Measurement of scapular positioning (resting position): Lateral Scapular Slide Test at 0° abduction; Measurement of scapular positioning (non resting position): Lateral Scapular Slide Test at 45° abduction and internal rotation, Lateral Scapular Slide Test at 90° abduction and maximal internal rotation	Interrater reliability; trial: prospective; order of subjects: not relevant; order of raters: fixed; blinding to other raters: yes; blinding to own results: not relevant; blinding to reference standard: yes; blinding to other clinical information: yes; blinding to additional cues: unclear	Measurement of scapular positioning (resting position/non resting position): intraclass correlation coefficient (ICC) of continuous outcomes (data: distance in cm); measures of uncertainty: standard error of measurements (SEM); results stratified: asymptomatic and symptomatic subjects

			diagnosis prohibiting the subject from participating in the study; prevalence: mixture; recruitment: convenience				
Juul-Kristensen, et al. (2011)	Intrarater	The aim of this study was to describe the intra-examiner reproducibility of specific clinical variables for scapular dyskinesis and report the presence of scapular dyskinesis in cases with trapezius myalgia and healthy controls, in addition to self-reported general health and work ability in these groups; setting: Unclear	PP: N=33, sym sub= 19; asym sub= 14; age: sym sub= (years: 44 ± 11.1); asym sub= (years: 41.5 ± 8.7); inclusion criteria: symptomatic subjects: pain in the neck-shoulder area, tightness of the trapezius muscle during passive lateral head flexion, palpable tenderness, carried out with an intended 4–kilogram pressure with a flat thumb perpendicular to the surface of the horizontal frontal border of the upper trapezius muscle; asymptomatic subjects: neck-shoulder trouble (ache, pain, or discomfort) for less than 8 days during the last year; a maximum of three body regions with 30 days of trouble; exclusion criteria: asymptomatic subjects: pain intensity within the last year of at least “quite a lot”, pain intensity during last 7 days, of at least “2” on a 10-point scale, pain frequency during last month, of at least “once a week”; prevalence: only	N= 1 (profession not reported); experience: 17; qualification: not reported; training: not reported; recruitment: not specified	Observation of scapular positioning / motion (static): Winging scapula during rest; Observation of scapular positioning / motion (dynamic): Winging scapula during slowly upper arm extension with a dumbbell (1.5 kg), Winging scapula during slowly upper arm extension; Measurement of scapular positioning (resting position): Static Alignment, Static Lower horizontal distance; Measurement of scapular positioning (non resting position): Dynamic Active proprioception/reposition error, Dynamic Maximum passive shoulder internal rotation, Dynamic Weakness of the scapula stabilizing muscles, Dynamic Upper arm angle at initial scapula movement, Dynamic Travelling distance	Intrarater reliability; trial: prospective; order of subjects: fixed; order of raters: not relevant; blinding to other raters: not relevant; blinding to own results: unclear; blinding to reference standard: yes; blinding to other clinical information: unclear; blinding to additional cues: unclear	Measurement of scapular positioning (resting position/non resting position): intraclass correlation coefficient (ICC) of continuous outcomes (data: angle in degree °, distance in cm and strength in Nm) and Observation of scapular positioning / motion (static(dynamic): Cohen’s kappa (k) of dichotomous outcomes (data: positive/negative); measures of uncertainty: confidence intervals (CI); results not stratified

			asymptomatic; recruitment: consecutive				
Kibler, et al. (2002)	Interrater & Intrarater	The aim of the study was to determine whether the observable characteristics of altered scapular motion could be delineated in a reliable manner among health care providers; setting: Unclear	PP: N= 26; age: (years: 29.5 ± 9); height: (cm: 178 ± 11.9); weight: (kg: 81.2 ± 15.95) ; inclusion criteria: not reported; exclusion criteria: bilateral symptoms, previous surgery, fracture in the shoulder girdle, or adhesive capsulitis; prevalence: mixture; recruitment: unclear	N= inter 4, intra 2 (2 physical therapists and 2 physicians); experience: not reported; qualification: not reported; training: additional training conducted, 10-minute visual and verbal presentation and a written description and video example of each abnormal pattern and a symmetric pattern; recruitment: not specified	Observation of scapular positioning / motion (dynamic): Observational classification 4-type method	Inter/Intra reliability; trial: prospective; order of subjects: variable, randomized; order of raters: not relevant; blinding to other raters: yes; blinding to own results: unclear; blinding to reference standard: unclear; blinding to other clinical information: unclear; blinding to additional cues: unclear	Observation of scapular positioning / motion (dynamic): Cohen’s kappa (k) of quadrotomous outcomes (data: type I to IV dyskinesis); measures of uncertainty: NR; results stratified: physicians and physical therapists rater
J. S. Lewis and Valentine (2007)	Intrarater	Primary purpose: investigate intra-rater reliability of the pectoralis minor length test in subjects without symptoms; secondary purpose: investigate intra-rater reliability of the pectoralis minor length test in subjects with symptoms; tertiary purpose: diagnostic accuracy of this clinical test of pectoralis minor length against the 'gold standard' recommendation of normal range being no greater than 2.6 cm above the table; setting: Secondary Care	PP: N= 90, sub sym= 45, asym sub= 45; sex: sub sym= M:F (22:23), asym sub= M:F (21:24); age: sub sym= (years: 42.8 ± 16.6, 19 - 84), asym sub= (years: 32.1 ± 7.3, 23 - 56); height: sym sub= (cm: 170 ± 10, 149 - 190), asym sub= (cm: 170 ± 10, 158 - 191); weight: sym sub= (kg: 71.4 ± 11.8, 49 - 90), asym sub= (kg: 70.4 ± 14.2, 50 - 111) ; inclusion criteria: subjects with symptoms: unilateral pain and/or restriction of movement arising from the area of the shoulder (C4/C5 dermatome); subjects without symptoms: no lumbar, thoracic, cervical or shoulder or upper limb symptoms; exclusion	N= 1 (profession not reported); experience: not reported; qualification: not reported; training: not reported; recruitment: not specified	Measurement of scapular positioning (resting position): The pectoralis minor length test / posterior acromion border to table	Intrarater reliability; trial: prospective; order of subjects: fixed; order of raters: not relevant; blinding to other raters: not relevant; blinding to own results: yes; blinding to reference standard: no; blinding to other clinical information: no; blinding to additional cues: unclear	Measurement of scapular positioning (resting position): intraclass correlation coefficient (ICC) of continuous outcomes (data: distance in mm); measures of uncertainty: confidence intervals (CI), standard error of measurements (SEM) for ICC; results stratified: asymptomatic and symptomatic subjects affected and non affected side, left and right side dominant and non dominant side

			criteria: for both groups: inability to fully communicate in English, subjects younger than 18 years of age, cardiac, respiratory, kidney, circulatory problems, systemic disease, diabetes, pregnancy, and, for female subjects pregnancy or suspicion of pregnancy; subjects without symptoms: a history of fractures, treatment or surgery to the lumbar, thoracic, cervical spine and upper limbs; prevalence: mixture; recruitment: referred				
J. S. Lewis and Valentine (2008)	Intrarater	The aim of this investigation was to determine the intratester reliability of a series of static linear and angular positions of the scapula during natural posture in subjects with and without shoulder symptoms; setting: Primary Care	PP: N= 90, sym sub= 45, asym sub= 44; sex: M:F (43:47) sym sub= M:F (22:23), asym sub= M:F (21:24); age: sym sub= (years: 43, 19 - 84), asym sub= (years: 32, 23 - 56); height: sym sub= (cm: 170, 150 - 190), sym sub= (cm: 170, 160 - 190); weight: sym sub= (kg: 71.4, 49 - 90), sym sub= (kg: 71.4, 49 - 90); inclusion criteria: symptomatic subjects: men and women over the age of 18 years, acute (6wk), subacute (6 –12wk), and chronic (3mo) pain conditions, and/or restriction of movement	N= 1 (physiotherapist); experience: 20; qualification: not reported; training: not reported; recruitment: not specified	Measurement of scapular positioning (resting position): Angular scapular measurement rotation, Angular scapular measurement tilt; Static scapular positioning using instruments: Linear measurement AC, BC, AB, DE, FE, FD with Anatomic reference points: A: Spinous process corresponding with the root of the spine of the scapula, B: root of the spine of the scapula, C: lateral end of the spine of the scapula, D: spinous process corresponding with the inferior angle of the	Intrarater reliability; trial: prospective; order of subjects: unclear; order of raters: unclear; blinding to other raters: not relevant; blinding to own results: unclear; blinding to reference standard: unclear; blinding to other clinical information: unclear; blinding to additional cues: unclear	Measurement of scapular positioning (resting position): intraclass correlation coefficient (ICC) of continuous outcomes (data: distance in cm and angle in degree °); measures of uncertainty: confidence intervals (CI); standard error of measurements (SEM) for ICC; results stratified asymptomatic and symptomatic subjects

