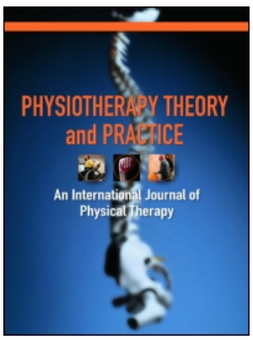


Taylor & Francis
Taylor & Francis Group

Physiotherapy Theory and Practice
An International Journal of Physical Therapy

ISSN: 0959-3985 (Print) 1532-5040 (Online) Journal homepage: https://www.tandfonline.com/loi/iptp20

Interrater agreement and reliability of clinical tests for assessment of patients with shoulder pain in primary care

Adri T Apeldoorn, Marjolein C Den Arend, Ruud Schuitemaker, Dick Egmond, Karin Hekman, Tjeerd Van Der Ploeg, Steven J Kamper, Maurits W Van Tulder & Raymond W Ostelo

To cite this article: Adri T Apeldoorn, Marjolein C Den Arend, Ruud Schuitemaker, Dick Egmond, Karin Hekman, Tjeerd Van Der Ploeg, Steven J Kamper, Maurits W Van Tulder & Raymond W Ostelo (2019): Interrater agreement and reliability of clinical tests for assessment of patients with shoulder pain in primary care, Physiotherapy Theory and Practice, DOI: 10.1080/09593985.2019.1587801

To link to this article: https://doi.org/10.1080/09593985.2019.1587801

Published online: 22 Mar 2019.

PHYSIOTHERAPY THEORY AND PRACTICE
https://doi.org/10.1080/09593985.2019.1587801

Taylor & Francis
Taylor & Francis Group

Interrater agreement and reliability of clinical tests for assessment of patients with shoulder pain in primary care

Adri T Apeldoorn[a,b], Marjolein C Den Arend[c], Ruud Schuitemaker[d], Dick Egmond[e], Karin Hekman[f], Tjeerd Van Der Ploeg[g], Steven J Kamper[h], Maurits W Van Tulder[i], and Raymond W Ostelo[i,j]

[a]Rehabilitation Department, Noordwest Ziekenhuisgroep, Alkmaar, Netherlands; [b]Breederode Hogeschool, Rotterdam, Netherlands; [c]Arcus Group Physiotherapy, Hoeven, Netherlands; [d]Schuitemaker and van Schaik Physiotherapy and Manual Therapy, Amsterdam, Netherlands; [e]Institute for Applied Manual Therapy, Wolfsburg, Germany; [f]IBC Amstelland, Amstelveen, Netherlands; [g]Hogeschool Inholland, Alkmaar, Netherlands; [h]School of Public Health, University of Sydney, Sydney, Australia; [i]Department of Health Sciences, Faculty of Science, Vrije Univerteit Amsterdam, Amsterdam, Netherlands; [j]Department of Epidemiology and Biostatistics (Amsterdam UMC), location VUmc & Amsterdam Movement Sciences

ABSTRACT

Background: There is limited information about the agreement and reliability of clinical shoulder tests. **Objectives**: To assess the interrater agreement and reliability of clinical shoulder tests in patients with shoulder pain treated in primary care. **Methods**: Patients with a primary report of shoulder pain underwent a set of 21 clinical shoulder tests twice on the same day, by pairs of independent physical therapists. The outcome parameters were observed and specific interrater agreement for positive and negative scores, and interrater reliability (Cohen's kappa (κ)). Positive and negative interrater agreement values of ≥0.75 were regarded as sufficient for clinical use. For Cohen's κ, the following classification was used: <0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, 0.81–1.00 very good reliability. Participating clinics were randomized in two groups; with or without a brief practical session on how to conduct the tests. **Results**: A total of 113 patients were assessed in 12 physical therapy practices by 36 physical therapists. Positive and negative interrater agreement values were both sufficient for 1 test (the Full Can Test), neither sufficient for 5 tests, and only sufficient for either positive or negative agreement for 15 tests. Interrater reliability was fair for 11 tests, moderate for 9 tests, and good for 1 test (the Full Can Test). An additional brief practical session did not result in better agreement or reliability. **Conclusion**: Clinicians should be aware that interrater agreement and reliability for most shoulder tests is questionable and their value in clinical practice limited.

ARTICLE HISTORY
Received 9 June 2018
Revised 3 December 2018
Accepted 18 January 2019

KEYWORDS
Physical therapy; diagnostic tests; accuracy

Introduction

Shoulder pain is one of the three most common musculoskeletal disorders (Picavet and Schouten, 2003; Pope, Croft, Pritchard, and Silman, 1997), and in the top three health problems of patients treated by Dutch physical therapists (Barten and Koppes, 2016). The incidence of shoulder pain in the Netherlands has been estimated to be 19 per 1000 registered patients per year (Greving et al., 2012). About 13% of persons with shoulder pain who visit a general practitioner are referred for physical therapy (Kooijman et al., 2013). Physical therapists also treat patients without a referral (direct access). In 2015, 51% of patients visited a physical therapist without a referral (Barten and Koppes, 2016). Shoulder pain is often recurrent and frequently persists long term. Karel et al. (2016) showed a recovery rate of 60% for the total population and 65% for the working population treated with physical therapy after 26 weeks.

The assessment of the shoulder can be challenging as sensitivity and specificity of symptom provocation tests are often insufficient to rule-in, or rule-out responsibility of the structure for patient symptoms. The relation between signs and symptoms, and impaired motor control or structural failure observed on imaging or intraoperatively is poor. For example, Tempelhof, Rupp, and Seil (1999) found a high rate of rotator cuff tears with increasing age in patients with asymptomatic shoulders. Due to these limitations, several researchers advise to use a comprehensive clinical examination including history and a combination of shoulder tests (Cools, Cambier, and Witvrouw, 2008; Hegedus et al., 2015; Michener, Walsworth, Doukas, and Murphy, 2009). Clusters of tests and classification

CONTACT Adri T Apeldoorn a.t.apeldoorn@nwz.nl Rehabilitation Department, Noordwest Ziekenhuisgroep, Wilhelminalaan 12, Alkmaar 1815 JD, Netherlands

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/iptp.
This article has been republished with minor changes. These changes do not impact the academic content of the article.

algorithms may be more beneficial for diagnosis in clinical practice compared to the use of single tests (Hegedus et al., 2015).

One of the classification systems that have been proposed is the classification algorithm of Cools, Cambier, and Witvrouw (2008). This algorithm uses not only generally accepted symptom-provoking shoulder tests like the Neer Test, Hawkins-Kennedy Test and the Empty Can Test, but also relatively new symptom altering tests based on movement dysfunction like the modified Scapular Assistance Test (mSAT) and the Scapular Retraction Test (SRT). While several authors have provided valuable analyses of diagnostic accuracy and validity of clinical shoulder tests, agreement and reliability of these single tests are also of concern. A recent systematic review of the intra and interrater reliability of clinical shoulder tests concluded that only a few tests have acceptable levels of reliability (Lange et al., 2017). The authors stressed that their findings may be inaccurate and need to be interpreted with caution due to the heterogeneity among studies, a lack of high quality studies, and the use of variable estimates of reliability. They concluded that agreement and reliability studies using appropriate methodology and statistical analysis are needed (Lange et al., 2017).

It can be argued that for the clinicians the interrater agreement and reliability is not very important, as they generally establish diagnosis and treatment strategy by combining findings of the patient history and clinical assessment procedures. However, sufficient interrater agreement and reliability are essential requirements as poor interrater agreement and reliability leads to different results on the same tests, different clusters, and different treatment plans. In the absence of adequate reliability there is little value in including a test in a clinical assessment schedule.

Several designs are available for the assessment of interrater agreement and reliability (e.g., independent and consecutive assessments, simultaneous assessment procedures with one rater and one observer, videotaped recordings, and vignette studies). Each design assesses a different aspect of agreement and reliability. For example, in vignette studies, ratings are based on the same signs and symptoms but do not account for patients that present differently in separate assessments and testing procedures. Independent and consecutive assessments include variation in patient presentations and testing procedures, but the downside is that the first assessment might have an influence on the second assessment. Nevertheless, we contend that this latter method is an appropriate design for the assessment of interrater agreement and reliability, because in clinical practice the label a patient receives is based on these tests and should not depend on what the therapist he/she sees or does not see. So in order to test if two therapists would arrive at the same conclusion, this design is the most appropriate. The aim of this study was to assess interrater agreement and reliability of physical shoulder tests by independent and consecutive assessments twice on the same day, by pairs of independent physical therapists. The secondary aim was to evaluate whether agreement and reliability differs between physical therapists who did or did not receive an additional brief practical training session.

Methods

The study was approved by the EMGO^{+} Science Committee on July 1, 2014, number WC2014-043 and the Medical Ethical Committee of the VU University Medical Center in Amsterdam approved the study on October 19, 2015, registration number 2014.482 (NL47668.029.14) The trial was registered in the Dutch trial register, number NTR5905. The presentation of the results follows the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (Kottner et al., 2011). This study design is a single-group test–retest design.

Subjects

Patients with a primary report of shoulder pain were recruited by participating physical therapy clinics. Inclusion criteria were: shoulder (girdle) pain, with or without radiation into the arm, age over 17, and adequate command of the Dutch language. Exclusion criteria were: recent (< 3 months) surgery or shoulder fracture, shoulder pain as a result of possible cervical nerve root entrapment (e.g., positive Spurling and/or traction test) (Bertilson, Grunnesjo, and Strender, 2003; Tong, Haig, and Yamakawa, 2002), possible total cuff tear (i.e., positive lag signs like the external rotation lag sign and the Hornblower's sign) (Jain et al., 2017), serious diseases (e.g., malignancies), rheumatic and/or neurological disorders (e.g., CVA, MS, Parkinson's disease), organ pathologies which affect the shoulder pain, dementia, and/or psychological disorders. Patients with partial cuff ruptures or calcifications as established by MRI or ultrasound were not excluded.

Design

Participating physical therapists informed all eligible patients who attended their primary care clinic for their shoulder pain about the study. Eligible patients

were further instructed about the study through a participant information letter. Patients were given at least 2 days' time to consider whether to participate in study or not. The first rater collected informed consent of the patient, and checked the inclusion and exclusion criteria. Information about the patient was collected regarding demographic characteristics (i.e., age and gender) and duration of symptoms, previous history of shoulder pain, pain intensity, functional status, education, employment and psychological status by means of a questionnaire. Disability was measured with the Dutch version of the Shoulder Pain and Disability Index (SPADI-D) (Roach, Budiman-Mak, Songsiridej, and Lertratanakul, 1991), which is a reliable and valid measure of shoulder disability in primary care (Thoomes-de Graaf et al., 2015).

