Clinimetric evaluation of measurement tools used in hand therapy to assess activity and participation

Karin Schoneveld, MsC, PT
Utrecht University, The Netherlands
Rehabilitation Department, Medical Centre Alkmaar, The Netherlands

Harriet Wittink, PhD, MsC, PT
Health and Lifestyle Research, University of Applied Sciences Utrecht, The Netherlands

Tim Takken, PhD, MsC, PT
Utrecht University, The Netherlands

ABSTRACT:
Study design: Systematic review.
Introduction: A number of measurement tools with strong clinimetric properties address activities and participation in hand-injured persons.
Purpose of the study: To evaluate clinimetric quality of measurement tools assessing activities and participation in patients with hand injuries.
Methods: The electronic databases Medline, Cochrane library, EMBASE, PEDro, Cumulative Index to Nursing and Allied Health Literature (ClNAHL), and Allied and Complementary Medicine Database (AMED) were searched for appropriate literature. Descriptive data of included tools were given, and their clinimetric quality was scored with specific criteria.
Results: The literature search identified 696 publications, referring to 15 measurement tools that met the inclusion criteria. For most of the included tests, applicability was good, whereas information on clinimetric properties was often lacking, especially for the performance tests. Overall, the Disabilities of the Arm, Shoulder, and Hand (DASH) was the most extensively studied tool with positive ratings for all criteria, closely followed by the Michigan Hand Outcomes Questionnaire (MHQ). Of the performance tests, the Functional Dexterity Test (FDT) received the best ratings.
Conclusions: Optimal measurement tool selection depends highly on the purpose of measurement and the type of hand injury.
Level of Evidence: 2a-.
J HAND THER. 2009;■■:■–■.

Hand therapists use these measures to determine a patient's health status, to predict subsequent events, and to evaluate change over time. Measurement activities compromise 20% of the therapist's time and are ranked as the most critical part of daily practice.[1,2]

Evidence-based clinical decision making requires measurement tools with good clinimetric properties covering all domains of the International Classification of Functioning, Disability, and Health (ICF).[3–6]

Traditionally, outcome assessment in hand therapy is focused on measures of range of motion (ROM), strength, and sensation. However, in the last decade the focus has shifted toward assessing health at the activity and participation level.[7,8] This patient-centered approach matches the main goal of hand therapy, which is to maximize activities and participation in life situations for individuals with disease or injuries of the upper extremity.[9,10]

Although the importance of measuring activities and participation is generally recognized, implementation in clinical practice and research needs to improve. Two systematic reviews of outcome assessment in upper-extremity trials showed that only a small number of studies used measurements of activity and/or participation.[11,12] Although these reviews classified the measurement tools according to the ICF domains, studies comparing their clinimetric quality are few. Strengths and weaknesses of objective measures[13] and self-reported questionnaires[14] used in hand surgery, including activities and participation, were described. Both authors stress the importance of clinimetric properties of measurement tools such as reliability, validity, and applicability, but a systematic overview of their quality was not provided.

The purpose of this paper is to systematically review the content and clinimetric quality of published measurement tools assessing activities and participation in patients with hand injuries.

Correspondence and reprint requests to K. Schoneveld, Medical Centre Alkmaar, Rehabilitation Department 081, Wilhelminalaan 12, Postbus 501, 1800 AM Alkmaar, The Netherlands.

doi:10.1016/j.jht.2008.11.005

Therefore, this paper attempts to answer the following questions:

- Which measurement tools to assess activities and participation in patients with hand injuries are described in the literature?
- What are the psychometric properties of these measurement tools?
- What is the applicability for use in daily clinical practice of these measurement tools?

METHODS

Literature search

The following electronic databases were searched by the first author (K S) for appropriate citations; Medline (1980–October 2007), the Cochrane Library (1980–October 2007), EMBASE (1980–October 2007), PEDro (1980–October 2007), ClNAHL (1982–October 2007), and AMED (1985–October 2007).