			arising from the area of the shoulder (C5-6 dermatome, not reproduced with movement of the cervical spine); asymptomatic subjects: no lumbar, thoracic, cervical, shoulder, or upper-limb symptoms; exclusion criteria: for both groups, exclusion criteria were an inability to communicate fully in English; cardiac, respiratory, kidney, or circulatory problems, systemic disease, diabetes, and pregnancy; asymptomatic subjects: additional exclusion criteria were a history of fractures, treatment, or surgery to the lumbar spine, thoracic spine, cervical spine, and upper limbs; prevalence: mixture; recruitment: convenience		scapula, E: inferior angle of the scapula, F: T12 spinous process		
P. McClure, et al. (2009)	Interrater	The aim of the study was to determine the interrater reliability of a newly developed scapular dyskinesis test (SDT) that is visually based and uses dynamic, loaded tasks; setting: Primary Care	PP: N= 142; sex: M:F (111:31); inclusion criteria: subjects were required to complete all test movements to be included in the study; exclusion criteria: a current pain rating of 7/10 or greater on a numeric rating scale where 0 represents no pain and 10 represents the worst pain possible, a history of	N= live (2 von 5); video (3 pairs à 2 Rater) (athletic trainer, physical therapist, physical therapy student); experience: live observation: not reported; video observation: (years: 1-2, 7–20); qualification: not reported; training: additional training conducted, not specified; recruitment: not specified	Observation of scapular positioning / motion (dynamic): Observational classification 3-type method	Interrater reliability; trial: prospective; order of subjects: unclear; order of raters: not relevant; blinding to other raters: yes; blinding to own results: not relevant; blinding to reference standard: unclear; blinding to other clinical information: unclear; blinding to additional cues: unclear	Observation of scapular positioning / motion (dynamic): weighted kappa (weighted k) of trichotomous outcomes (data: grade: normal, subtle, obvious); measures of uncertainty: confidence intervals (CI), maximum kappa coefficients (kmax); results stratified: left and right side

			rotator cuff or glenoid labral tear, shoulder dislocation, fracture, or shoulder surgery within the past year, history of direct contact injury to the neck or upper extremities within the past 30 days, allergy to adhesives, or body mass index ≥ 30.0; prevalence: mixture; recruitment: convenience				
Nijs, et al. (2005)	Interrater	Primary purpose: examining the interobserver reliability; secondary purpose: examining the internal consistency; tertiary purpose: examining the clinical importance of 3 clinical tests for the assessment of scapular positioning in patients with shoulder pain; setting: Primary Care	PP: N= 29; sex: M:F (10:19); age: (years: 56.6.3 ± 14.9, 18 - 81); symptom duration: (months: 13.7 ± 27.1); pain: (VAS (0 - 100): 11.3 ± 16.3, 4 - 62); inclusion criteria: diagnosed as having a shoulder disorder by a physician, be referred by a physician for PT, and have shoulder pain at the time of the study; exclusion criteria: recent surgical intervention, visible injury related to the shoulder pain; prevalence: only symptomatic; recruitment: convenience	N= 2 (physiotherapists); experience: not reported; qualification: bachelor's degree in PT and final year of the master's degree program in PT; training: additional training conducted, 2-hour training session under supervision of 2 manual therapists; recruitment: not specified	Measurement of scapular positioning (resting position): Lateral Scapular Slide Test at 0° abduction, Medial scapular border to T4 retraction, The pectoralis minor length test / posterior acromion border to table relaxed, The pectoralis minor length test / posterior acromion border to table retraction; Measurement of scapular positioning (non resting position): Lateral Scapular Slide Test at 45° abduction and internal rotation, Lateral Scapular Slide Test at 90° abduction and maximal internal rotation	Interrater reliability; trial: prospective; order of subjects: not relevant; order of raters: unclear; blinding to other raters: yes; blinding to own results: not relevant; blinding to reference standard: unclear; blinding to other clinical information: unclear; blinding to additional cues: unclear	Measurement of scapular positioning (resting position/non resting position): intraclass correlation coefficient (ICC) of continuous outcomes (data: angle in degree ° and distance in cm); measures of uncertainty: NR; results stratified: left and right side
Odom, et al. (2001)	Interrater & Intrarater	Primary purpose: investigate the intrarater and interrater reliability of measurements obtained with the LSST; secondary purpose: examine the	PP: N= 46; age: (years: 30 ± 11.1, 18 - 65); inclusion criteria: age between 18 and 65 year, actively assume and maintain a position of at least 90 degrees of	N= 6 (physical therapists); experience: 5.8 ± 1.16, 4 - 7; qualification: not reported; training: additional training conducted, one session;	Measurement of scapular positioning (resting position): Lateral Scapular Slide Test at 0° abduction; Measurement of scapular positioning (non resting	Inter/Intra reliability; trial: prospective; order of subjects: unclear; order of raters: unclear; blinding to other raters: yes; blinding to own results: yes; blinding to	Measurement of scapular positioning (resting position/non resting position): intraclass correlation coefficient (ICC) of continuous

		validity of the LSST for classifying subjects based on the presence or absence of diagnosed shoulder impairment; setting: Primary Care	shoulder abduction in the coronal plane; exclusion criteria: postural or bony deformities, surgery within previous year, history of systemic disease that would affect neuromusculoskeletal function; prevalence: mixture; recruitment: convenience	recruitment: not specified	position): Lateral Scapular Slide Test at 45° abduction and internal rotation, Lateral Scapular Slide Test at 90° abduction and maximal internal rotation	reference standard: yes; blinding to other clinical information: no; blinding to additional cues: unclear	outcomes (data: distance in cm); measures of uncertainty: confidence intervals (CI), standard error of measurements (SEM) for ICC; results stratified: asymptomatic and symptomatic subjects
Rabin, et al. (2006)	Interrater	The aim of the study was to determine the interrater reliability of the modified SAT; setting: Primary Care	PP: N= 43; sex: M:F (30:16); age: (years: 44.5 ± 14.3); symptom duration: (months: 32.2 ± 58.7); dominant side: L:R (5:41); inclusion criteria: age 18 years or older, had a referral for physical therapy evaluation and treatment with a primary diagnosis of a shoulder disorder, and had reproduction of pain with active shoulder elevation in the scapular or sagittal plane; exclusion criteria: not have reproduction of their pain with active shoulder elevation in the scapular or sagittal plane, if they could not elevate their shoulder at least 90° in the scapular or sagittal plane, or if active elevation of the shoulder was contraindicated (i.e., early postoperative stages following rotator cuff	N= 2 of 11 (9 physical therapists, 2 physical therapy students); experience: for physical therapists 2 to more than 20; qualification: not reported; training: one formal training session, if necessary further training was conducted; recruitment:	Symptom alterations test: Modified Scapular Assistance Test	Interrater reliability; trial: prospective; order of subjects: not relevant; order of raters: unclear; blinding to other raters: yes; blinding to own results: not relevant; blinding to reference standard: unclear; blinding to other clinical information: unclear; blinding to additional cues: unclear	Symptom alterations test: Cohen's kappa (k) of dichotomous outcomes (data: positive/negative); measures of uncertainty: NR; results not stratified