Patients were assigned to two physical therapists successively. Both physical therapists performed each individual test according to the same procedure, and they followed a particular order that was established a priori. The second physical therapist was blinded to the results of the first physical therapist and performed the tests within 30 minutes of the first assessment. The patient was asked not to communicate any details or outcome of the first assessment with the second rater.

Patients could have been referred by their general practitioner or medical specialist or visited the clinic without referral (direct access). Patients could be assessed at the first visit at the clinic, or at a subsequent visit. In case of the first visit, both raters had no or minor information from the referring general practitioner about the clinical diagnosis or patient history. Examples of minor information are "shoulder problems," "it might be impingement," "patient does not want an injection." In case the patient was included after the first visit, the rater who had seen the patient before had information about the patient history and clinical diagnosis. All tests were scored positive, negative or not applicable on a standardized data form. Tests were regarded not applicable if the test could not be performed (i.e., the patient could not accomplish the shoulder test position due to a restricted range of motion or pain), or the test was not informative (e.g., the mSAT is not applicable if the patient does not report symptoms with active elevation).

Provocation tests were applied with a minimal amount of stress. When a patient reported (increase of) pain the test was ended immediately. The written evaluations were placed in envelopes, and the raters remained blind to the other's classification decision during the study. Figure 1 shows a detailed description of the study design.

Before each examination, current shoulder pain was assessed on a numerical rating scale (NRS: 0–10) to check stability of the patient's pain between examinations. Unstable patients were a priori defined as those having > 2 points change.

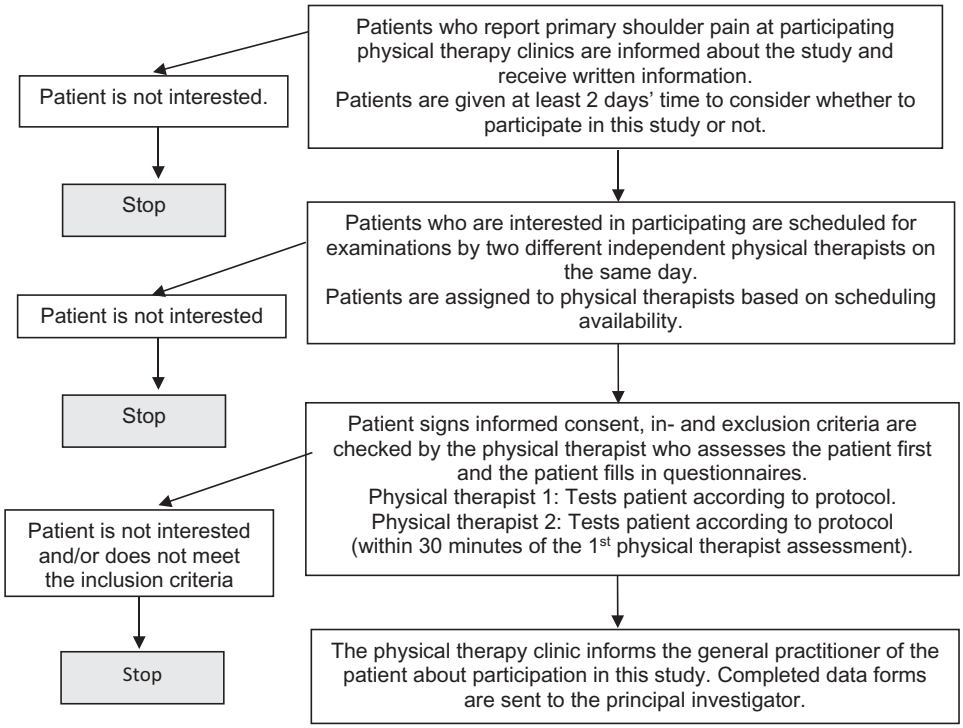


Figure 1. Study design.

Therapists and clinics

Private physical therapy clinics of the Shoulder Network Amsterdam and of the personal network of the principal investigator (AA) were invited to participate. Physical therapy clinics could participate if at least two physical therapists of that clinic were interested in participating in the study.

Prior to recruitment of study patients, three patients were pilot-tested by participating physical therapists to standardize techniques and interpretations. Due to limited resources for this study physical therapists were not asked to record the number of patients who refused to participate or reasons for not participating.

Physical therapy clinics received instructions by the principal investigator consisting of leaflets and access to a video demonstrating how all tests are performed. The leaflet and the video detailed the performance and scoring of the tests (positive or negative). Half of the clinics received an additional brief face-to-face practical session of approximately 45 minutes, delivered by the principal investigator in order to assess if this brief practical session had an influence on results.

Randomization

Participating physical therapy clinics were randomized in two groups regarding the level of instructions they received prior to the study. Clinics were randomized using a computer-generated randomization list. To conceal the allocation, randomization was conducted by an independent person unaware of the purpose of the study.

Tests

The selection of shoulder tests was mainly based on the classification algorithm of Cools, Cambier, and Witvrouw (2008). This algorithm uses not only generally accepted symptom-provoking shoulder tests like the Neer Test, Hawkins-Kennedy Test, and the Empty Can Test, but also relatively new symptom altering tests based on movement dysfunction like the mSAT and the SRT. There is very little or no information about the agreement or reliability of these latter tests available (Lewis, 2009).

Prior to this study, the principal investigator discussed the design of the study and the operational definitions of the clinical tests in detail with two core members of the Shoulder Network Amsterdam (coauthors RS and KH) and coauthor DE. After consultation the Shoulder Network Amsterdam (a part of the Dutch shoulder Network), four extra tests were added; the Scapula Position (visual observation of the scapula in neutral standing), Internal Rotation Resistance Strength Test, Impingement Relief Test (IRT), and the Combined Reduction Test (CRT). The CRT combines elements of the mSAT, SRT, and the IRT. The Shoulder Network Amsterdam promotes the use of symptom reduction tests. If reduction of symptoms is found with a reduction test, the same technique can be used to treat symptoms.

The test procedures recommended in a Dutch standard textbook for the clinical examination and treatment of extremities (Egmond and Schuitemaker, 2014) slightly differ from the instructions of Cools, Cambier, and Witvrouw (2008). Therefore, RS and DE produced an instruction video of the performance of all 21 shoulder tests exclusively for the present study according to the Dutch textbook. Operational definitions of the selected tests of the present study are provided in Appendix 1.

Table 1 presents a summary of the interrater agreement and reliability findings of the 21 selected clinical shoulder tests. The results are based on an extensive search in Medline, from inception till 1 March 2018, and performed by the principal investigator (AA). Studies were included if they selected subjects with shoulder pain or a mix of subjects with and without shoulder pain. Studies that included asymptomatic subjects or subjects with shoulder pain due to neurological problems or neck or upper limb symptoms were excluded. Studies that based their results on the assessment on the affected and the nonaffected limb were also excluded. The search strategy was restricted to English, French, Dutch, and German-language papers. The reference list of identified articles was checked for additional studies. Also, three relevant systematic reviews on measurement properties of shoulder tests were checked (D'Hondt et al., 2017; Lange et al., 2017; May et al., 2010).

Sample-size estimation

Sample size estimations for reliability parameters are not a matter of statistical significance (de Vet, Terwee, Mokkink, and Knol, 2011; Tooth and Ottenbacher, 2004). Sample size however does affect the precision of the reliability estimates, and therefore de Vet, Terwee, Mokkink, and Knol (2011) advise to include at least 50 patients to fill a 2 × 2 table. Because physical therapists were randomized in two groups, the plan was to recruit at least 100 patients in total.

Data analyses

Descriptive statistics and frequency distributions of all baseline variables were assessed. As recommended by

Table 1. Overview of results of the literature search of interrater agreement and interrater reliability of clinical shoulder tests used in the present study.

Test	Study	Subjects, n	Observed agreement, %	Positive specific agreement	Negative specific agreement	Kappa coefficient (95%CI)
Scapula Position						
External Rotation Resistance Test	Nørregaard, Krogsgaard, Lorenzen, and Jensen, 2002	68	86			0.67
	Hayes and Petersen, 2003	18	61			0.37 (0.01-0.74) (grading included 4 classifications)
	Ostor et al, 2004	136	NR			0.45; 0.18; 0.38 (three pairs of examiners ab; bc; ac)
	Nanda et al, 2008	63	80			0.44
	Michener, Walsworth, Doukas, and Murphy, 2009	55	87			0.67 (0.40-0.94)
Empty Can Test (Jobe Test)	Ostor et al, 2004	136	NR			0.49; 0.44; 0.46 (three pairs of examiners ab; bc; ac)
	Holtby and Razmjou, 2004	152	NR			0.43
	Nanda et al, 2008	63	77			0.44
	Johansson and Ivarson, 2009	33	NR			0.94
	Michener, Walsworth, Doukas, and Murphy, 2009	55	76			0.47 (0.22-0.72)
	Vind et al, 2011	44	95			0.90 (0.76-1.00)
	Burns, Cleland, Carpenter, and Mintken, 2016	21	91 (pain) 67 (weakness)			0.69 (0.28-1) (pain) 0.35 (0-0.74) (weakness)
Full Can Test	Burns, Cleland, Carpenter, and Mintken, 2016	21	81			0.23 (0-0.65) (pain) 0.38 (0.01-0.76) (weakness)
Active Compression Test (O'Brien's Test)	Walsworth et al, 2008	55	64			0.24 (-0.02-0.50)
	Cadogan et al, 2011	40	88 70			0.22 (-0.24-0.68) (pain on 'top' of the schoulder) 0.38 (0.1-0.65) (pain 'inside' the shoulder)
	Burns, Cleland, Carpenter, and Mintken, 2016	21	81			0.57 (0.19-0.95)
Neer Test	Razmjou, Holtby, and Myhr, 2004	149	77	0.67	0.82	0.51 (0.36-0.65)
	Nanda et al, 2008	63	75			0.10
	Johansson and Ivarson, 2009	33	NR			1.0
	Michener, Walsworth, Doukas, and Murphy, 2009	55	71			0.40 (0.13-0.67)
	Vind et al, 2011	44	58			0.95 (0.86-1.00)
	Burns, Cleland, Carpenter, and Mintken, 2016	21	75			0.51 (0.13-0.88)
Hawkins-Kennedy Test	Nørregaard, Krogsgaard, Lorenzen, and Jensen, 2002	68	NR			0.07
	Razmjou, Holtby, and Myhr, 2004	150	60	0.57	0.63	0.29 (0.15-0.43)
	Ostor et al, 2004	136	NR			0.29; 0.18; 0.43 (three pairs of examiners ab; bc; ac)
	Nanda et al, 2008	63	95			0.55
	Johansson and Ivarson, 2009	33	NR			0.91
	Michener, Walsworth, Doukas, and Murphy, 2009	55	69			0.39 (0.12-0.65)
	Cadogan et al, 2011	40	68			0.38 (0.10-0.63)
	Vind et al, 2011	44	82			0.60 (0.34-0.85)
	Burns, Cleland, Carpenter, and Mintken, 2016	21	86			0.71 (0.41-1)
Kim Test	Kim, Park, Jeong, and Shin, 2005	172	NR			0.91 (unclear if this is the kappa coefficient or the percentage of agreement)
	Cadogan et al, 2011	40	85			-0.04 (-0.12-0.03)
Biceps Load II Test	Kim et al, 2001	127	NR			0.82
Internal Rotation Resistance Strength Test						
Load and Shift Test	Tzannes, Paxinos, Callanan, and Murrell, 2004	13	NR			0.42-0.72* (four examiners)
	Burns, Cleland, Carpenter, and Mintken, 2016	21	95-100			0.64 (0-1) right shoulder 1.0 (1.0-1.0) left shoulder
	Eshoj et al, 2018	40	95	0.50	0.97	0.48 (0.00-1.00)
Acromioclavicular Joint Stress Test						
Modified Scapular Assistance Test	Rabin, Irrgang, Fitzgerald, and Eubanks, 2006	46	77-91			0.53-0.62
	Kopkow, Lange, Schmitt, and Kasten, 2015	110	89	0.75	0.93	0.68 (0.50-0.85)
Scapular Retraction Test						

(*Continued*)

Table 1. (Continued).