A two-stage search strategy was used. In stage one, studies using or describing measurement tools for therapists on activities and participation used in persons with hand injuries were searched. The following text words and Mesh-terms were combined with "OR" in four steps: 1. Hand Injuries [Mesh], hand injuries, hand injury; 2. Physical Therapy [Mesh], physiotherapy, Occupational Therapy [Mesh], hand therapy; 3. Treatment Outcome [Mesh], Outcome Assessment [Mesh], Outcome and Process Assessment [Mesh], measure, assess, scale, monitor, score; 4. Activity, participation, disability, Quality of Life [Mesh], functional status, functional ability, hand function, dexterity, Activities of Daily Living [Mesh], capability, performance, Task Performance and Analysis [Mesh]. Then, search outcomes for each step were combined by "AND."

In stage two, the names of the identified measurement tools found in stage one were entered in the electronic databases as search terms. Abbreviations and shorts forms commonly used for these tools were also entered as search terms. Publications found during this second stage were screened for information on clinimetric quality of the measurement tools. References of retrieved articles were screened for additional relevant literature in both stages.

Terminology

- "Hand injury" is defined as "a disease or injury of the upper extremity, which is not caused by injury to the central nervous system."
- The terms "activity" and "participation" are defined according to the ICF.[15]

Activity is the execution of a task or action by an individual. Participation is someone's involvement in a life situation. The domains of activities and participation cover the full range of life areas, such as "mobility," "self care," and "interpersonal interactions."

Inclusion and exclusion criteria

Studies with the following properties were included:

- The main focus of the study is the development or clinimetric evaluation of a measurement tool described as above.
- The study is written as a full report in English or Dutch.
- The study is published between January 1, 1980 and October 31, 2007.

Measurement tools were included if they were specifically developed or validated for use in persons with a broad range of hand injuries and used in more than one study concerning hand injuries. In addition, they had to include items of activity and/or participation. Both performance tests and questionnaires in the public domain could satisfy these criteria. Measurement tools were excluded if they were developed for use in persons with hand injuries due to central neurological disorders (e.g., stroke, spinal cord injury) and for use in children (age < 18 years) with hand injuries. In addition, measurement tools used for one specific hand condition only (e.g., osteoarthritis, rheumatoid arthritis, carpal tunnel) and tools with unknown clinimetric properties were excluded.

Analysis of measurement tools

Descriptive data extracted from the studies included the: purpose of the measurement tool, target population, number of scales or sub scores, number of items (questionnaires), range of scores, and study populations used to investigate the quality of the measurement tool.

Clinimetric quality of the questionnaires was scored with the "quality criteria for health status questionnaires" described by Terwee et al.[16] and the "checklist for rating clinimetric quality of self-assessment questionnaires" from Bot et al.,[17] which both contain criteria on validity, reproducibility, responsiveness, floor and ceiling effects, and interpretability. These were complemented with pragmatic criteria on applicability described by Auger et al.[18] The same criteria were used to score the quality of the performance tests, except that the questionnaire-specific criterion "readability" was left out. In addition, the criterion "instruction or schooling of examiner needed" was added, because of its importance to performance tests.[19–21]

The first author rated the criteria with the system described in the Appendix.

RESULTS

The search identified 696 potentially relevant publications, of which 63 full text publications were screened. Finally, a total of 44 publications were included, referring to 36 measurement tools on activity and participation level (Figure 1). Using many different search terms was necessary for an exhaustive search. This explains the large number of potentially relevant publications, of which the majority could be excluded on title alone. Of the 36 identified measurement tools, 15 (5 questionnaires and 10 performance tests) met the eligibility criteria. Descriptive data and quality assessment of included measurement tools are given in Tables 1–4. The names of excluded measurement tools and reason for exclusion are listed in Table 5.