			repair); prevalence: only symptomatic; recruitment: referred				
Shadmehr, et al. (2010)	Interrater & Intrarater	The aim of the study was to determine the reliability and diagnostic accuracy of the LSST using the caliper; setting: Primary Care	PP: N= 57; age: (years: 18 - 65); inclusion criteria: diagnosed shoulder pathology by an orthopedic surgeon; had to be able to actively perform 0°, 45° and 90° of abduction and full internal rotation of the shoulder; pathology on one shoulder; exclusion criteria: previous shoulder surgery, a history of systemic disease, neuromuscular dysfunction and obesity (body mass index 29.9 or greater); prevalence: mixture; recruitment: referred	N= 3 (physical therapists); experience: not reported; qualification: not reported; training: 2h single training sessions; recruitment: not specified	Measurement of scapular positioning (resting position): Lateral Scapular Slide Test at 0° abduction; Measurement of scapular positioning (non resting position): Lateral Scapular Slide Test at 45° abduction and internal rotation, Lateral Scapular Slide Test at 90° abduction and maximal internal rotation	Inter/Intra reliability; trial: prospective; order of subjects: variable, randomized; order of raters: variable, randomized; blinding to other raters: unclear; blinding to own results: unclear; blinding to reference standard: unclear; blinding to other clinical information: unclear; blinding to additional cues: unclear	Measurement of scapular positioning (resting position/non resting position): intraclass correlation coefficient (ICC) of continuous outcomes (data: distance in cm); measures of uncertainty: confidence intervals (CI), standard error of measurements (2 SEM) for ICC; results stratified: asymptomatic and symptomatic subjects affected and non affected side, left and right side
Struyf, Meeus, et al. (2014)	Interrater & Intrarater	The purpose of this study was to investigate the reliability of the PML measurement in patients with and without SIS; setting: Primary Care	PP: N= 50, sym sub= 25, asym sub= 25; sex: M:F (24: 26), sym sub= M:F (8:17), asym sub= M:F (16:9); age: sym sub= (years, 50.8 ± 16.3), asym sub= (years, 20.8 ± 1.5); height: sym sub= (cm, 169.2 ± 9.9), asym sub= (cm, 175.7 ± 9.8); weight: sym sub= (kg, 74.5 ± 15.5), asym sub= (kg, 72.6 ± 13.7); inclusion criteria: (1) age over 18, (2) shoulder pain present for at least 15 consecutive days (3) a pain score of at least 10% on a Visual Analogue Scale	N= 2 (physiotherapists); experience: 1 year; qualification: not reported; training: 2 hours training session; recruitment: unclear	Measurement of scapular positioning (resting position): The pectoralis minor length test / posterior acromion border to table	Inter/Intra reliability; trial: prospective; order of subjects: unclear; order of raters: variable, randomized; blinding to other raters: yes; blinding to own results: yes; blinding to reference standard: unclear; blinding to other clinical information: yes; blinding to additional cues: unclear	Measurement of scapular positioning (resting position): intraclass correlation coefficient (ICC) of continuous outcomes (data percentage); measures of uncertainty: confidence intervals (CI), standard error of measurements (SEM) for ICC, minimal detectable change at the 95% confidence level (MDC95); results stratified asymptomatic and symptomatic subjects affected and non affected side, dominant and non

			(VAS), a score of at least 10% on the Shoulder Disability Questionnaire (SDQ), and at least 2 out of 3 impingement tests had to be positive for inclusion (Hawkins test, Neer test & Jobe test were performed) Healthy controls were recruited by the following inclusion criteria: (1) age over 18, (2) the absence of shoulder pain (<5% on VAS or SDQ) during the last year and (3) the absence of a history of fractures, treatment or surgery of the shoulder girdle; exclusion criteria: (1) the inability to communicate using the Dutch language, (2) presence of a systemic disease, influencing shoulder pathology, (3) pregnancy, (4) traumatic shoulder pathologies and (5) shoulder pain radiating from the cervical spine; prevalence: mixture; recruitment: convenience				dominant side
Terwee, et al. (2005)	Interrater	Primary purpose: evaluate, in a large population of patients, interobserver reproducibility (agreement and reliability) with regard to active elevation and several passive shoulder movements; secondary	PP: N= 201; sex: M:F (68:133); age: (years: 48 ± 12); symptom duration: (months: < 3, N= 27), (months: 3 - 6, N= 16), (months: 6 - 12, N= 22), (months: > 12, N= 35); inclusion criteria: between	N= 2 (physiotherapists); experience: 3 and 10; qualification: not reported; training: additional training conducted, visually assessment compared with goniometric measurement; recruitment: not specified	Measurement of scapular positioning (non resting position): Scapulohumeral abduction	Interrater reliability; trial: prospective; order of subjects: not relevant; order of raters: variable, randomized; blinding to other raters: yes; blinding to own results: not relevant; blinding to reference	Measurement of scapula positioning (non resting position): intraclass correlation coefficien (ICC) of continuous outcomes (data: angle in degree °); measures of uncertainty: confidence

		purpose: evaluate whether clinical characteristics are associated with the level of interobserver reproducibility; setting: Primary Care	18 and 75 years of age, were able to complete questionnaires (e.g., no dementia); exclusion criteria: shoulder problems resulting from neurologic, vascular, or internal disorders; systemic rheumatic diseases; fractures; or dislocations; prevalence: only symptomatic; recruitment: consecutive			standard: unclear; blinding to other clinical information: no; blinding to additional cues: unclear	intervals (CI); results stratified: affected and non affected side
Uhl, et al. (2009)	Interrater	Primary purpose: assess the inter-rater reliability and validity of 2 clinical assessment methods of categorizing scapular dyskinesis by observation; secondary purpose: quantify the prevalence of asymmetry of bilateral scapular motion in individuals with and without shoulder symptoms by use of 3D kinematic analysis; setting: Primary Care	PP: N= 56; sex: M:F (21:35), sym sub= M:F (24:11), asym sub= M:F (11:10); age: sym sub= (years: 32 ± 11), asym sub= (years: 24 ± 3); dominant side: L:R (7:49), sym sub= L:R (5:30), asym sub= L:R (2:19); height: sym sub= (cm: 175 ± 9), asym sub= (cm: 172 ± 9); weight: sym sub= (kg: 80 ± 16), asym sub= (kg: 75 ± 18) ; inclusion criteria: not reported; exclusion criteria: bilateral shoulder pain, previous fracture of the scapula, humerus, or clavicle, history or evidence of injury to the long thoracic, spinal accessory, or cervical root nerve; prevalence: mixture; recruitment: convenience	N= 2 (physical therapist and physician); experience: not reported; qualification: not reported; training: not reported; recruitment: not specified	Observation of scapular positioning / motion (dynamic): Observational classification 4-type and 2-type method	Interrater reliability; trial: prospective; order of subjects: unclear; order of raters: unclear; blinding to other raters: yes; blinding to own results: not relevant; blinding to reference standard: unclear; blinding to other clinical information: unclear; blinding to additional cues: unclear	Observation of scapular positioning / motion (dynamic): Cohen's kappa (k) of quadrotomous outcomes (data: type I to IV dyskinesis) and Cohen's kappa (k) of dichotomous outcomes (data: positive/negative); measures of uncertainty: NR; results not stratified

Watson, et al. (2005)	Intrarater	The aim of the study was to determine if the Plurimeter-V gravity referenced inclinometer was reliable for measuring scapular upward rotation positions during total shoulder abduction (TSA) in a group of patients with shoulder pathology; setting: Primary Care	PP: N= 26; sex: M:F (11:15); age: (years: 29, SEM= 2.5); symptom duration: (months: 45, SEM= 11.3); dominant side: L:R (1:25); inclusion criteria: pain and/or dysfunction in one or both shoulders, a diagnosis of shoulder pathology determined by an orthopaedic surgeon, >140˚ of humeral elevation in the coronal plane (abduction) as measured by the Plurimeter-V inclinometer (using a previously developed protocol); exclusion criteria: congenital defect of the scapula or thorax; history of trauma, fracture, or surgery to the scapula, rib cage, or thoracic spine; any evidence of peripheral or central nerve lesion, such as long thoracic nerve or accessory nerve; known neuromuscular disorders, <140˚ of combined TSA, an irritable shoulder condition, defined as one with shoulder pain increasing appreciably with repeated motion and being slow to settle once aggravated; prevalence: mixture; recruitment: convenience	N= 1 (profession not reported); experience: previous experience using the Plurimeter-V inclinometer; qualification: not reported; training: not reported; recruitment: not specified	Measurement of scapular positioning (non resting position): Scapula upward rotation: total shoulder abduction at rest, Scapula upward rotation: total shoulder abduction at 45°, Scapula upward rotation: total shoulder abduction at 90°, Scapula upward rotation: total shoulder abduction at 135°, Scapula upward rotation: total shoulder abduction at end of range	Intrarater reliability; trial: prospective; order of subjects: fixed; order of raters: not relevant; blinding to other raters: not relevant; blinding to own results: unclear; blinding to reference standard: unclear; blinding to other clinical information: unclear; blinding to additional cues: unclear	Measurement of scapula[r] positioning (non restin[g] position): intraclas[s] correlation coefficien[t] (ICC) of continuou[s] outcomes (data: angle i[n] degree °); measures o[f] uncertainty: confidenc[e] intervals (CI), standar[d] error of measurement[s] (SEM) for ICC; results no[t] stratified

Abbreviations:

BI = Bias index; F = Female; ICC = Intraclass correlation coefficient; k = Kappa; kmax = Maximum kappa coefficients; kw = Weighted Kappa; L = Left side; LSST = Lateral Scapular Slide Test; M = Male; MDC = Minimal detectable change; N= Sample size; PAPAK = Prevalence - adjusted bias - adjusted Kappa; PI = Prevalance index; PP = Per protocol; PT = Physical therapy; R = Right side; SAT = Scapular assistance test; SEM = Standard error of the measurement; sym sub = Symptomatic subjects; VAS = Visual analog scale; wk = Week

Table 2: Overview of risk of bias assessment utilized with Quality Appraisal of Reliability Studies (QAREL) checklist