Test	Study	Subjects, n	Observed agreement, %	Positive specific agreement	Negative specific agreement	Kappa coefficient (95%CI)
Impingement Relief Test	Nørregaard, Krogsgaard, Lorenzen, and Jensen, 2002	68	NR			0.34
Sulcus Sign Test	Nørregaard, Krogsgaard, Lorenzen, and Jensen, 2002	68	NR			0.13
	Tzannes, Paxinos, Callanan, and Murrell, 2004	13	NR			0.60* (four examiners)
	Eshoj et al, 2018	40	75	0.58	0.82	0.43 (0.17-0.72)
Apprehension Test	Vind et al, 2011	44	86			0.71 (0.59-0.98)
	Tzannes, Paxinos, Callanan, and Murrell, 2004	13	NR			0.31* (pain) 0.47* (apprehension) 0.44* (pain and/or apprehension) (four examiners)
	Eshoj et al, 2018	40	83	0.80	0.84	0.65 (0.38-0.85)
Relocation Test	Tzannes, Paxinos, Callanan, and Murrell, 2004	13	NR			0.31* (pain) 0.71* (apprehension) 0.44* (pain and/or apprehension) (four examiners)
	Eshoj et al, 2018	40	75	0.55	0.83	0.39 (0.07-0.68)
Release Test	Tzannes, Paxinos, Callanan, and Murrell, 2004	13	NR			0.31* (pain) 0.63* (apprehension) 0.45* (pain and/or apprehension) (four examiners)
	Eshoj et al, 2018	40	65	0.80	0.84	0.65 (0.38-0.85)
Combined Reduction Test						
Glenohumeral Internal Rotation Deficit Test						

Abbreviations: NR: not reported
*Intra Class Correlation

the GRRAS guidelines (Kottner et al., 2011) proportions of observed and specific interrater agreement and kappa statistics were used to determine interrater agreement and reliability. Observed agreement is the portion of cases for which the raters agree. Specific agreement quantifies the degree of agreement for positive and negative scores separately. Positive agreement (PA) is calculated in a 2 × 2 table by 2a/2a+b + c. Negative agreement (NA) by 2d/2d+b + c. Cell "a" contains the number of scores for which raters agree on positive scores, cells "b" and "c" contain the number of scores for which they disagree, and cell "d" the number for which raters agree on negative scores (Table 2). Specific agreement is a useful measure because it highlights where the difficulties lie (de Vet et al., 2013; Hripcsak and Heitjan, 2002). Interrater reliability is the observed proportion of agreement corrected for chance. A standard measure of interrater reliability is known as Cohen's κ.

The degree to which interrater agreement and reliability values are sufficient for clinical decision-making depends on the purpose and consequences of test results (Kottner et al., 2011). We decided that a positive and negative agreement value of ≥ 0.75 was relevant to consider in this setting. In other words, we assume that the interrater agreement is clinically relevant and acceptable if the probability that a physical therapist will agree with the opinion of a colleague physical therapist is ≥ 0.75. However, we realize that this cut-off value is based on arguments and lacks clear foundation. For Cohen's κ, the following classification was used: < 0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, 0.81–1.00 very good reliability (Altman, 1997).

Table 2. Two-by-two table to calculate specific agreement.

	Rater 2		
Rater 1	Positive	Negative	Total rater 2
Positive	a	b	a + b
Negative	c	d	c + d
Total rater 1	a + c	b + d	a + b + c + d

Interrater agreement and reliability values of shoulder tests were calculated separately for the two groups of physical therapists with different levels of instruction, that is, with or without a brief practical session on how to conduct the tests. Differences between characteristics of physical therapists who received a brief practical session or not were analyzed with appropriate methods (i.e., chi-square tests, Mann-Whitney tests, and unpaired *t*-tests). Also, differences between baseline characteristics of patients assessed by physical therapists who received a brief practical session or not were analyzed. In a further analysis, the interrater agreement and reliability of shoulder tests were calculated excluding patients who

were 'unstable' (> 2 points of change on an 11-point NRS) between examinations.

A *P* value of less than 0.05 (two-tailed) was considered as statistically significant. The data were analyzed using IBM SPSS Statistics Version 20.0 (IBM Corporation, Armonk, NY, USA), VasserStats (http://vassarstats.net/).

Results

From July 2016 until December 2016, 113 patients were assessed in 12 physical therapy practices by 36 physical therapists. Physical therapists' mean age was 38.3 years (standard deviation [SD] 11.1), and their mean duration of experience in physical therapy was 14.4 years (SD 14.4). Seventeen participating physical therapists (47.2%) had formal postgraduate manual therapy training. The median number of patients assessed by each practice was 10 (inter quartile range [IQR] 6–12). Patients were assessed during their first (74.8%), second (16.8%), third (2.8%), fifth (3.7%), 15th (0.9%), or 25th (0.9%) visit at their physical therapy clinic for their shoulder pain.

Of the 113 participating patients, there were no withdrawals during the assessment procedure. In all patients, the time between the two assessment procedures was not more than 30 minutes. The majority of patients were female (61.2%), had chronic shoulder pain (>12 weeks, 72.6%), and the mean SPADI-D was 45.2 (SD 21.4). Table 3 presents the clinical and demographic characteristics of the patients.

The proportion of missing values and not applicable scores for all clinical shoulder tests was 8.7% (411/4746). Most not applicable scores were found for the Relocation Test (46.0%, 104/226), Release Test (46.9%, 106/226), and the CRT (49.6%, 112/226).

For seven shoulder tests the positive agreement was regarded as sufficient (≥ 0.75); Empty Can Test, Full Can Test, Active Compression Test, Neer Test, mSAT, Relocation Test, and the Release Test. For 10 shoulder tests the negative agreement was regarded as sufficient (≥ 0.75): External Rotation Resistance Test, Full Can Test, Kim Test, Biceps Load II Test, Internal Rotation Resistance Strength Test, Load and Shift Test, Acromioclavicular Joint Stress Test, Sulcus Sign Test, CRT, and the Glenohumeral Internal Rotation Deficit Test. Only the Full Can Test scored sufficient on both positive and negative agreement (Table 4). A total of 11 tests scored fair, 9 moderate, and 1 good (Full Can Test) on interrater reliability (Table 4).

Randomization of the clinics resulted in 7 out the 12 practices receiving an additional brief practical session of 45 minutes. In the group without this additional training, 16 physical therapists assessed 73 patients, and in the group with additional training 20 physical therapists assessed 40 patients. Patients assessed by physical therapists in the group without additional training had a lower education level compared to the high-level instruction group (i.e., patients with a low, middle or high-level of education in percentages: 25, 60, 15 vs. 15, 35, 50, $P < 0.001$). The two groups of patients were not different on the other evaluated characteristics. Physical therapists

Table 3. Characteristics of patients (n = 113).

Characteristics	
Age, years, mean (SD) (range)	50.7 (14.4) (18–84)
Sex, *n* male (%)	45 (39.8)
Regard themselves as being Dutch, *n* (%)	110 (97.3)
Shoulder affected, *n* right (%)	53 (46.9)
Arm affected, *n* dominant (%)	63 (55.8)
Duration of current shoulder pain (weeks), median (IQR)	36.0 (12.0–136.0)
Acute (0–6 weeks), *n* (%)	16 (14.2)
Sub-acute (7–12 weeks), *n* (%)	15 (13.3)
Chronic (> 12 weeks), *n* (%)	82 (72.6)
Previous shoulder surgery, *n* (%)	8 (7.1)
Previous shoulder fracture, *n* (%)	1 (0.9)
Medical images (X-ray, CT, MRI, ultrasound), *n* (%)	38 (33.6)
Pain past week, Numerical Rating Scale, (0–10), mean (SD) (range)	5.4 (2.1) (1–9)
SPADI-D (0–100), mean (SD) (range)	45.2 (21.4) (4–94)
Currently taking pain medication, *n* (%)	11 (9.7)
Marital status	
With a partner, *n* (%)	80 (70.8)
Single, *n* (%)	33 (29.2)
Educational level	
Low, *n* (%)	24 (21.2)
Middle, *n* (%)	58 (51.3)
High, *n* (%)	31 (27.4)
Short-Form 36	
Physical Component Summary (0–100), mean (SD) (range)	43.6 (8.7) (18.8–64.7)
Mental Component Summary (0–100), mean (SD) (range)	52.0 (10.5) (14.8–67.6)

Abbreviations: SD: standard deviation; IQR: interquartile range; CT: computed tomography; MR: magnetic resonance imaging; SPADI-D: Shoulder Pain and Disability Index (Dutch version)

Table 4. Interrater agreement and interrater reliability (n = 113).