Description of measurement tools

Questionnaires

The Disabilities of the Arm, Shoulder and Hand (DASH)[22] and Hand Function Sort (HFS)[23] were designed for upper-extremity disorders, whereas the Michigan Hand Outcomes Questionnaire (MHQ)[24] and Patient Evaluation Measure (PEM)[13] focus on the hand and wrist. The Patient Rated Wrist Hand Evaluation (PRWHE) was originally designed for wrist disorders and formerly known as the Patient Rated Wrist Evaluation (PRWE)[25] but was modified in 2004 for use in wrist and hand disorders.[26] In this paper, we use the name of the modified version, PRWHE, to describe findings on both versions. The HFS is the only questionnaire that contains pictures to clarify the questions and the main focus of which is work performance. The DASH and the MHQ both assess physical health and pain symptoms. Social and emotional health is measured more extensively by the DASH. On the other hand, the MHQ measures the left and right hand and has six separate subscales. The PEM consists of three parts, of which only part two (Hand Health Profile) is specific for hand injuries. The PEM and MHQ are the only tools that contain a specific question or scale on aesthetics. For all the questionnaires, the populations used for the clinimetric studies consisted of a broad range of hand injuries (Table 1).

Performance tests

Five performance tests were developed for all types of hand injuries (Functional Dexteriy Test [FDT][27], Jebsen Taylor Hand Function test [JTHT],[28] NK Hand Dexterity Test [NKHDT],[31] Purdue