Study	Design	QAREL Items											Total
		1	2	3	4	5	6	7	8	9	10	11	
Juul-Kristensen et al. (2011)	Intrarater	Y	Y	NA	U	Y	U	U	N	Y	Y	Y	6/11
Kibler et al. (2002)	Intrarater	Y	Y	Y	U	U	U	U	Y	Y	Y	Y	7/11
J. S. Lewis and Valentine (2007)	Intrarater	Y	U	NA	Y	N	N	U	N	N	Y	Y	4/11
J. S. Lewis and Valentine (2008)	Intrarater	Y	Y	NA	U	U	U	U	U	Y	Y	Y	5/11
Odom et al. (2001)	Intrarater	Y	Y	Y	Y	Y	N	U	U	U	Y	Y	7/11
Shadmehr et al. (2010)	Intrarater	Y	Y	NA	U	U	U	U	Y	N	Y	Y	5/11
Struyf, Meeus, et al. (2014)	Intrarater	Y	Y	Y	Y	U	Y	U	U	U	Y	Y	7/11

Watson et al. (2005)	Intrarater	Y	U	NA	U	U	U	U	N	Y	Y	Y	4/11
Baertschi et al. (2013)	Interrater	Y	Y	U	NA	Y	Y	U	U	Y	Y	Y	7/11
Curtis and Roush (2006)	Interrater	Y	Y	Y	NA	Y	Y	U	N	U	Y	Y	7/11
Kibler et al. (2002)	Interrater	Y	Y	Y	U	U	U	U	NA	U	Y	Y	5/11
P. McClure et al. (2009)	Interrater*	Y	Y	N	NA	U	U	U	U	U	Y	Y	4/11
P. McClure et al. (2009)	Interrater#	Y	Y	Y	NA	U	U	U	U	U	Y	Y	5/11
Nijs et al. (2005)	Interrater	Y	Y	Y	NA	Y	U	U	U	U	Y	Y	6/11
Odom et al. (2001)	Interrater	Y	Y	Y	Y	Y	N	U	U	U	Y	Y	7/11
Rabin et al. (2006)	Interrater	Y	Y	Y	NA	U	U	U	U	U	Y	Y	5/11
Shadmehr et al. (2010)	Interrater	Y	Y	U	NA	U	U	U	Y	N	Y	Y	5/11
Struyf, Meeus, et al. (2014)	Interrater	Y	Y	Y	Y	U	Y	U	U	U	Y	Y	7/11
Terwee et al. (2005)	Interrater	Y	Y	Y	NA	U	N	U	Y	U	Y	Y	6/11
Uhl et al. (2009)	Interrater	Y	Y	Y	NA	U	U	U	U	U	Y	Y	5/11

* = live observation; # = video observation

Abbreviation: Y = yes; N = no; U = unclear; NA = not applicable

QAREL Items: 1. Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied? 2. Was the test performed by raters who were representative of those to whom the authors intended the results to be applied? 3. Were raters blinded to the findings of other raters during the study? 4. Were raters blinded to their own prior findings of the test under evaluation? 5. Were raters blinded to the results of the reference standard for the target disorder (or variable) being evaluated? 6. Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design? 7. Were raters blinded to additional cues that were not part of the test? 8. Was the order of examination varied? 9. Was the time interval between repeated measurements compatible with the stability (or theoretical stability) of the variable being measured? 10. Was the test applied correctly and interpreted appropriately? 11. Were appropriate statistical measures of agreement used?

Table 3: Results of intrarater reliability.

Indextest	Study	Instruments	Test position	Prevalence	$N_{subject}$	$N_{shoulder}$	N_{rater}	N_{trial}	Estimates of reliability (95 % CI)	Po (%)	Measures of uncertainty	Strength of agreement	QAR Score

Observation of scapular positioning / motion (static)													
Winging scapula during rest	Juul-Kristensen, et al. (2011)	NA	standing	0.58	33	33	1	2	k= 1.00 (0.00, 1.00)	NR	NR	Almost Perfect[L]	6/11
Observation of scapular positioning / motion (dynamic)													
Observational classification 4-type method	Kibler, et al. (2002)	NA	standing	0.23	26	26	2	2	Rater A: k= 0.59 (NR); Rater B: k= 0.49 (NR)	NR	NR	Moderate[L]	7/11
Winging scapula during slowly upper arm extension	Juul-Kristensen, et al. (2011)	NA	standing	0.58	33	33	1	2	k= 1.00 (0.00, 1.00)	NR	NR	Almost Perfect[L]	6/11
Winging scapula during slowly upper arm extension with a dumbbell (1.5 kg)	Juul-Kristensen, et al. (2011)	NA	standing	0.58	33	33	1	2	k= 0.84 (0.53, 1.15)	NR	NR	Almost Perfect[L]	6/11
Measurement of scapular positioning (resting position)													
Angular scapular measurement rotation	J. S. Lewis and Valentine (2008)	inclinometer	standing	1.00	90	180	1	3	L: ICC(2,1)= 0.92 (0.86 - 0.95); ICC(2,3)= 0.96 (0.92 - 0.98); R: ICC(2,1)= 0.94 (0.89 - 0.97); ICC(2,3)= 0.97 (0.95 - 0.98)	NA	L: SEM for ICC(2,3)= 0.9; 2 SEM for ICC(2,3)= 1.8; R: SEM for ICC(2,3)= 0.7; SEM for ICC(2,3)= 1.4	Excellent[F]	5/11
	J. S. Lewis and Valentine (2008)	inclinometer	standing	0.00	90	180	1	3	L: ICC(2,1)= 0.90 (0.82 - 0.94); ICC(2,3)= 0.95 (0.90, 0.97); R: ICC(2,1)= 0.84 (0.72 - 0.91); ICC(2,3)= 0.92 (0.83, 0.96)	NA	L: SEM for ICC(2,3)= 0.9; 2 SEM for ICC(2,3)= 1.8; R: SEM for ICC(2,3)= 1.2; SEM for ICC(2,3)= 2.4	Excellent[F]	5/11
Angular scapular measurement tilt	J. S. Lewis and Valentine (2008)	inclinometer	standing	1.00	90	180	1	3	L: ICC(2,1)= 0.94 (0.89 - 0.97); ICC(2,3)= 0.97 (0.94 - 0.98); R: ICC(2,1)= 0.95 (0.93 - 0.98); ICC(2,3)= 0.98 (0.96 - 0.99)	NA	L: SEM for ICC(2,3)= 1.0; 2 SEM for ICC(2,3)= 2.0; R: SEM for ICC(2,3)= 0.8;	Excellent[F]	5/11

										SEM for ICC(2,3)= 1.6		
	J. S. Lewis and Valentine (2008)	inclinometer	standing	0.00	90	180	1	3	L: ICC(2,1)= 0.94 (0.88 - 0.97); ICC(2,3)= 0.97 (0.94, 0.98); R: ICC(2,1)= 0.95 (0.95 - 0.99); ICC(2,3)= 0.98 (0.95, 0.99)	NA	L: SEM for ICC(2,3)= 1.0; 2 SEM for ICC(2,3)= 2.0; R: SEM for ICC(2,3)= 0.9; SEM for ICC(2,3)= 1.8	Excellent[F]	5/11
Lateral Scapular Slide Test at 0° abduction	Odom, et al. (2001)	tape measure	standing	1.00	20	40	6	2	ICC(1,1)= 0.52 (0.10, 0.74), within-day reliability[#]	NA	SEM for ICC(1,1)= 0.78	Fair to good[F]	7/11
	Odom, et al. (2001)	tape measure	standing	0.00	26	54	6	2	ICC(1,1)= 0.75 (0.56, 0.85), within-day reliability[#]	NA	SEM for ICC(1,1)= 0.61	Fair to good[F]	7/11
	Shadmehr, et al. (2010)	scoliometer	standing	1.00	27	54	3	2	A: ICC(2,1)= 0.87 (NR)[ICC]; Non-A: ICC(2,1)= 0.88 (NR) [ICC]	NA	A: 2 SEM for ICC(2,1)= 7.9 (0.73, 0.93); Non-A: 2 SEM for ICC(2,1)= 7.6 (0.75, 0.94)	Excellent[F]	5/11
	Shadmehr, et al. (2010)	scoliometer	standing	0.00	30	60	3	2	L: ICC(2,1)= 0.95 (NR)[ICC]; R: ICC(2,1)= 0.95 (NR)[ICC]	NA	L: 2 SEM for ICC(2,1)= 7.9 (0.90, 0.97); R: 2 SEM for ICC(2,1)= 7.4 (0.91, 0.98)	Excellent[F]	5/11
Linear measurement AB	J. S. Lewis and Valentine (2008)	tape measure	standing	1.00	90	180	1	3	L: ICC(2,1)= 0.83 (0.72 - 0.91); ICC(2,3)= 0.91 (0.83 - 0.95); R: ICC(2,1)= 0.90 (0.83 - 0.94); ICC(2,3)= 0.95 (0.90 - 0.97)	NA	L: SEM for ICC(2,3)= 0.3; 2 SEM for ICC(2,3)= 0.6; R: SEM for ICC(2,3)= 0.3; SEM for ICC(2,3)= 0.6	Excellent[F]	5/11
	J. S. Lewis and Valentine (2008)	tape measure	standing	0.00	90	180	1	3	L: ICC(2,1)= 0.90 (0.83 - 0.95); ICC(2,3)= 0.95 (0.91 - 0.97); R: ICC(2,1)= 0.97 (0.95 - 0.98); ICC(2,3)=	NA	L: SEM for ICC(2,3)= 0.2; 2 SEM for ICC(2,3)= 0.4; R:	Excellent[F]	5/11