Test	Observed agreement, % (95%CI)	Positive specific agreement	Negative specific agreement	Cohen's kappa (95% CI)
Scapula Position	71.8 (62.3–79.8)	0.71	0.73	0.44 (0.27–0.60)
External Rotation Resistance Test	75.2 (66.1–82.7)	0.72	0.78*	0.50 (0.34–0.66)
Empty Can Test (Jobe test)	76.1 (67.0–83.4)	0.79*	0.72	0.51 (0.34–0.66)
Full Can Test	83.0 (74.5–89.2)	0.75*	0.87*	0.62 (0.47–0.78)
Active Compression Test (O'Brien's test)	74.1 (64.8–81.7)	0.79*	0.67	0.46 (0.29–0.63)
Neer Test	76.2 (66.9–83.6)	0.83*	0.59	0.43 (0.23–0.62)
Hawkins-Kennedy Test	67.9 (58.3–76.2)	0.74	0.59	0.33 (0.15–0.51)
Kim Test	70.6 (61.0–78.8)	0.56	0.78*	0.34 (0.14–0.53)
Biceps Load II Test	83.2 (74.5–89.5)	0.40	0.90*	0.31 (0.01–0.60)
Internal Rotation Resistance Strength Test (Zaslav Test)	76.8 (67.7–84.0)	0.64	0.83*	0.47 (0.30–0.65)
Load and Shift Test	85.0 (76.7–90.7)	0.48	0.91*	0.40 (0.13–0.66)
Acromioclavicular Joint Stress Test	76.1 (67.0–83.4)	0.64	0.82*	0.47 (0.29–0.64)
Modified Scapular Assistance Test	71.4 (61.3–79.9)	0.77*	0.62	0.39 (0.20–0.58)
Scapular Retraction Test	62.9 (52.4–72.3)	0.56	0.68	0.25 (0.06–0.45)
Impingement Relief Test	67.7 (57.3–76.7)	0.62	0.72	0.35 (0.16–0.54)
Sulcus Sign Test	84.8 (76.5–90.7)	0.45	0.91*	0.36 (0.09–0.64)
Apprehension Test	66.7 (57.0–75.2)	0.70	0.62	0.32 (0.14–0.50)
Relocation Test	67.4 (51.3–80.5)	0.76*	0.50	0.27 (0.00–0.58)
Release Test	75.6 (59.4–87.1)	0.81*	0.64	0.46 (0.17–0.75)
Combined Reduction Test	69.0 (52.8–81.9)	0.48	0.78*	0.26 (0.00–0.60)
Glenohumeral Internal Rotation Deficit Test	80.9 (72.1–87.5)	0.68	0.86*	0.54 (0.37–0.72)

Abbreviations: CI: confidence interval; *Sufficient agreement for positive agreement or negative agreement (≥ 0.75)

in the group without additional training had less experience in clinical practice compared to the group with additional training (median 8.0 years [IQR 4.0–12.0] vs. 16.0 years [IQR 10.3–28.8], P = .01), and were younger (mean 32.7 year [SD 7.9] vs. 42.5 year [SD 11.5], P = .01). The proportion of manual therapists was similar between the two groups. In general, the group that received an additional brief practical session did not score better on interrater agreement and reliability values than the group without an additional brief practical session (Table 5). In a second additional analysis, 21 patients that reported a change of pain of > 2 points between the two examinations were excluded. Of these 21 patients, 10 patients reported a reduction and 11 patients an increase of pain. In general, excluding these 21 patients did not result in higher agreement values. Positive agreement was regarded as sufficient for 7 tests, and negative agreement for 10 tests. Cohen's κ ranged from 0.24 to 0.66 (Table 6).

Discussion

The present study investigated the interrater agreement and reliability of 21 clinical shoulder tests commonly used in the assessment of patients with shoulder pain in primary care. The observed interrater agreement values ranged from 63% to 85% for all shoulder tests. The positive interrater agreement was sufficient for 7 shoulder tests, and the negative interrater agreement sufficient for 10 shoulder tests. Only one test (Full Can Test) scored sufficient positive and negative agreement values for clinical use. The Full Can Test was also the only test with a good interrater reliability. For the other 20 tests, the interrater reliability was fair (11 tests) or moderate (9 tests).

Previous interrater agreement and reliability studies of shoulder tests reported observed agreement values between 58 and 100% (Table 1). Although our agreement values were somewhat lower for the Kim Test, mSAT, and the Apprehension Test, in general, they were similar to previously reported values. It should be noted that comparison with other studies is hampered due to several reasons including no or a small number of studies available for each test, the variety of agreement and reliability parameters reported, and due differences of the populations studied. For example, Eshoj et al. (2018) reported sufficient interrater agreement and reliability values for the Apprehension Test and the Release Test, however their population sample was small (n = 40), and most of the subjects were normal shoulder individuals (68%). In the present study only patients with a primary report of shoulder pain were included. Besides, the values in Table 1 are not necessarily valid due to poor methodological quality of the majority of the previous studies (Lange et al., 2017). Improved comparability of studies in the future will be facilitated by concordance with minimal standards as outlined in the GRRAS (Kottner et al., 2011).

In general, agreement and reliability studies report observed agreement values and Cohen's κ values.

Table 5. Interrater agreement and interrater reliability for physical therapists that received high-level and low-level instruction.

	High-level instruction (20 physical therapists, 40 patients)				Low-level instruction (16 physical therapists, 73 patients)			
Test	Observed agreement % (95%CI)	Positive specific agreement	Negative specific agreement	Cohen's kappa (95%CI)	Observed agreement % (95%CI)	Positive specific agreement	Negative specific agreement	Cohen's kappa (95% CI)
Scapula Position	62.2 (44.8–77.1)	0.56	0.67	0.23 (−0.09–0.55)	76.7 (65.1–85.5)	0.77*	0.76*	0.54 (0.31–0.76)
External Rotation Resistance Test	77.5 (61.1–88.6)	0.71	0.82*	0.53 (0.22–0.84)	74.0 (62.2–83.2)	0.72	0.75*	0.48 (0.25–0.71)
Empty Can Test (Jobe test)	72.5 (55.9–84.9)	0.65	0.78*	0.42 (0.11–0.73)	78.1 (66.6–86.6)	0.84*	0.65	0.49 (0.26–0.72)
Full Can Test	77.5 (61.1–88.6)	0.61	0.84*	0.45 (0.15–0.76)	86.1 (75.5–92.8)	0.81*	0.89*	0.70 (0.47–0.93)
Active Compression Test (O'Brien's test)	74.4 (57.6–86.4)	0.67	0.79*	0.46 (0.15–0.77)	74.0 (62.2–83.2)	0.82*	0.54	0.36 (0.13–0.59)
Neer Test	87.5 (72.4–95.3)	0.91*	0.78*	0.70 (0.45–0.95)	70.0 (57.2–79.8)	0.78*	0.49	0.27 (0.04–0.51)
Hawkins-Kennedy Test	60.0 (43.4–74.7)	0.67	0.50	0.17 (−0.14–0.48)	72.2 (60.2–81.8)	0.77*	0.64	0.42 (0.18–0.65)
Kim Test	65.0 (48.3–78.9)	0.56	0.71	0.28 (−0.02–0.58)	73.9 (61.7–83.4)	0.55	0.82*	0.37 (0.13–0.60)
Biceps Load II Test	81.1 (64.3–91.4)	0.22	0.89*	0.13 (−0.17–0.43)	84.3 (73.2–91.5)	0.48	0.91*	0.39 (0.16–0.62)
Internal Rotation Resistance Strength Test (Zaslav Test)	72.5 (55.9–84.9)	0.48	0.81*	0.29 (−0.02–0.60)	79.2 (67.7–87.5)	0.71	0.84*	0.55 (0.33–0.77)
Load and Shift Test	72.5 (55.9–84.9)	0.27	0.83*	0.11 (−0.19–0.41)	91.8 (82.4–96.6)	0.67	0.95*	0.62 (0.40–0.84)
Acromioclavicular Joint Stress Test	82.5 (66.6–92.1)	0.59	0.89*	0.48 (0.18–0.78)	72.6 (60.7–82.1)	0.66	0.77*	0.43 (0.21–0.66)
Modified Scapular Assistance Test	45.2 (27.8–63.7)	0.51	0.37	−0.10 (−0.45–0.24)	83.6 (72.1–91.1)	0.87*	0.77*	0.64 (0.40–0.88)
Scapular Retraction Test	61.3 (42.3–77.6)	0.55	0.67	0.23 (−0.09–0.56)	63.6 (50.8–74.9)	0.57	0.68	0.26 (0.02–0.50)
Impingement Relief Test	70.0 (50.4–84.6)	0.67	0.73	0.39 (0.04–0.75)	66.7 (53.9–77.5)	0.59	0.72	0.33 (0.11–0.56)
Sulcus Sign Test	74.4 (57.6–86.4)	0.50	0.83*	0.33 (0.01–0.64)	90.4 (80.7–95.7)	0.36	0.95*	0.31 (0.08–0.54)
Apprehension Test	60.0 (43.4–74.7)	0.67	0.50	0.20 (−0.08–0.48)	70.4 (58.3–80.4)	0.73	0.68	0.41 (0.19–0.64)
Relocation Test	66.7 (38.7–87.0)	0.78*	0.29	0.12 (−0.31–0.55)	67.9 (47.6–83.4)	0.74	0.57	0.32 (−0.05–0.68)
Release Test	64.3 (35.6–86.0)	0.78*	<0.01	−0.21 (−0.72–0.30)	81.5 (61.3–93.0)	0.84*	0.78*	0.63 (0.26–0.99)
Combined Reduction Test	71.4 (51.1–86.1)	0.60	0.78*	0.38 (0.02–0.75)	64.3 (35.6–86.0)	<0.01	0.78*	−0.21 (−0.72–0.30)
Glenohumeral Internal Rotation Deficit Test	67.5 (50.8–80.9)	0.70	0.65	0.36 (0.06–0.66)	88.6 (78.2–94.6)	0.64	0.93*	0.57 (0.34–0.80)

Abbreviations: CI: confidence interval; *Sufficient agreement for positive agreement or negative agreement (≥ 0.75)

Table 6. Interrater agreement and interrater reliability for symptomatically stable patients between examinations (n = 92).

Test	Observed agreement % (95% CI)	Positive specific agreement	Negative specific agreement	Cohen's kappa (95%CI)
Scapula Position	73.0 (62.4–81.6)	0.72	0.74	0.47 (0.26–0.67)
External Rotation Resistance Test	77.2 (67.0–85.0)	0.73	0.80*	0.53 (0.33–0.74)
Empty Can Test (Jobe test)	72.8 (62.4–81.3)	0.75*	0.70	0.45 (0.25–0.66)
Full Can Test	84.6 (75.2–91.0)	0.78*	0.88*	0.66 (0.46–0.87)
Active Compression Test (O'Brien's test)	71.4 (60.9–80.2)	0.75*	0.66	0.41 (0.21–0.62)
Neer Test	76.1 (65.7–84.3)	0.83*	0.60	0.44 (0.23–0.64)
Hawkins-Kennedy Test	68.1 (57.4–77.3)	0.74	0.59	0.33 (0.13–0.54)
Kim Test	68.5 (57.7–77.7)	0.53	0.76*	0.30 (0.09–0.50)
Biceps Load II Test	81.6 (71.6–88.8)	0.27	0.89*	0.18 (−0.02–0.38)
Internal Rotation Resistance Strength Test (Zaslav Test)	75.8 (65.5–83.9)	0.61	0.82*	0.44 (0.24–0.64)
Load and Shift Test	84.8 (75.4–91.1)	0.46	0.91*	0.37 (0.17–0.58)
Acromioclavicular Joint Stress Test	76.1 (65.9–84.1)	0.61	0.83*	0.44 (0.24–0.64)
Modified Scapular Assistance Test	71.6 (60.3–80.8)	0.78*	0.61	0.39 (0.17–0.60)
Scapular Retraction Test	63.8 (52.2–74.0)	0.57	0.69	0.28 (0.08–0.48)
Impingement Relief Test	63.8 (52.2–74.0)	0.60	0.67	0.28 (0.07–0.49)
Sulcus Sign Test	82.4 (72.7–89.3)	0.38	0.90*	0.28 (0.08–0.49)
Apprehension Test	65.6 (54.7–75.1)	0.68	0.63	0.31 (0.10–0.51)
Relocation Test	68.8 (49.9–83.3)	0.78*	0.44	0.24 (−0.09–0.57)
Release Test	77.4 (58.5–89.7)	0.84*	0.63	0.47 (0.13–0.82)
Combined Reduction Test	68.6 (50.6–82.6)	0.52	0.77*	0.29 (−0.03–0.62)
Glenohumeral Internal Rotation Deficit Test	81.3 (71.5–88.4)	0.71	0.86*	0.57 (0.37–0.78)

Abbreviations: CI: confidence interval; *Sufficient agreement for positive agreement or negative agreement (≥ 0.75)

However, the GRASS also recommend to calculate specific agreement (Kottner et al., 2011). Despite these recommendations specific agreement values are rarely reported. In our literature review only one study reported specific agreement values (Kopkow, Lange, Schmitt, and Kasten, 2015), and for two studies we were able to calculate specific agreement values (Eshoj et al., 2018; Razmjou, Holtby, and Myhr, 2004) (Table 1).