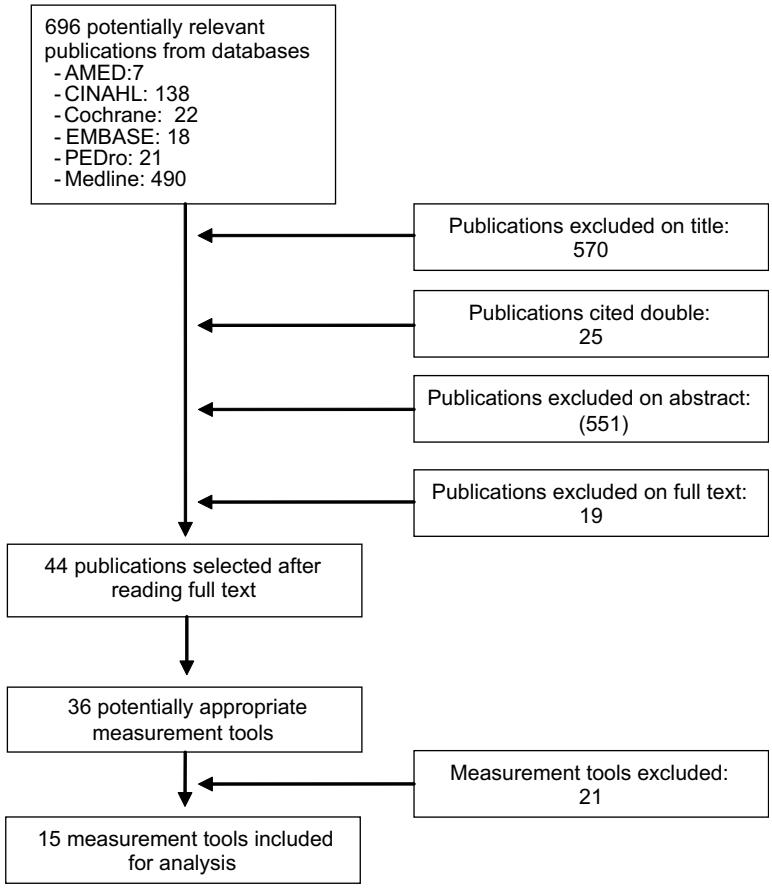

FIGURE 1. *Flowchart search strategy.*

TABLE 1. **Descriptive data of the questionnaires**

Questionnaire	*Abbreviation*	*Purpose*	*Target Population*	*Number of Scales*	*Number of Items*	*Response Options*	*Range of Scores*	*Study Populations*	*References*
Disabilities of the Arm, Shoulder and Hand	DASH	Tto assess symptoms and functional status	Upper-extremity musculoskeletal conditions	1	30	5	0–100	Shoulder and hand/wrist disorders, acute hand/wrist trauma, diverse hand/wrist surgery, OA	21,25,43–48
Hand Function Sort	HFS	To assess hand function focusing on work performance	Upper-extremity musculoskeletal conditions	1	62	5	0–248	Musculoskeletal disorders, acute injuries to hand/wrist/elbow, repetitive strain and overuse injuries	22,49
Michigan Hand Outcomes Questionnaire	MHQ	To assess patient's perception of one or both hands	All types of hand/wrist conditions	6; ADL, pain, work, function, aesthetics, satisfaction	37	5	0–100	Hand/wrist disorders referred for surgery, acute distal radius surgery	23,43,50,51
Patient Evaluation Measure	PEM	to assess outcomes for hand disorders	All types of hand/wrist conditions	3; Opinion on delivery of care, HHP, overall health	18 (total) 10 (HHP)	7	10–126 (total) 10–70 (HHP)	Hand/wrist disorders, e.g., scaphoid and distal radius fracture; CTS	52–55
Patient Rated Wrist Hand Evaluation	PRWHE	To assess wrist/hand related pain and disabilities	All types of wrist/hand conditions	2; Pain, function	15	10	0–100	Hand/wrist disorders, e.g., OA, carpal instabilities; radius and scaphoid fractures	24,25,56

ADL: Activities of Daily Living; HHP: Hand Health Profile; OA: Osteoarthritis; CTS: Carpal Tunnel Syndrome.

TABLE 2. **Descriptive data of the performance tests**

Performance Test	*Abbreviation*	*Purpose*	*Target Population*	*Number of Subscores*	*(Range of) Scores*	*Study Populations*	*References*
Box and Blocks Test	BBT	To assess gross manual dexterity and eye-hand coordination	Physically impaired persons	1	Number of blocks displaced within 60 sec	Adults with mild ID (39–49 yr), healthy adults (22–53 yr)	57,58
Functional Dexterity Test	FDT	To assess the ability to use the hand for functional tasks	Persons with hand injuries	2; Time to complete, total time plus penalty	Movement times (sec) converted into functional score	Hand-injured persons, healthy persons of all ages	26
Grooved Pegboard Test	GPT	To assess manual dexterity & eye-hand coordination	Originally for neurological and cognitive disorders	2; Place task, remove task	Movement times (sec)	Adults with mild ID, healthy persons 7–30 yr	57,59–61
Jebsen-Taylor Hand Function Test	JTHT	To assess hand function regarding ADL	All types of hand injuries	7; ADLs	Movement times (sec) converted into functional score	OA, RA; healthy adults; healthy women freehanded and wearing different types of orthosis	27,62,63
Moberg Pick Up Test	Moberg test	To evaluate functional sensibility of the hand	Specifically suitable for peripheral nerve injuries	2; Eyes open; eyes closed	Movement times (300 sec is max)	Healthy persons (11–89 yr), median nerve injuries	35,64,65
Nine Hole Peg Test	NHPT	To assess finger dexterity	Physically handicapped persons	1	Movement times (sec)	Healthy adults (20–94 yr)	32,53,66
NK Hand Dexterity Test	NKHDT	To assess hand dexterity	All types of (hand) injuries	3; Small, medium, large objects	Movement times (sec)	Healthy adults (18–75 yr); distal radius fracture; variety of hand injuries	28,67,68
Purdue Pegboard Test	PPT	To assess fine manual dexterity	All types of hand injuries	2; One hand, both hands	Movement times (sec)	Traumatic hand injury; healthy adults; healthy students	32,69–71
South Hampton Assessment Procedure	SHAP	To assess pathologic or prosthetic hand function	All type of hand injuries	26; 12 Object tasks, 14 ADLs	Movement times (sec)	Hand prosthesis users, healthy students	30,72
Sollerman Hand Function Test	Sollerman test	To assess functional grip of the hand	Originally for tetraplegia	40; 20 ADLs, for each ability and time to perform	0–80	Median and ulnar nerve injuries	34,73

ADL: Activities of Daily Living; CTS: Carpal Tunnel Syndrome; sec: seconds; ID: Intellectual Disability; OA: Osteoarthritis; RA: Rheumatoid Arthritis.