									0.98 (0.97 - 0.99)		SEM for ICC(2,3)= 0.2; SEM for ICC(2,3)= 0.4		
Linear measurement AC	J. S. Lewis and Valentine (2008)	tape measure	standing	1.00	90	180	1	3	L: ICC(2,1)= 0.94 (0.90 - 0.97); ICC(2,3)= 0.97 (0.95 - 0.98); R: ICC(2,1)= 0.94 (0.90 - 0.97); ICC(2,3)= 0.97 (0.95 - 0.98)	NA	L: SEM for ICC(2,3)= 0.3; 2 SEM for ICC(2,3)= 0.6; R: SEM for ICC(2,3)= 0.3; SEM for ICC(2,3)= 0.6	Excellent[F]	5/11
	J. S. Lewis and Valentine (2008)	tape measure	standing	0.00	90	180	1	3	L: ICC(2,1)= 0.95 (0.92 - 0.98); ICC(2,3)= 0.98 (0.96 - 0.99); R: ICC(2,1)= 0.89 (0.80 - 0.94); ICC(2,3)= 0.94 (0.89 - 0.97)	NA	L: SEM for ICC(2,3)= 0.3; 2 SEM for ICC(2,3)= 0.6; R: SEM for ICC(2,3)= 0.5; SEM for ICC(2,3)= 1.0	Excellent[F]	5/11
Linear measurement BC	J. S. Lewis and Valentine (2008)	tape measure	standing	1.00	90	180	1	3	L: ICC(2,1)= 0.89 (0.80 - 0.94); ICC(2,3)= 0.94 (0.89 - 0.97); R: ICC(2,1)= 0.83 (0.71 - 0.90); ICC(2,3)= 0.91 (0.83 - 0.95)	NA	L: SEM for ICC(2,3)= 0.3; 2 SEM for ICC(2,3)= 0.6; R: SEM for ICC(2,3)= 0.3; SEM for ICC(2,3)= 0.6	Excellent[F]	5/11
	J. S. Lewis and Valentine (2008)	tape measure	standing	0.00	90	180	1	3	L: ICC(2,1)= 0.89 (0.82 - 0.94); ICC(2,3)= 0.95 (0.90 - 0.97); R: ICC(2,1)= 0.90 (0.82 - 0.94); ICC(2,3)= 0.95 (0.90 - 0.97)	NA	L: SEM for ICC(2,3)= 0.3; 2 SEM for ICC(2,3)= 0.6; R: SEM for ICC(2,3)= 0.3; SEM for ICC(2,3)= 0.6	Excellent[F]	5/11
Linear measurement DE	J. S. Lewis and Valentine	tape measure	standing	1.00	90	180	1	3	L: ICC(2,1)= 0.89 (0.81 - 0.94); ICC(2,3)= 0.93 (0.89	NA	L: SEM for ICC(2,3)= 0.4; 2	Excellent[F]	5/11

	(2008)								- 0.97); R: ICC(2,1)= 0.92 (0.86 - 0.95); ICC(2,3)= 0.96 (0.92 - 0.98)		SEM for ICC(2,3)= 0.8; R: SEM for ICC(2,3)= 0.3; SEM for ICC(2,3)= 0.6		
	J. S. Lewis and Valentine (2008)	tape measure	standing	0.00	90	180	1	3	L: ICC(2,1)= 0.79 (0.66 - 0.88); ICC(2,3)= 0.88 (0.79 - 0.94); R: ICC(2,1)= 0.87 (0.73 - 0.94); ICC(2,3)= 0.93 (0.84 - 0.97)	NA	L: SEM for ICC(2,3)= 0.5; 2 SEM for ICC(2,3)= 1.0; R: SEM for ICC(2,3)= 0.4; SEM for ICC(2,3)= 0.8	Excellent[F]	5/11
Linear measurement FD	J. S. Lewis and Valentine (2008)	tape measure	standing	1.00	90	180	1	3	L: ICC(2,1)= 0.91 (0.84 - 0.95); ICC(2,3)= 0.95 (0.92 - 0.97); R: ICC(2,1)= 0.89 (0.81 - 0.94); ICC(2,3)= 0.94 (0.89 - 0.98)	NA	L: SEM for ICC(2,3)= 0.5; 2 SEM for ICC(2,3)= 1.0; R: SEM for ICC(2,3)= 0.5; SEM for ICC(2,3)= 1.0	Excellent[F]	5/11
	J. S. Lewis and Valentine (2008)	tape measure	standing	0.00	90	180	1	3	L: ICC(2,1)= 0.91 (0.84 - 0.95); ICC(2,3)= 0.95 (0.91 - 0.97); R: ICC(2,1)= 0.90 (0.83 - 0.95); ICC(2,3)= 0.95 (0.91 - 0.97)	NA	L: SEM for ICC(2,3)= 0.5; 2 SEM for ICC(2,3)= 1.0; R: SEM for ICC(2,3)= 0.5; SEM for ICC(2,3)= 1.0	Excellent[F]	5/11
Linear measurement FE	J. S. Lewis and Valentine (2008)	tape measure	standing	1.00	90	180	1	3	L: ICC(2,1)= 0.94 (0.89 - 0.97); ICC(2,3)= 0.97 (0.94 - 0.98); R: ICC(2,1)= 0.92 (0.86 - 0.96); ICC(2,3)= 0.96 (0.93 - 0.98)	NA	L: SEM for ICC(2,3)= 0.4; 2 SEM for ICC(2,3)= 0.8; R: SEM for ICC(2,3)= 0.4; SEM for ICC(2,3)= 0.8	Excellent[F]	5/11

	J. S. Lewis and Valentine (2008)	tape measure	standing	0.00	90	180	1	3	L: ICC(2,1)= 0.93 (0.88 - 0.96); ICC(2,3)= 0.97 (0.94 - 0.98); R: ICC(2,1)= 0.92 (0.85 - 0.95); ICC(2,3)= 0.96 (0.92 - 0.98)	NA	L: SEM for ICC(2,3)= 0.4; 2 SEM for ICC(2,3)= 0.8; R: SEM for ICC(2,3)= 0.5; SEM for ICC(2,3)= 1.0	Excellent[F]	5/11
Scapula upward rotation: total shoulder abduction at rest	Watson, et al. (2005)	inclinometer	standing	1.00	26	26	1	2	ICC(NR)= 0.94 (0.90, 0.99), within-day reliability	NA	SEM for ICC(NA)= 1.7	Excellent[F]	4/11
Static Alignment	Juul-Kristensen, et al. (2011)	NA	standing	0.58	33	33	1	2	ICC(2,1)= 0.85 (0.71, 0.92)	NA	NR	Excellent[F]	6/11
Static Lower horizontal distance	Juul-Kristensen, et al. (2011)	tape measure	standing	0.58	33	33	1	2	ICC(2,1)= 0.91 (0.82, 0.95)	NA	NR	Excellent[F]	6/11
The pectoralis minor length test / posterior acromion border to table	J. S. Lewis and Valentine (2007)	rigid standard plastic transparent right angle	supine	1.00	45	90	1	3	L: ICC(2,3)= 0.92 (0.94, 0.98), ICC(2,1)= 0.93 (0.88, 0.96); R: ICC(2,3)= 0.96 (0.93, 0.98), ICC(2,1)= 0.92 (0.86, 0.96); D: ICC(2,3)= 0.96 (0.94, 0.98), ICC(2,1)= 0.93 (0.86, 0.96); Non-D: ICC(2,3)= 0.96 (0.93, 0.98), ICC(2,1)= 0.92 (0.87, 0.96)	NA	L: SEM for ICC(2,3)= 3, SEM for ICC(2,1)= 4; R: SEM for ICC(2,3)= 3, SEM for ICC(2,1)= 4 ; D: SEM for ICC(2,3)= 3, SEM for ICC(2,1)= 4; Non-D: SEM for ICC(2,3)= 3, SEM for ICC(2,1)= 4	Excellent[F]	4/11
	J. S. Lewis and Valentine (2007)	rigid standard plastic transparent right angle	supine	0.00	45	90	1	3	A: ICC(2,3)= 0.95 (0.91, 0.97), ICC(2,1)= 0.91 (0.84, 0.95); A: ICC(2,3)= 0.97 (0.94, 0.98), ICC(2,1)= 0.93 (0.87, 0.96); L: ICC(2,3)= 0.97 (0.94, 0.98), ICC(2,1)= 0.92 (0.86, 0.96); R: ICC(2,3)= 0.95 (0.91, 0.97), ICC(2,1)= 0.90 (0.81, 0.94); D: ICC(2,3)= 0.95 (0.92, 0.98), ICC(2,1)= 0.92 (0.86,	NA	A: SEM for ICC(2,3)= 4, SEM for ICC(2,1)= 5; Non-A: SEM for ICC(2,3)= 3, SEM for ICC(2,1)= 4; L: SEM for ICC(2,3)= 3, SEM for ICC(2,1)= 4; R: SEM for ICC(2,3)= 4, SEM	Excellent[F]	4/11