The present study showed that an additional, brief face-to-face practical session did not improve interrater agreement and reliability. That is, agreement and reliability between pairs of raters who received this brief practical session on how to conduct clinical tests was not better than the agreement and reliability of two raters who only received basic instruction (i.e., leaflet and a video of the performance of all clinical shoulder tests). Although the groups of patients and physical therapists were different with respect to some characteristics, we think that the explanation for not finding a difference between raters who received a brief practical session and those who didn't, is that most of the participating physical therapists were experienced. Besides, almost 50% of the participating physical therapists had post-graduate qualifications. It is reasonable to believe that an additional brief practical session will not add much to the proficiency of experienced clinicians.

Limitations

The findings of this investigation need to be interpreted in the light of certain limitations. Patients were subjected to a large number of tests, on two separate sessions with a maximum time interval of 30 minutes. We chose independent examinations in order not to eliminate the variability in patient status and therapist method. An alternate method involves making videos of patients with one physical therapist that performs the clinical shoulder tests (Lewis et al., 2016). These videos can be assessed by several examiners. These videos, however, do not mimic clinical practice. We realized that an extensive test procedure performed by two different physical therapists can cause exacerbation of symptoms which can lead to poor agreement and reliability values. Therefore, we expected that a small subgroup of patients would report more pain after the first examination, although the physical therapists were instructed to avoid unnecessary pain with the shoulder testing procedure. In the present study, 21 patients reported a change of > 2 points on an 11-point NRS after the first examination. To our surprise, 10 of these 21 patients reported a reduction of pain rather than an increase. The test procedure involved provocative tests and symptom reduction tests. Although it is speculation, it might be that some patients had benefit of the reduction tests, and reported less pain after the first examination. An additional analysis in which patients with a clinically relevant change in pain between the two examinations (> 2 points of change on an 11-point NRS) were excluded did not lead to other agreement or reliability values. Therefore, we do not think that the extensive test procedure had a major effect on our results.

Patients were included with acute, subacute and chronic shoulder pain. The usefulness of clinical

shoulder tests in patients with chronic pain can be questioned. Contrary to acute pain in which nociception is presumed to be the primary source of pain, chronic pain is a more complex condition. For example, several studies have provided evidence for altered central nociceptive processing in subgroups of patients with persisting pain which is suggestive of central sensitization (Nijs et al., 2014). Due to this process of generalized hypersensitivity of the somatosensory system, patients can react with an increased response on clinical shoulder tests. There is reason to believe, that in these patients, clinical shoulder tests are less valid because of a high potential for false positive testing. However, as the aim of the present study was related to agreement and reliability, and not validity, we do not think that the possibility of dominant central sensitization pain in our patients has had a major influence on our results.

Pairing of raters and the order of testing were not randomized due to organizational constraints. Although we do not think that this has substantially influenced our results, the possibility of bias cannot be excluded.

The proportion of missing values and not applicable scores was low (8.7%). Several of the not applicable scores were logical (e.g., the Relocation Test is not applicable when the Apprehension Test is negative). Although the impact of missing values on our results is uncertain, the not applicable scores did not influence our interrater agreement and reliability values, because in these cases, calculations could not be performed.

The majority of the patients (75%) were assessed and reassessed at the first visit at the clinic where physical therapists had no or minor information about the patient. In the remaining cases (25%), the physical therapist who treated the patient was informed about the patient history and clinical diagnosis. It is uncertain if this has influenced our results.

A last point of concern is that participating physical therapists did not collect characteristics of patients that refused to participate. Although potential selection bias cannot be excluded, we consider the risk to be low.

Clinical implications

It is important to realize that for the performance of most clinical shoulder tests several modifications have been proposed. We advise clinicians to discuss advantages and disadvantages of the original performance and their modifications with colleagues, to assess patients together occasionally and to discuss disagreements. Tests that cannot be performed with acceptable reliability offer little or no value in the clinical assessment of patients.

Conclusion

The present study is in line with the GRRAS guidelines and provides data for shoulder tests that has never been published before in patients with shoulder pain. The findings of the present study show that most clinical shoulder tests do not show sufficient interrater agreement and reliability values for clinical use.

Acknowledgments

This work was supported by the Dutch Scientific College of Physiotherapy (WCF) of the Royal Dutch Society for Physical Therapy (KNGF).

Funding

This work was supported by the Dutch Scientific College of Physiotherapy (WCF) of the Royal Dutch Society for Physical Therapy (KNGF) (no grant number).

Disclosure Statement

The authors declare that they have no conflict of interest.

References

Altman DG 1997 Practical Statistics for Medical Research. London: Chapman and Hall.

Barten D, Koppes L 2016 Zorg door de fysiotherapeut. Jaarcijfers 2015 en trendcijfers 2011–2015 (Care delivered by the physical therapist. In: The year 2015 and trend between 2011–2015). Utrecht: Nivel.

Bertilson BC, Grunnesjo M, Strender LE 2003 Reliability of clinical tests in the assessment of patients with neck/shoulder problems-impact of history. Spine 28: 2222–2231.

Boon AJ,Smith J 2000 Manual scapular stabilization: Its effect on shoulder rotational range of motion. Archives of Physical Medicine and Rehabilitation 81: 978–83.

Burns SA, Cleland JA, Carpenter K, Mintken PE 2016 Interrater reliability of the cervicothoracic and shoulder physical examination in patients with a primary complaint of shoulder pain. Physical Therapy in Sport : Official Journal of the Association of Chartered Physiotherapists in Sports Medicine 18: 46–55.

Cadogan A, Laslett M, Hing W, McNair P, Williams M 2011 Interexaminer reliability of orthopaedic special tests used in the assessment of shoulder pain. Manual Therapy 16: 131–135.

Cools AM, Cambier D, Witvrouw EE 2008 Screening the athlete's shoulder for impingement symptoms: A clinical reasoning algorithm for early detection of shoulder pathology. British Journal of Sports Medicine 42: 628–635.

Corso G 1995 Impingement relief test: an adjunctive procedure to traditional assessment of shoulder impingement syndrome. Journal of Orthopaedic and Sports Physical Therapy 22: 183–92.

Cyriax J 1982 Textbook of Orthopaedic Medicine Volume I. Diagnosis of soft Tissue Lesions (8th ed). London, Ballière Tindall.

D'Hondt NE, Kiers H, Pool JJ, Hacquebord ST, Terwee CB, Veeger DH 2017 Reliability of performance-based clinical measurements to assess shoulder girdle kinematics and positioning: Systematic review. Physical Therapy 97: 124–144.

de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL 2013 Clinicians are right not to like Cohen's kappa. British Medical Journal 346: f2125.

de Vet HC, Terwee CB, Mokkink LB, Knol DL 2011 Measurement in Medicine. New York: Cambridge University Press.

Egmond DL, Schuitemaker R 2014 Extremiteiten Manuele Therapie in Enge en Ruime Zin (Extremities Manual Therapy Including both Broad and Narrow Definition) (11th ed), Houten: Bohn Stafleu van Loghum.

Eshoj H, Ingwersen KG, Larsen CM, Kjaer BH, Juul-Kristensen B 2018 Intertester reliability of clinical shoulder instability and laxity tests in subjects with and without self-reported shoulder problems. British Medical Journal Open 8: e018472.

Greving K, Dorrestijn O, Winters JC, Groenhof F, van der Meer K, Stevens M, Diercks RL 2012 Incidence, prevalence, and consultation rates of shoulder complaints in general practice. Scandinavian Journal of Rheumatology 41: 150–155.

Hawkins RJ, Kennedy JC 1980 Impingement syndrome in athletes. American Journal of Sports Medicine 8: 151–8.

Hayes KW, Petersen CM 2003 Reliability of classifications derived from Cyriax's resisted testing in subjects with painful shoulders and knees. Journal of Orthopaedic and Sports Physical Therapy 33: 235–246.

Hegedus EJ, Cook C, Lewis J, Wright A, Park JY 2015 Combining orthopedic special tests to improve diagnosis of shoulder pathology. Physical Therapy in Sport : Official Journal of the Association of Chartered Physiotherapists in Sports Medicine 16: 87–92.

Holtby R, Razmjou H 2004 Validity of the supraspinatus test as a single clinical test in diagnosing patients with rotator cuff pathology. Journal of Orthopaedic and Sports Physical Therapy 34: 194–200.

Hripcsak G, Heitjan DF 2002 Measuring agreement in medical informatics reliability studies. Journal of Biomedical Informatics 35: 99–110.

Jain NB, Luz J, Higgins LD, Dong Y, Warner JJ, Matzkin E, Katz JN 2017 The diagnostic accuracy of special tests for rotator cuff tear: The ROW cohort study. American Journal of Physical Medicine and Rehabilitation 96: 176–183.

Jobe FW, Jobe CM 1983 Painful athletic injuries of the shoulder. Clinical orthopaedics and related research: 117–24.

Johansson K, Ivarson S 2009 Intra- and interexaminer reliability of four manual shoulder maneuvers used to identify subacromial pain. Manual Therapy 14: 231–239.

Karel YH, Verhagen AP, Thoomes-de Graaf M, Duijn E, van Den Borne MP, Beumer A, Ottenheijm RP, Dinant GJ, Koes BW, Scholten-Peeters GG 2016 Development of a prognostic model for patients with shoulder complaints in physical therapist practice. Physical Therapy 97: 72–80.