TABLE 3. **Summary of quality assessment of the questionnaires**

Questionnaire	*Content Validity*	*Construct Validity*	*Internal consistency*	*Floor/ceiling Effect*	*Test–retest Reliability*	*Agreement*	*Responsiveness*	*Interpretability*	*Ease of Scoring*	*Readability*	*Time to Administer*
DASH (Disabilities of the Arm, Shoulder and Hand)	++	+	+	+	+	+	++	+	+	++	++
HFS (Hand Function Sort)	++	+	+	?	?	?	+	–	++	++	++
MHQ (Michigan Hand Outcomes Questionnaire)	++	+	++	?	+	+	+	+	+	++	++
PEM (Patient Evaluation Measure)	?	+	+	?	+	?	+	–	++	++	++
PRWHE (Patient Rated Wrist Hand Evaluation)	++	+	?	?	+	?	+	–	++	++	++

Result or method was rated as: ++ very good; + good; +/– doubtful; – poor;? no information found; n.a.: not applicable.

TABLE 4. **Summary of quality assessment of the performance tests**

Performance Test	*Content Validity*	*Construct Validity*	*Internal Consistency*	*Floor/Ceiling Effect*	*Intrarater Reliability*	*Interrater Reliability*	*Agreement*	*Responsiveness*	*Interpretability*	*Instructions/ Schooling*	*Ease of Scoring*	*Time to Perform*
BBT	?	+	?	?	+	+	?	?	–	++	++	++
FDT	?	+	n.a.	?	+	+	?	?	+	++	++	++
GPT	?	+	n.a.	?	+	?	+	?	–	++	++	++
JTHT	++	+	?	?	–	?	?	?	+	+	++	+
Moberg test	?	–	?	?	?	–	?	+	+	++	++	++
NHPT	?	+	n.a.	?	–	+/–	?	?	–	++	++	++
NK test	**?**	+	**n.a.**	**?**	–	–	**?**	+	+	++	++	–
PPT	?	+	+	?	–	?	?	?	–	++	++	++
SHAP	+	?	–	?	+	+	?	?	+	–	++	–
Sollerman test	?	–	?	?	?	+	?	?	–	++	++	–

Result or method was rated as: ++ very good; + good; +/– doubtful; – poor;? no information found; n.a.: not applicable.

TABLE 5. **Excluded measurement tools**

Measurement Tool	*Reason Exclusion*	*Reference*
Performance tests		
Arthritis Hand Function Test	Developed and used for RA and OA patients	74
Bennet Hand Tool Dexterity Test	Unknown clinimetric properties	75
Crawford Small Parts Dexterity Test	Unknown clinimetric properties	75
Greenseid & McCormack Test	Unknown clinimetric properties	76
Minnesota Rate of Manipulation	Unknown clinimetric properties	77
Motor Assessment Scale	Developed and validated for stroke patients	78
O'Neill Hand Function Assessment	Used in only one (validation) study	79
Sequential Occupational Dexterity Assessment	Validated and used for RA patients only	80
Work Environment Scale	Not developed and validated for hand injuries	81
Questionnaires		
AUSCAN osteoarthritis hand index	Developed and validated for RA and OA patients only	82
Boston Carpal Tunnel Questionnaire	Developed and validated for CTS patients	83
Groningen Activity Restriction Scale	Generic instrument	84
Hand Clinic Questionnaire	Used in only one study (validation study)	55
Health Assessment Questionnaire	Developed for RA patients; not specific for hand injuries	85
Patient-Rated Tennis Elbow Evaluation (revised version of Patient-Rated Forearm Questionnaire)	Developed and validated for LE patients (both versions)	86
Short Form-36 Health Survey	Generic instrument, overall HRQoL	36
Test d'Evaluation des Membres Supérieurs de personnes Agées	Generic instrument, developed for persons >60 yr of age	87
Upper-Extremity Functional Scale	Developed and used for a workers' compensation population	88
Semistructured interviews		
Adaptation Interview Schedule	Not validated; used in only one study	89
Canadian Occupational Performance Measure	Generic instrument	90
Goal Attainment Scaling	Generic instrument	91

RA: Rheumatoid Arthritis; OA: Osteoarthritis; CTS: Carpal Tunnel Syndrome; LE: Lateral Epicondylitis; HRQoL: Health-Related Quality of Life.