									0.95); Non-D: ICC(2,3)= 0.96 (0.92, 0.98),ICC(2,1)= 0.91 (0.84, 0.95)		for ICC(2,1)= 5; D: SEM for ICC(2,3)= 4, SEM for ICC(2,1)= 5; Non-D: SEM for ICC(2,3)= 3, SEM for ICC(2,1)= 4		
	Struyf, Meeus, et al. (2014)	protractor	supine	1.00	25	50	2	2	Rater A: A: ICC (2,1)= 0.91 (0.83 - 0.95)[*]; Non-A: ICC (2,1)= 0.93 (0.87 - 0.97)[*]; Rater B: A: ICC (2,1)= 0.87 (0.76 - 0.93)[*]; Non-A: ICC (2,1)= 0.93 (0.95 - 0.99)[*]	NA	Rater A: A: SEM for ICC= 0.21, MDC(95)= 0.57; Non-A: SEM for ICC= 0.19, MDC(95)= 0.54; Rater B: A: SEM for ICC= 0.27, MDC(95)= 0.75; Non-A: SEM for ICC= 0.30, MDC(95)= 0.84	Excellent[F]	7/11
	Struyf, Meeus, et al. (2014)	protractor	supine	0.00	25	50	2	2	Rater A: D: ICC (2,1)= 0.76 (0.59 - 0.87)[*]; Non-D: ICC (2,1)= 0.87 (0.77 - 0.94)[*]; Rater B: D: ICC (2,1)= 0.80 (0.65 - 0.90)[*]; Non-D: ICC (2,1)= 0.78 (0.63 - 0.89)[*]	NA	Rater A: D: SEM for ICC= 0.29, MDC(95)= 0.80; Non-D: SEM for ICC= 0.21, MDC(95)= 0.59; Rater B: D: SEM for ICC= 0.32, MDC(95)= 0.89; Non-D: SEM for ICC= 0.32, MDC(95)= 0.88	Excellent[F]	7/11
Measurement of scapular positioning (non resting position)													
Dynamic Active proprioception/reposition error	Juul-Kristensen, et al. (2011)	laser pointer	sitting	0.58	33	33	1	2	ICC(2,1)= 0.73 (0.47, 0.87)	NA	NR	Fair to good[F]	6/11
Dynamic Maximum passive shoulder internal	Juul-Kristensen, et al. (2011)	inclinometer	sitting	0.58	33	33	1	2	ICC(2,1)= 0.76 (0.56, 0.87)	NA	NR	Excellent[F]	6/11

rotation													
Dynamic Travelling distance	Juul-Kristensen, et al. (2011)	tape measure	standing	0.58	33	33	1	2	ICC(2,1)= 0.64 (0.39, 0.81)	NA	NR	Fair to good[F]	6/11
Dynamic Upper arm angle at initial scapula movement	Juul-Kristensen, et al. (2011)	inclinometer	sitting	0.58	33	33	1	2	ICC(2,1)= 0.74 (0.53, 0.86)	NA	NR	Fair to good[F]	6/11
Dynamic Weakness of the scapula stabilising muscles	Juul-Kristensen, et al. (2011)	NA	prone	0.58	33	33	1	2	ICC(2,1)= 0.42 (0.10, 0.66)	NA	NR	Fair to good[F]	6/11
Lateral Scapular Slide Test at 45° abduction and internal rotation	Odom, et al. (2001)	tape measure	standing	1.00	20	40	6	2	ICC(1,1)= 0.66 (0.36, 0.82), within-day reliability[#]	NA	SEM for ICC(1,1)= 0.57	Fair to good[F]	7/11
	Odom, et al. (2001)	tape measure	standing	0.00	26	54	6	2	ICC(1,1)= 0.77 (0.60, 0.86), within-day reliability[#]	NA	SEM for ICC(1,1)= 0.58	Excellent[F]	7/11
	Shadmehr, et al. (2010)	scoliometer	standing	1.00	27	54	3	2	A: ICC(2,1)= 0.91 (NR)[ICC]; Non-A: ICC(2,1)= 0.96 (NR)[ICC]	NA	A: 2 SEM for ICC(2,1)= 6.4 (0.83, 0.96); Non-A: 2 SEM for ICC(2,1)= 4.6 (0.84, 0.96)	Excellent[F]	5/11
	Shadmehr, et al. (2010)	scoliometer	standing	0.00	30	60	3	2	L: ICC(2,1)= 0.94 (NR)[ICC]; R: ICC(2,1)= 0.95 (NR)[ICC]	NA	L: 2 SEM for ICC(2,1)= 7.6 (0.88, 0.97); R: 2 SEM for ICC(2,1)= 6.8 (0.90, 0.97)	Excellent[F]	5/11
Lateral Scapular Slide Test at 90° abduction and maximal internal rotation	Odom, et al. (2001)	tape measure	standing	1.00	20	40	6	2	ICC(1,1)= 0.62 (0.27, 0.79), within-day reliability[#]	NA	SEM for ICC(1,1)= 0.86	Fair to good[F]	7/11
	Odom, et al. (2001)	tape measure	standing	0.00	26	54	6	2	ICC(1,1)= 0.80 (0.65, 0.88), within-day reliability[#]	NA	SEM for ICC(1,1)= 0.80	Excellent[F]	7/11
	Shadmehr, et al. (2010)	scoliometer	standing	1.00	27	54	3	2	A: ICC(2,1)= 0.94 (NR)[ICC]; Non-A: ICC(2,1)= 0.90 (NR)[ICC]	NA	A: 2 SEM for ICC(2,1)= 5.6 (0.87, 0.97); Non-A: 2 SEM for ICC(2,1)= 6.5 (0.80, 0.95)	Excellent[F]	5/11

	Shadmehr, et al. (2010)	scoliometer	standing	0.00	30	60	3	2	L: ICC(2,1)= 0.97 (NR)[ICC]; R: ICC(2,1)= 0.95 (NR)[ICC]	NA	L: 2 SEM for ICC(2,1)= 5.6 (0.93, 0.98); R: 2 SEM for ICC(2,1)= 7.3 (0.90, 0.97)	Excellent[F]	5/11
Scapula upward rotation: total shoulder abduction at 45°	Watson, et al. (2005)	inclinometer	standing	1.00	26	26	1	2	ICC(NR)= 0.88 (0.79, 0.97), within-day reliability	NA	SEM for ICC(NA)= 4.0	Excellent[F]	4/11
Scapula upward rotation: total shoulder abduction at 90°	Watson, et al. (2005)	inclinometer	standing	1.00	26	26	1	2	ICC(NR)= 0.90 (0.82, 0.97), within-day reliability	NA	SEM for ICC(NA)= 3.8	Excellent[F]	4/11
Scapula upward rotation: total shoulder abduction at 135°	Watson, et al. (2005)	inclinometer	standing	1.00	26	26	1	2	ICC(NR)= 0.81 (0.67, 0.94), within-day reliability	NA	SEM for ICC(NA)= 5.2	Excellent[F]	4/11
Scapula upward rotation: total shoulder abduction at end of range	Watson, et al. (2005)	inclinometer	standing	1.00	26	26	1	2	ICC(NR)= 0.94 (0.89, 0.98), within-day reliability	NA	SEM for ICC(NA)= 2.9	Excellent[F]	4/11

[*] = Data for CI obtained from author of primary studies; [$]= Data for ICC and CI obtained from author of primary studies; [#] = Study authors provided time between tests; [ICC] = ICC form self named; [L] = Landis and Koch (1977); [F] = Fleiss (1999)

Abbreviations: 2 SEM = Double standard error of the measurement; A = Affected side; D = Dominant shoulder; ICC = Intraclass correlation coefficient; k =

Kappa; L = Left side; MDC = Minimal detectable change; NA = Not applicable; Non-A = nonaffected side; Non-D = Non dominant shoulder; NR = Not reported; R = Right side; SEM = Standard error of the measurement