Kelly BT, Kadrmas WR, Speer KP 1996 The manual muscle examination for rotator cuff strength. An electromyographic investigation. American Journal of Sports Medicine 24: 581–8.

Kibler WB, Sciascia A, Dome D 2006 Evaluation of apparent and absolute supraspinatus strength in patients with shoulder injury using the scapular retraction test. American Journal of Sports Medicine 34: 1643-7.

Kibler WB, Uhl TL, Maddux JW, Brooks PV, Zeller B, McMullen J 2002 Qualitative clinical evaluation of scapular dysfunction: a reliability study. Journal of Shoulder and Elbow Surgery 11: 550–6.

Kim SH, Ha KI, Ahn JH, Kim SH, Choi HJ 2001 Biceps load test II: A clinical test for SLAP lesions of the shoulder. Arthroscopy 17: 160–164.

Kim SH, Park JS, Jeong WK, Shin SK 2005 The Kim test: A novel test for posteroinferior labral lesion of the shoulder–A comparison to the jerk test. American Journal of Sports Medicine 33: 1188–1192.

Kooijman M, Swinkels I, van Dijk C, de Bakker D, Veenhof C 2013 Patients with shoulder syndromes in general and physiotherapy practice: An observational study. BMC Musculoskeletal Disorders 14: 128.

Kopkow C, Lange T, Schmitt J, Kasten P 2015 Interrater reliability of the modified scapular assistance test with and without handheld weights. Manual Therapy 20: 868–874.

Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL 2011 Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Journal of Clinical Epidemiology 64: 96–106.

Lange T, Matthijs O, Jain NB, Schmitt J, Lutzner J, Kopkow C 2017 Reliability of specific physical examination tests for the diagnosis of shoulder pathologies: A systematic review and meta-analysis. British Journal of Sports Medicine 51: 511–518.

Lewis JS 2009 Rotator cuff tendinopathy/subacromial impingement syndrome: Is it time for a new method of assessment? British Journal of Sports Medicine 43: 259–264.

Lewis JS, McCreesh K, Barratt E, Hegedus EJ, Sim J 2016 Interrater reliability of the shoulder symptom modification procedure in people with shoulder pain. British Medical Journal Open Sport and Exercise Medicine 2: e000181.

May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M 2010 Reliability of physical examination tests used in the assessment of patients with shoulder problems: A systematic review. Physiotherapy 96: 179–190.

Michener LA, Walsworth MK, Doukas WC, Murphy KP 2009 Reliability and diagnostic accuracy of 5 physical examination tests and combination of tests for subacromial impingement. Archives of Physical Medicine and Rehabilitation 90: 1898–1903.

Nanda R, Gupta S, Kanapathipillai P, Liow RYL, Rangan A 2008 An assessment of the inter examiner reliability of clinical tests for subacromial impingement and rotator cuff integrity. European Journal of Orthopaedic Surgery and Traumatology 18: 6.

Neer CS 2nd 1983 Impingement lesions. Clinical Orthopaedics and Related Research: 70–77.

Nijs J, Torres-Cueco R, van Wilgen CP, Girbes EL, Struyf F, Roussel N, van Oosterwijck J, Daenen L, Kuppens K, Vanwerweeen L et al. 2014 Applying modern pain neuroscience in clinical practice: Criteria for the classification of central sensitization pain. Pain Physician 17: 447–457.

Nørregaard J, Krogsgaard MR, Lorenzen T, Jensen EM 2002 Diagnosing patients with longstanding shoulder joint pain. Annals of the Rheumatic Diseases 61: 646–649.

O'Brien SJ, Pagnani MJ, Fealy S, McGlynn SR, Wilson JB 1998 The active compression test: A new and effective test

for diagnosing labral tears and acromioclavicular joint abnormality. American Journal of Sports Medicine 26: 610-3.

Ostor AJ, Richards CA, Prevost AT, Hazleman BL, Speed CA 2004 Interrater reproducibility of clinical tests for rotator cuff lesions. Annals of the Rheumatic Diseases 63: 1288–1292.

Picavet HS, Schouten JS 2003 Musculoskeletal pain in the Netherlands: Prevalences, consequences and risk groups, the DMC(3)-study. Pain 102: 167–178.

Pope DP, Croft PR, Pritchard CM, Silman AJ 1997 Prevalence of shoulder pain in the community: The influence of case definition. Annals of the Rheumatic Diseases 56: 308–312.

Rabin A, Irrgang JJ, Fitzgerald GK, Eubanks A 2006 The intertester reliability of the scapular assistance test. Journal of Orthopaedic and Sports Physical Therapy 36: 653–660.

Razmjou H, Holtby R, Myhr T 2004 Pain provocative shoulder tests: Reliability and validity of the impingement tests. Physiotherapy Canada 56: 229–236.

Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y 1991 Development of a shoulder pain and disability index. Arthritis Care and Research: The Official Journal of the Arthritis Health Professions Association 4: 143–149.

Silliman JF, Hawkins RJ 1993 Classification and physical diagnosis of instability of the shoulder. Clinical Orthopaedics and Related Research: 7–19.

Tempelhof S, Rupp S, Seil R 1999 Age-related prevalence of rotator cuff tears in asymptomatic shoulders. Journal of Shoulder and Elbow Surgery 8: 296–299.

Thoomes-de Graaf M, Scholten-Peeters GG, Duijn E, Karel Y, Koes BW, Verhagen AP 2015 The Dutch Shoulder Pain and Disability Index (SPADI): A reliability and validation study. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 24: 1515–1519.

Tong HC, Haig AJ, Yamakawa K 2002 The Spurling test and cervical radiculopathy. Spine 27: 156–159.

Tooth LR, Ottenbacher KJ 2004 The kappa statistic in rehabilitation research: An examination. Archives of Physical Medicine and Rehabilitation 85: 1371–1376.

Tzannes A, Paxinos A, Callanan M, Murrell GA 2004 An assessment of the interexaminer reliability of tests for shoulder instability. Journal of Shoulder and Elbow Surgery 13: 18–23.

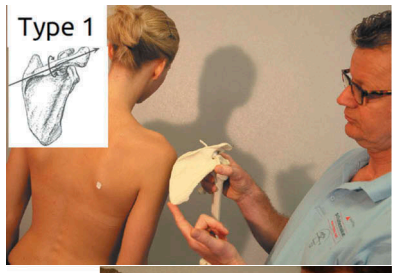
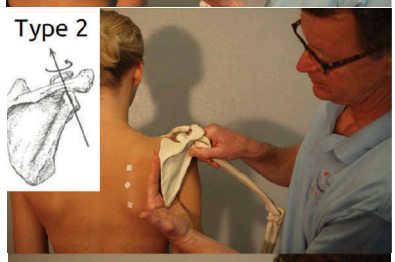
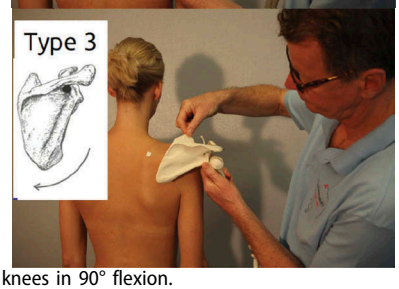
Vind M, Bogh SB, Larsen CM, Knudsen HK, Sogaard K, Juul-Kristensen B 2011 Inter-examiner reproducibility of clinical tests and criteria used to identify subacromial impingement syndrome. British Medical Journal Open 1: e000042.

Walsworth MK, Doukas WC, Murphy KP, Mielcarek BJ, Michener LA 2008 Reliability and diagnostic accuracy of history and physical examination for diagnosing glenoid labral tears. American Journal of Sports Medicine 36: 162–168.

Worcester JN, Green DP 1968 Osteoarthritis of the acromioclavicular joint. Clinical Orthopaedics and Related Research 58: 69-73.

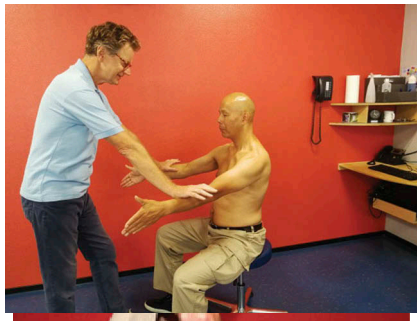
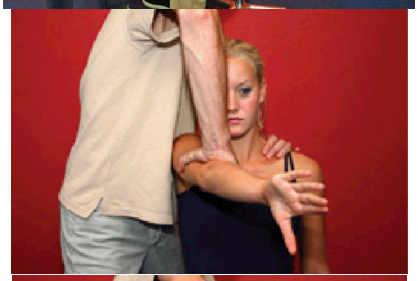
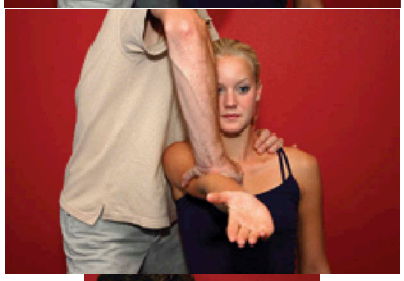
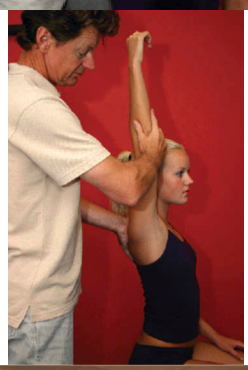
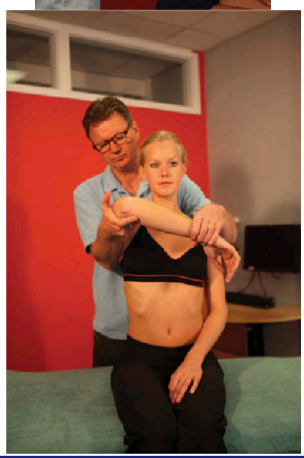
Zaslav KR 2001 Internal rotation resistance strength test: A new diagnostic test to differentiate intra-articular pathology from outlet (Neer) impingement syndrome in the shoulder. Journal of Shoulder and Elbow Surgery 10: 23–7.

Appendix 1. Test procedures

Test	Performance	Criteria	Photo
Patient in standardized standing position: Arms at the sides of the body, elbows straight and shoulders in neutral position. Feet shoulder width apart. Light contraction of m. transversus abdominis.			
Scapula Position Kibler et al, 2002	Examiner is standing behind the patient and observes the position of the scapula of the affected side.	Types 1, 2, 3, or normal. 1: Tipping: increased anterior tilt, inferior angle of the scapula is prominent.	
		2: Winging or scapula alatae: increased internal rotation (protraction shoulder girdle), medial border of the scapula is prominent.	
		3: Upward rotation of the superomedial border (***cavitas*** glenoidalis scapulae is pointed caudal), superior angle of the scapula is prominent.	
Patient in standardized seated position: Extended lumbar spine. Both feet flat on the ground. Hips and knees in 90° flexion.			
External Rotation Resistance Test Cyriax, 1982	Patient's elbow is at the side of the body, elbow 90° flexed, forearm in neutral position (thumbs up). Examiner is standing at the same side of the patient with the hip against the tested shoulder. One hand fixates the trunk by pulling the patient's shoulder against his side. The other hand is placed on the lateral surface of the forearm proximal to the patient's wrist, and applies pressure. Patient is asked to resist the slowly increasing pressure.	Test is considered positive if pain is reported.	
Empty Can Test (Jobe Test) Jobe and Jobe, 1983	Both arms in 90° abduction in the scapular plane. Elbows slightly flexed, thumbs pointing downwards (internal rotation shoulder and forearm). Examiner is standing in front of patient and asks the patient to hold this position against slowly increasing downward pressure applied to elbow (into adduction/internal rotation).	Test is considered positive if pain or weakness is reported. The test is also positive if the position itself produces pain.	

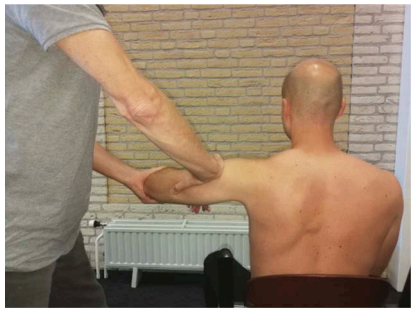
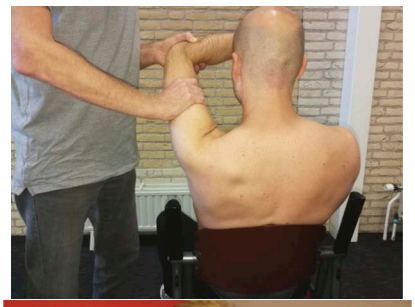
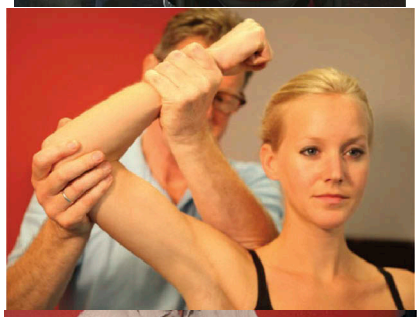
(*Continued*)

(Continued).

Test	Performance	Criteria	Photo
Full Can Test Kelly, Kadrmas, and Speer, 1996	Both arms in 90° abduction in the scapular plane. Elbows extended, forearm supinated and thumbs pointing upwards. Examiner is standing in front of patient and asks the patient to remain this position against slowly increasing downward pressure applied to the elbows (into adduction).	Test is considered positive if pain or weakness is reported. The test is also positive if the position itself produces pain.	
Active Compression Test (O'Brien's Test) O'Brien et al, 1998	Patient's arm in 90° flexion, 10–15° horizontal adduction. Examiner is standing behind or at the same side of the patient and provides pressure downwards with one hand on the distal part of the humerus.	Test is considered positive if pain is reported in the first position (internal rotation) and is significant less or disappears in the second position (external rotation).	
	First, the patient is asked to maximally internally rotate the arm (thumb down) and resist slowly increasing pressure which is applied in a downward direction. The patient is then asked to maximally externally rotate the arm (thumb up) and is asked again to resist the downward pressure.		
Neer Test Neer, 1983	Examiner is standing behind the patient. One hand palpates and controls the scapula (no force is applied). The other hand holds patients extended elbow and forces the patient's arm into maximal elevation.	Test is considered positive if pain is produced.	
Hawkins-Kennedy Test Hawkins and Kennedy,1980	Patient's arm in 90° flexion, elbow in 90° flexion. Forearm is internally rotated and positioned in the transversal plane. Examiner is standing behind the patient and supports this position with both arms in order to relax the patient's arm. The shoulder is then passively gently rotated internally by lowering the forearm while supporting the elbow.	Test is considered positive if pain is produced during internal rotation. The test is also positive if the position itself produces pain.	

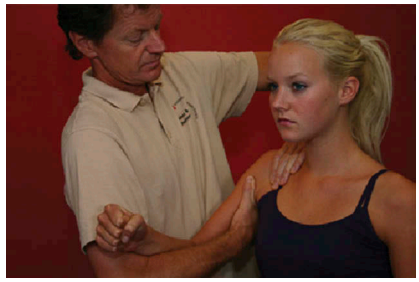
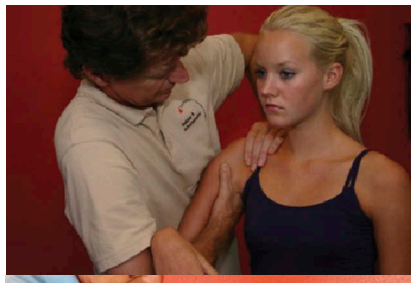
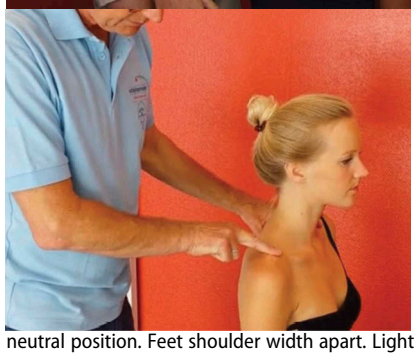
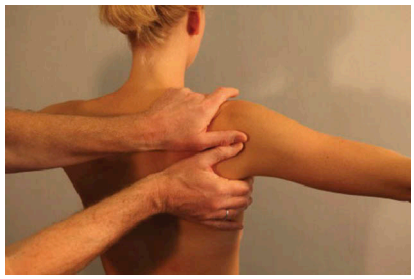
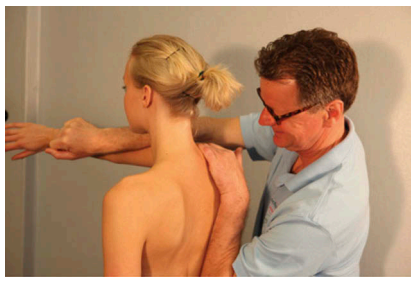
(*Continued*)

(Continued).

Test	Performance	Criteria	Photo
Kim Test Kim, Park, Jeong, and Shin, 2005	Patient's arm in 90° abduction, elbow in 90° flexion. Forearm is internally rotated and positioned in transversal plane. Examiner is standing next to the patient on the same side as the tested shoulder and supports this position with one hand at the patient's elbow, and the other hand on the proximal arm. The examiner applies an strong axial loading force along the line of the proximal arm from the elbow toward the glenoid. While a downward and backward force is applied to the proximal arm, the arm is elevated 45°diagonally upward (to 135° flexion-elevation). During the test, it is important to apply a firm axial compression force to the glenoid surface by the humeral head.	Test is considered positive if a sudden onset of posterior shoulder pain is produced during the test, regardless of accompanying posterior clunk of the humeral head.	 
Biceps Load II Test Kim et al., 2001	Patient's arm in 120° abduction and maximal external rotation. Elbow in 90° flexion. Examiner is standing behind the patient and applies an axial force along the line of the humerus from the elbow to glenoid with one hand while supporting the external rotation of the forearm. The patient is asked to flex the elbow against resistance.	Test is considered positive if pain is reported.	
Internal Rotation Resistance Strength Test Zaslav, 2001	Patient's arm in 90° abduction, elbow in 90° flexion and 80° (submaximal) external rotation. Examiner is standing at the side of the patient's tested shoulder. One hand is placed at medial border of the elbow, the other hand at the distal/dorsal part of the forearm. First, gradual force in the direction of internal rotation is applied. The patient responds with an isometric external rotation force. Then the examiner switches hands and applies a gradually force in the direction of external rotation. The patient responds with isometric internal rotation force.	Test is considered positive if the patient has good strength in external rotation but apparent weakness in internal rotation.	 

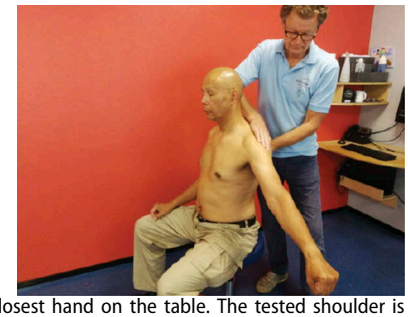
(*Continued*)

(Continued).

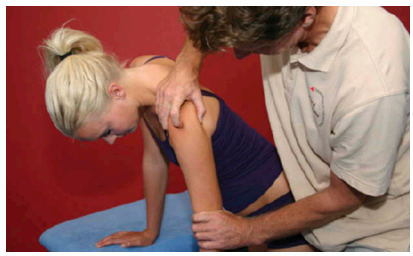
Test	Performance	Criteria	Photo
Load and Shift Test Silliman and Hawkins, 1993	Patient's arm in loose packed position. Examiner is standing on the same side as the tested shoulder and grasps the humerus as proximal as possible. The patient's forearm is rested on the examiner's forearm. The other hand of the examiner fixates the clavicle, acromion and coracoid process. The patient is asked to active 'load' the glenohumeral joint with an isometric contraction of the shoulder muscles. After relaxation the examiner shifts the humeral head in the anterior and posterior directions with a maximum of grade 1 translation. This is a slightly modified version of the load and shift test described by Silliman and Hawkins (1993) in which one hand grasp the humeral head, and load it into the glenoid.	Test is considered positive at grade 1. Grade 0: Barely no translation (<25% diameter humeral head). Grade 1: Translation within the rim of the glenoid (25–50% diameter humeral head). Note: Grade 2: Subluxation over the rim of the glenoid (>50% diameter humeral head) but spontaneous reduction (does not lock). Grade 3: Dislocation over the rim of the glenoid (>50% diameter humeral head) with no spontaneous reduction (locks out over the rim).	 
Acromioclavicular Joint Stress Test Worcester and Green, 1968	Patient's arm is relaxed at the side of the body. Examiner is standing behind the patient and gradually applies downward pressure with fore finger and middle finger directly over the acromioclavicular joint.	Test is considered positive if pain is reported on top of the acromioclavicular joint.	
Patient in standardized standing position: Arms at the sides of the body, elbows straight and shoulders in neutral position. Feet shoulder width apart. Light contraction of m. transversus abdominis.			
Modified Scapular Assistance Test Rabin, Irrgang, Fitzgerald, and Eubanks, 2006	The patient is asked to actively elevate the arm in a provocative direction (e.g., sagittal (anteflexion), frontal (abduction), or scapular (scaption) plane). Examiner is standing diagonally behind the patient and places one hand over the superior part of the scapula. The other hand is placed over the inferior part of the scapula. During the movement the examiner facilitates upward rotation and posterior tilting by pushing the lower part of the scapula against the thorax.	Test is considered positive if a significant reduction of pain, symptoms or effort occurs during assisted elevation, compared to elevation without assistance. Test is considered not applicable if the patient does not report symptoms with active elevation.	
Scapular Retraction Test Kibler, Sciascia, and Dome, 2006	A positive clinical provocation test (for example the Active Compression Test) is repeated. Examiner is standing behind the patient and places one hand on the upper trapezius and acromion of the tested shoulder and pulls the shoulder girdle in retraction. The scapula is lightly held in retraction by forearm pressure on the medial scapular border. With the other hand the clinical provocation test is repeated.	The test is positive when the initial pain, present in the clinical provocation test reduces significantly or disappears during the Scapular Retraction Test. Test is considered not applicable if the patient does not report symptoms with clinical provocation tests like the Active Compression Test, the Full Can Test and the Empty Can Test.	