Table 4: Results of interrater reliability

Indextest	Study	Instrument	Test position	Prevalence	$N_{subject}$	$N_{shoulder}$	N_{rater}	Estimates of reliability (95 % CI)	Po (%)	Measures of uncertainty	Strength of agreement	QAREL Score
Observation of scapular positioning / motion (dynamic)												
Clavicular movement during the first 60 degrees of flexion	Baertschi, et al. (2013)	NA	standing	1.00	19	19	2	k= 0.52 (0.12, 0.92); PABAK= 0.57 (0.21, 0.95)	79	BI= 0.10; PI= 0.36	Moderate[L]	7/11
	Baertschi, et al. (2013)	NA	standing	0.49	39	39	2	k= 0.63 (0.39, 0.87); PABAK= 0.64 (0.40, 0.88)	82	BI= 0.12; PI= 0.20	Substantial[L]	7/11
	Baertschi, et al. (2013)	NA	standing	0.00	20	20	2	k= 0.35 (0.00, 0.86); PABAK= 0.70 (0.39, 1.00)	85	BI= 0.15; PI= 0.85	Fair[L]	7/11
Movement of the cervicothoracic junction during the last phase of flexion	Baertschi, et al. (2013)	NA	standing	1.00	19	19	2	k= 0.64 (0.00, 1.00); PABAK= 0.90 (0.70, 1.00)	95	BI= 0.05; PI= 0.84	Substantial[L]	7/11
	Baertschi, et al. (2013)	NA	standing	0.49	39	39	2	k= 0.84 (0.67, 1.00); PABAK= 0.85 (0.68, 1.00)	92	BI= 0.02; PI= 0.15	Almost Perfect[L]	7/11
	Baertschi, et al. (2013)	NA	standing	0.00	20	20	2	k= 0.74 (0.40, 1.00); PABAK= 0.80 (0.54, 1.00)	90	BI= 0.10; PI= 0.50	Substantial[L]	7/11
Observational classification 2-type method	Uhl, et al. (2009)	NA	standing	0.36	56	56	2	k= 0.41 (NR)	79	NR	Moderate[L]	5/11
Observational classification 3-type method	P. McClure, et al. (2009)	NA	standing		142	142	2	L: weighted k= 0.58 (0.38, 0.79); R: weighted k= 0.55 (0.32, 0.78)	L: 81; R:	L: kmax= 0.75; R: kmax= 0.93	Moderate[L]	4/11

									80			
	P. McClure, et al. (2009)	NA	standing		90	90	6	L: weighted k= 0.48 (0.29, 0.67); R: weighted k= 0.61 (0.43, 0.78)	L: 75; R: 82	L: kmax= 0.79; R: kmax= 0.79	Moderate[L]	5/11
Observational classification 4-type method	Kibler, et al. (2002)	NA	standing	0.23	26	26	4	Ratergroup A: k= 0.31 (NR); Ratergroup B: k= 0.42 (NR)	NR	NR	Fair[L]	5/11
	Uhl, et al. (2009)	NA	standing	0.36	56	56	2	k= 0.44 (NR)	61	NR	Moderate[L]	5/11
Scapular axillary hair test at the end of flexion	Baertschi, et al. (2013)	NA	standing	1.00	19	19	2	k= 0.60 (0.11, 1.00); PABAK= 0.79 (0.51, 1.00)	89	BI= 0.00; PI= 0.68	Moderate[L]	7/11
	Baertschi, et al. (2013)	NA	standing	0.49	39	39	2	k= 0.79 (0.60, 0.98); PABAK= 0.79 (0.60, 0.99)	90	BI= 0.05; PI= 0.07	Substantial[L]	7/11
	Baertschi, et al. (2013)	NA	standing	0.00	20	20	2	k= 0.74 (0.40, 1.00); PABAK= 0.80 (0.54, 1.00)	90	BI= 0.10; PI= 0.50	Substantial[L]	7/11
Scapular posterior tilting during the last phase of flexion	Baertschi, et al. (2013)	NA	standing	1.00	19	19	2	k= 0.60 (0.11, 1.00); PABAK= 0.79 (0.51, 1.00)	89	BI= 0.00; PI= 0.68	Moderate[L]	7/11
	Baertschi, et al. (2013)	NA	standing	0.49	39	39	2	k= 0.74 (0.53, 0.95); PABAK= 0.74 (0.53, 0.95)	87	BI= 0.02; PI= 0.05	Substantial[L]	7/11
	Baertschi, et al. (2013)	NA	standing	0.00	20	20	2	k= 0.57 (0.14, 1.00); PABAK= 0.70 (0.39, 1.00)	85	BI= 0.05; PI= 0.55	Moderate[L]	7/11
Measurement of scapular positioning (resting position)												
Lateral Scapular Slide Test at 0° abduction	Curtis and Roush (2006)	scoliometer	standing	1.00	18	18	3	ICC(2,3)= 0.96 (0.93, 0.98)[$]	NA	NR	Excellent[F]	7/11
	Curtis and Roush (2006)	scoliometer	standing	1.00	18	18	3	Rater A/B: ICC(2,2)= 0.94 (0.87, 0.97)[$]; Rater A/C: ICC(2,2)= 0.93 (0.84, 0.97)[$]; Rater B/C: ICC(2,2)= 0.97 (0.94, 0.99)[$]	NA	Rater A/B: SEM for ICC(2,2)= 4.60; Rater A/C: SEM for ICC(2,2)= 4.76; Rater B/C: SEM	Excellent[F]	7/11

										for ICC(2,2)= 3.00		
	Curtis and Roush (2006)	scoliometer	standing	0.00	15	15	3	ICC(2,3)= 0.94 (0.89, 0.97)[$]	NA	NR	Excellent[F]	7/11
	Curtis and Roush (2006)	scoliometer	standing	0.00	15	15	3	Rater A/B: ICC(2,2)= 0.95 (0.90, 0.97)[$]; Rater A/C: ICC(2,2)= 0.89 (0.78, 0.94)[$]; Rater B/C: ICC(2,2)= 0.88 (0.77, 0.94)[$]	NA	Rater A/B: SEM for ICC(2,2)= 5.21; Rater A/C: SEM for ICC(2,2)= 5.58; Rater B/C: SEM for ICC(2,2)= 4.80	Excellent[F]	7/11
	Nijs, et al. (2005)	tape measure	standing	1.00	29	29	2	L: ICC(2,2)= 0.82 (NR); R: ICC(2,2)= 0.96 (NR)	NA	NR	Excellent[F]	5/11
	Odom, et al. (2001)	tape measure	standing	1.00	20	40	6	ICC(1,1)= 0.79 (0.46, 0.91)	NA	SEM for ICC(1,1)= 0.79	Excellent[F]	7/11
	Odom, et al. (2001)	tape measure	standing	0.00	26	54	6	ICC(1,1)= 0.67 (0.25, 0.85)	NA	SEM for ICC(1,1)= 0.79	Fair to good[F]	7/11
	Shadmehr, et al. (2010)	scoliometer	standing	1.00	27	54	3	A: ICC(2,3)= 0.79 (NR)[ICC]; Non-A: ICC(2,3)= 0.75 (NR)[ICC]	NA	A: 2 SEM for ICC(2,3)= 9.3 (0.65, 0.89); Non-A: 2 SEM for ICC(2,3)= 10.2 (0.59, 0.86)	Excellent[F]	5/11
	Shadmehr, et al. (2010)	scoliometer	standing	0.00	30	60	3	L: ICC(2,3)= 0.92 (NR)[ICC]; R: ICC(2,3)= 0.92 (NR)[ICC]	NA	L: 2 SEM for ICC(2,3)= 10 (0.86, 0.96); R: 2 SEM for ICC(2,3)= 9.4 (0.86, 0.96)	Excellent[F]	5/11
Medial scapular border to T4 relaxed	Nijs, et al. (2005)	scoliometer	standing	1.00	29	29	2	L: ICC(2,2)= 0.79 (NR); R: ICC(2,2)= 0.50 (NR)	NA	NR	Fair to good[F]	5/11
Medial scapular border to T4 retraction	Nijs, et al. (2005)	scoliometer	standing	1.00	29	29	2	L: ICC(2,2)= 0.70 (NR); R: ICC(2,2)= 0.80 (NR)	NA	NR	Fair to good[F]	5/11
The pectoralis minor length test / posterior acromion border to table	Struyf, Meeus, et al. (2014)	protractor	supine	1.00	25	50	2	A: ICC (2,1)= 0.48 (0.10 - 0.73)[*], ICC(2,3)= 0.65 (0.18 - 0.85)[*]; Non-A: ICC (2,1)= 0.56 (0.22 -	NA	A: SEM for ICC= 0.46, MDC(95)= 1.29; Non-A: SEM for	Fair to good[F]	7/11

								0.78)*, ICC(2,3)= 0.72 (0.36 - 0.88)*		ICC= 0.61, MDC(95)= 1.68		
	Struyf, Meeus, et al. (2014)	protractor	supine	0.00	25	50	2	D: ICC (2,1)= 0.53 (0.15 - 0.76)*, ICC(2,3)= 0.67 (0.26 - 0.87)*; Non-D: ICC (2,1)= 0.47 (0.09 - 0.73)*, ICC(2,3)= 0.64 (0.16 - 0.84)*	NA	D: SEM for ICC= 0.38, MDC(95)= 1.07; Non-D: SEM for ICC= 0.45, MDC(95)= 1.24	Fair to good[F]	7/11
The pectoralis minor length test / posterior acromion border to table relaxed	Nijs, et al. (2005)	tape measure	supine	1.00	29	29	2	L: ICC(2,2)= 0.94 (NR); R: ICC(2,2)= 0.88 (NR)	NA	NR	Excellent[F]	5/11
The pectoralis minor length test / posterior acromion border to table retraction	Nijs, et al. (2005)	tape measure	supine	1.00	29	29	2	L: ICC(2,2)= 0.92 (NR); R: ICC(2,2)= 0.91 (NR)	NA	NR	Excellent[F]	5/11
Measurement of scapular positioning (non resting position)												
Lateral Scapular Slide Test at 45° abduction and internal rotation	Curtis and Roush (2006)	scoliometer	standing	1.00	18	18	3	ICC(2,3)= 0.93 (0.87, 0.96)$	NA	NR	Excellent[F]	7/11
	Curtis and Roush (2006)	scoliometer	standing	1.00	18	18	3	Rater A/B: ICC(2,2)= 0.87 (0.73, 0.94)$; Rater A/C: ICC(2,2)= 0.89 (0.77, 0.95)$; Rater B/C: ICC(2,2)= 0.93 (0.85, 0.97)$	NA	Rater A/B: SEM for ICC(2,2)= 5.38; Rater A/C: SEM for ICC(2,2)= 4.72; Rater B/C: SEM for ICC(2,2)= 4.18	Excellent[F]	7/11
	Curtis and Roush (2006)	scoliometer	standing	0.00	15	15	3	ICC(2,3)= 0.94 (0.89, 0.97)$	NA	NR	Excellent[F]	7/11
	Curtis and Roush (2006)	scoliometer	standing	0.00	15	15	3	Rater A/B: ICC(2,2)= 0.80 (0.57, 0.90)$; Rater A/C: ICC(2,2)= 0.84 (0.66, 0.92)$; Rater B/C: ICC(2,2)= 0.76 (0.49, 0.88)$	NA	Rater A/B: SEM for ICC(2,2)= 6.37; Rater A/C: SEM for ICC(2,2)= 7.16; Rater B/C: SEM for ICC(2,2)=	Excellent[F]	7/11

										4.38		
	Nijs, et al. (2005)	tape measure	standing	1.00	29	29	2	L: ICC(2,2)= 0.85 (NR); R: ICC(2,2)= 0.95 (NR)	NA	NR	Excellent[F]	5/11
	Odom, et al. (2001)	tape measure	standing	1.00	20	40	6	ICC(1,1)= 0.45 (-0.38, 0.78)	NA	SEM for ICC(1,1)= 0.79	Fair to good[F]	7/11
	Odom, et al. (2001)	tape measure	standing	0.00	26	54	6	ICC(1,1)= 0.43 (-0.29, 0.75)	NA	SEM for ICC(1,1)= 1.08	Fair to good[F]	7/11
	Shadmehr, et al. (2010)	scoliometer	standing	1.00	27	54	3	A: ICC(2,3)= 0.70 (NR)[ICC]; Non-A: ICC(2,3)= 0.82 (NR)[ICC]	NA	A: 2 SEM for ICC(2,3)= 11 (0.53, 0.84); Non-A: 2 SEM for ICC(2,3)= 8.8 (0.69, 0.90)	Fair to good[F]	5/11
	Shadmehr, et al. (2010)	scoliometer	standing	0.00	30	60	3	L: ICC(2,3)= 0.93 (NR)[ICC]; R: ICC(2,3)= 0.93 (NR) [ICC]	NA	L: 2 SEM for ICC(2,3)= 8.3 (0.87, 0.96); R: 2 SEM for ICC(2,3)= 8 (0.89, 0.96)	Excellent[F]	5/11
Lateral Scapular Slide Test at 90° abduction and maximal internal rotation	Curtis and Roush (2006)	scoliometer	standing	1.00	18	18	3	ICC(2,3)= 0.86 (0.74, 0.93)[$]	NA	NR	Excellent[F]	7/11
	Curtis and Roush (2006)	scoliometer	standing	1.00	18	18	3	Rater A/B: ICC(2,2)= 0.79 (0.60, 0.90)[$]; Rater A/C: ICC(2,2)= 0.78 (0.57, 0.89)[$]; Rater B/C: ICC(2,2)= 0.89 (0.78, 0.94)[$]	NA	Rater A/B: SEM for ICC(2,2)= 6.86; Rater A/C: SEM for ICC(2,2)= 6.36; Rater B/C: SEM for ICC(2,2)= 7.20	Fair to good[F]	7/11
	Curtis and Roush (2006)	scoliometer	standing	0.00	15	15	3	Rater A/B: ICC(2,2)= 0.79 (0.60, 0.90)[$]; Rater A/C: ICC(2,2)= 0.78 (0.57, 0.89)[$]; Rater B/C: ICC(2,2)= 0.89 (0.78, 0.94)[$]	NA	Rater A/B: SEM for ICC(2,2)= 7.54; Rater A/C: SEM for ICC(2,2)= 8.26; Rater B/C: SEM for ICC(2,2)=	Fair to good[F]	7/11

										6.22		
	Curtis and Roush (2006)	scoliometer	standing	0.00	15	15	3	ICC(2,3)= 0.88 (0.78, 0.93)[$]	NA	NR	Excellent[F]	7/11
	Nijs, et al. (2005)	tape measure	standing	1.00	28	28	2	L: ICC(2,2)= 0.70 (NR); R: ICC(2,2)= 0.85 (NR)	NA	NR	Fair to good[F]	5/11
	Odom, et al. (2001)	tape measure	standing	1.00	20	40	6	ICC(1,1)= 0.57 (-0.23, 0.85)	NA	SEM for ICC(1,1)= 1.10	Fair to good[F]	7/11
	Odom, et al. (2001)	tape measure	standing	0.00	26	54	6	ICC(1,1)= 0.74 (0.41, 0.88)	NA	SEM for ICC(1,1)= 1.20	Fair to good[F]	7/11
	Shadmehr, et al. (2010)	scoliometer	standing	1.00	27	54	3	A: ICC(2,3)= 0.63 (NR)[ICC]; Non-A: ICC(2,3)= 0.95 (NR)[ICC]	NA	A: 2 SEM for ICC(2,3)= 13.4 (0.42, 0.79); Non-A: 2 SEM for ICC(2,3)= 6.8 (0.76, 0.93)	Excellent[F]	5/11
	Shadmehr, et al. (2010)	scoliometer	standing	0.00	30	60	3	L: ICC(2,3)= 0.94 (NR)[ICC]; R: ICC(2,3)= 0.95 (NR)[ICC]	NA	L: 2 SEM for ICC(2,3)= 7.9 (0.90, 0.97); R: 2 SEM for ICC(2,3)= 7.3 (0.91, 0.97)	Excellent[F]	5/11
Scapulohumeral abduction	Terwee, et al. (2005)	NA	sitting	1.00	181	362	2	A: ICC(2,2)= 0.67 (0.35, 0.81)[ICC]; Non-A: ICC(2,2)= 0.15 (0.02, 0.29)[ICC]	NA	NR	Poor[F]	6/11
Symptom alteration tests												
Modified Scapular Assistance Test sagittal plane	Rabin, et al. (2006)	NA	standing	1.00	43[+]	43[+]	2	k= 0.62 (NR)	91	NR	Substantial[L]	5/11
Modified Scapular Assistance Test scapular plane	Rabin, et al. (2006)	NA	standing	1.00	43[+]	43[+]	2	k= 0.53 (NR)	77	NR	Moderate[L]	5/11

[*] = Data for CI obtained from author of primary studies; [$]= Data for ICC and CI obtained from author of primary studies; [+] = Incomplete subjects; [ICC] = ICC form self named; [L] = Landis and Koch (1977); [F] = Fleiss (1999)

Abbreviations: 2 SEM = Double standard error of the measurement; A = Affected side; BI = Bias index; D = Dominant shoulder; ICC = Intraclass correlation coefficient; k = Kappa; kmax = Maximum Kappa; L = Left side; MDC = Minimal detectable change; NA = Not applicable; Non-A = nonaffected side; Non-D = Non dominant shoulder; NR = Not reported; PAPAK = Prevalence - adjusted bias - adjusted Kappa; PI = Prevalence index; R = Right side; SEM = Standard error of the measurement