(*Continued*)

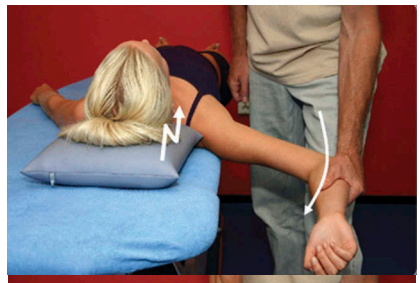
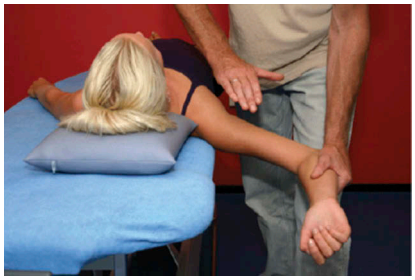
(Continued).

Impingement Relief Test Corso, 1995	The patient is asked to actively elevate the arm in a provocative direction (e.g., sagittal [anteflexion], frontal [abduction], or scapular [scaption] plane). Examiner is standing diagonally behind the patient and applies a gentle inferior glide to the humeral head just prior to the onset of the painful arc or "impingement sign".	Test is considered positive if a reduction of pain, symptoms or effort occurs during assisted elevation, compared to elevation without inferior glide pressure. Test is considered not applicable if the patient does not report symptoms with active elevation.	

Patient in standing position next to examiner's table: Trunk 45° flexed. One foot in front of the other. Closest hand on the table. The tested shoulder is hanging relaxed in 20° flexion.

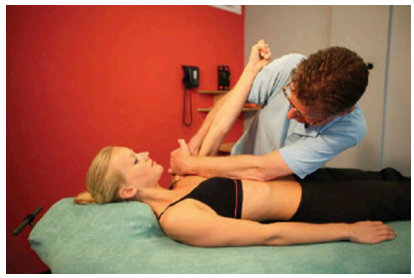
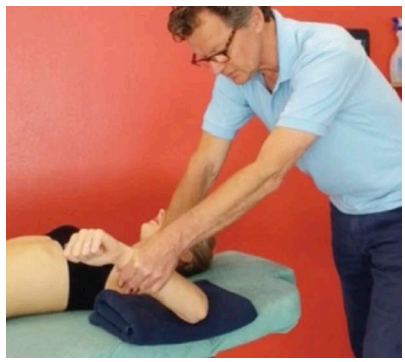
Sulcus Sign Test Silliman and Hawkins, 1993	Patient's arm is relaxed in loose packed position. The examiner is standing diagonally behind the patient and grasps the acromion, clavicle, coracoid process and spinae scapula with one hand. This hand palpates the size of the gap between acromion and caput humeri (sulcus). The other hand holds the distal part of the humerus just above the lateral epicondyl and applies a downward force.	Test is considered positive at grade 1 or 2: Grade 0: No sulcus sign (no enlarged gap) Grade 1: Sulcus sign, enlarged gap of one centimeter Grade 2: Sulcus sign, enlarged gap of two centimeter	

Patient in supine position at the side of the table but with the scapula fully supported by the table.

Apprehension Test Tzannes, Paxinos, Callanan, and Murrell, 2004	Examiner is standing next to the side of the patient. Patient's arm is in 90° abduction and maximal external rotated (submaximal closed-packed position), elbow in 90° flexion. The examiner applies a horizontal abduction- external rotation to the end range of motion (or until the patient's request for the examiner to stop).	Test is considered positive if pain or apprehension is reported.	
Relocation Test Tzannes, Paxinos, Callanan, and Murrell, 2004	Examiner is standing next to the side of the patient. Patient's arm is in 90° abduction and maximal external rotated (submaximal closed-packed position), elbow in 90° flexion. The examiner applies a horizontal abduction- external rotation to the end range of motion (or until the patient's request for the examiner to stop). The examiner's other hand applies a posterior translation (downward) force on the head of the humerus.	Test is considered positive if pain or apprehension is reduced with the applied posterior translation force. Test is not applicable if the Apprehension Test is negative.	
Release Test Tzannes, Paxinos, Callanan, and Murrell, 2004	Examiner is standing next to the side of the patient. Patient's arm is in 90° abduction and maximal external rotated (submaximal closed-packed position), elbow in 90° flexion. The examiner applies a horizontal abduction-external rotation to the end range of motion (or until the patient's request for the examiner to stop). The examiner's other hand applies a posterior translation (downward) force on the head of the humerus. The examiner suddenly release the posterior/downward pressure of the humeral head while holding the patient's arm in the position of apprehension.	Test is considered positive if pain or apprehension is returned after removal of the applied posterior translation force. Test is not applicable if the Apprehension Test and Relocation Test are negative.	

(*Continued*)

(Continued).

Combined Reduction Test Egmond and Schuitemaker, 2014	The examiner is standing at the side of the patient between the trunk and the arm of the patient. Patient places the arm at the side of the examiner's shoulder. To center the caput humerus into the cavitas glenoidalis the patient is asked to make a fist in order to establish an isometric contraction of the shoulder muscles. This arm is guided actively to 60° flexion by the examiner. The arm of the examiner which supports the patient's arm is placed on the lateral part of the m. deltoideus. The other arm of the examiner applies light pressure with the forearm on the sternum of the patient (to stabilize the scapula on the table) and this hand is placed on top of the other hand. Both hands of the examiner (hand-over-hand technique) are applying facilitating, proximal resistance on the m. deltoideus region during the complete movement from 60° to 90° flexion. This pressure will result in a safe feeling of stability. Rotation is neutral. From this point (90° flexion) the examiner moves his body to cranial to continue the actively guided final goal of 180° flexion. The patient flex the arm active and the examiner partly resist this movement. This should be pain free. After this procedure, the patient moves the arm back to neutral position against resistance. The test is repeated three to five times. The technique is an attempt to combine the several reduction techniques, i.e., Modified Scapular Assistance Test, Scapular Retraction Test and Impingement Relief Test.	Examiner asks and observes the patient if 'centering' is successful. The patient determines the maximum pain free range of motion. Test is considered positive if pain free range of motion increases with at least 20° or a significant reduction of symptoms occurs after 3–5 repetitions.	
Glenohumeral Internal Rotation Deficit Test Boon and Smith, 2000	Patient's arm is in 90° abduction. Elbow in 90° flexion. Forearm is neutral. A towel is placed under the upper arm of the patient to prevent horizontal abduction. Examiner is standing next to the head of the patient and fixates the scapula on the examiner's table by pressing the frontal side of the anterior chest wall downwards. The other hand performs internal rotation of the shoulder (in 90° abduction) until resistance or pain occurs. The examiner measures the ROM with a goniometer or by visual inspection.	Test is considered positive if the passive ROM is 20° or more restricted compared to the other side.	

REFERENCES

Boon AJ, Smith J 2000 Manual scapular stabilization: its effect on shoulder rotational range of motion. Archives of Physical Medicine and Rehabilitation 81: 978–983.

Corso G 1995 Impingement relief test: an adjunctive procedure to traditional assessment of shoulder impingement syndrome. Journal of Orthopaedic and Sports Physical Therapy 22: 183–192.

Cyriax J 1982 Textbook of Orthopaedic Medicine Volume I. Diagnosis of Soft Tissue Lesions (8th ed). London, Ballière Tindall.

Egmond DL, Schuitemaker R 2014 Extremiteiten manuele therapie in enge en ruime zin (Extremities manual therapy including both broad and narrow definition) (11th ed). Houten: Bohn Stafleu van Loghum.

Hawkins RJ, Kennedy JC 1980 Impingement syndrome in athletes. American Journal of Sports Medicine 8: 151–158.

Jobe FW, Jobe CM 1983 Painful athletic injuries of the shoulder. Clinical Orthopaedics and Related Research: 117–124.

Kelly BT, Kadrmas WR, Speer KP 1996 The manual muscle examination for rotator cuff strength. An electromyographic investigation. American Journal of Sports Medicine 24: 581–588.

Kibler WB, Sciascia A, Dome D 2006 Evaluation of apparent and absolute supraspinatus strength in patients with shoulder injury using the scapular retraction test. American Journal of Sports Medicine 34: 1643–1647.

Kibler WB, Uhl TL, Maddux JW, Brooks PV, Zeller B, McMullen J 2002 Qualitative clinical evaluation of scapular dysfunction: a reliability study. Journal of Shoulder and Elbow Surgery 11: 550–556.

Kim SH, Ha KI, Ahn JH, Kim SH, Choi HJ 2001 Biceps load test II: A clinical test for SLAP lesions of the shoulder. Arthroscopy 17: 160–164.

Kim SH, Park JS, Jeong WK, Shin SK 2005 The Kim test: A novel test for posteroinferior labral lesion of the shoulder–a comparison to the jerk test. American Journal of Sports Medicine 33: 1188–1192.

Neer CS 2nd 1983 Impingement lesions. Clinical Orthopaedics and Related Research: 70–77.

O'Brien SJ, Pagnani MJ, Fealy S, McGlynn SR, Wilson JB 1998 The active compression test: a new and effective test for diagnosing labral tears and acromioclavicular joint abnormality. American Journal of Sports Medicine 26: 610–613.

Rabin A, Irrgang JJ, Fitzgerald GK, Eubanks A 2006 The intertester reliability of the Scapular Assistance Test. Journal of Orthopaedic and Sports Physical Therapy 36: 653–660.

Silliman JF, Hawkins RJ 1993 Classification and physical diagnosis of instability of the shoulder. Clinical Orthopaedics and Related Research: 7–19.

Tzannes A, Paxinos A, Callanan M, Murrell GA 2004 An assessment of the interexaminer reliability of tests for shoulder instability. Journal of Shoulder and Elbow Surgery 13: 18–23.

Worcester JN Jr, Green DP 1968 Osteoarthritis of the acromioclavicular joint. Clinical Orthopaedics and Related Research 58: 69–73.

Zaslav KR 2001 Internal rotation resistance strength test: A new diagnostic test to differentiate intra-articular pathology from outlet (Neer) impingement syndrome in the shoulder. Journal of Shoulder and Elbow Surgery 10: 23–27.