Pegboard test [PPT],[29] South Hampton Assessment Procedure [SHAP][30]), whereas two others were designed for physically impaired persons in general (Box and Blocks Test [BBT],[31] Nine Hole Peg Test [NHPT][32]). The Grooved Pegboard Test (GPT)[33] was developed as part of a neuropsychological test battery. The Sollerman Hand Function Test (Sollerman test)[34] was originally developed for tetraplegia and the Moberg Pick Up Test (Moberg test)[35] for peripheral nerve injuries (Table 2).

The JTHT, SHAP, and Sollerman test consist of tasks concerning different activities of daily living (ADL). The NKHDT is a computerized test that requires the manipulation of different objects. The other tests consist of a pegboard with pegs (FDT, GPT, PPT, NHPT) or a box with blocks (BBT). The objects, pegs, or blocks have to be placed and removed in the board or box as quickly as possible. The Moberg test, which is designed to assess functional sensation, is an exception. In this test 12 small common items (e.g. coin, key) have to be placed and removed. The test is divided in an eyes-open and an eyes-shut trial. All but one test use performance time or number of items placed within a certain time as test outcome. In the Sollerman test, the examiner scores the patient's ability and time needed to perform the test. This test contains bilateral and unilateral tasks, whereas the others contain only unilateral tasks. All tests are standardized regarding performance and scoring and are commercially available. Specific instructions to construct the test are also available. The populations used for the clinimetric studies included different types of hand injuries. Healthy subjects were used as well, especially to develop normative data on scores.

Summary of the quality assessment

Questionnaires

Content validity was good for all questionnaires, except for the PEM. Construct validity was studied for all questionnaires by correlating their scores with scores on other health questionnaires measuring the same concept and with impairments in ROM and grip strength. Correlation between the DASH, joint-specific questionnaires, and ability to work were high. Total DASH scores correlated moderately with several subscales of the 36-item short-form health survey (SF-36),[36] a generic health status questionnaire. Scales of the MHQ correlated moderately with related questions on the SF-12,[37] a short version of the SF-36, as was expected. Correlations between the PRWHE and related scales of the SF-36 and an overall impairment score were respectively high and moderate. The HFS, PEM, and PRWHE all correlated strongly with DASH scores. A weak correlation was found between the PEM and impairments scores,

whereas correlations between the DASH and impairment scores were not significant or low (Table 3).

Information on internal consistency was found on four questionnaires. Factor analysis and a high Cronbach's alpha were reported for the MHQ. For the HFS only factor analysis was reported, while for the PEM and DASH Cronbach's alpha was reported and rated as adequate.

Information on floor and ceiling effects was found only for the DASH, which showed no such effect. The DASH and MHQ scored highest on reproducibility, with good reliability and adequate methods used to calculate agreement (Standard Error of Measurement [SEM] for the DASH; Limits of Agreement [LoA] for the MHQ). The HFS, PEM, and PRWHE had good reliability. No data on agreement were found for these tests.

Responsiveness of the HFS, MHQ, PEM, and PRWHE was good, with information on Standardized Response Means (SRMs) and effect size (ES) reported. Responsiveness of the DASH was high; SRMs, ES, and area under the curve (AUC) were reported. However, for none of the questionnaires the minimal important change (MIC) was presented. The responsiveness of the DASH was compared with the other questionnaires in some studies. The DASH scored slightly better than the HFS, while the opposite was true for the PRWHE and PEM. The DASH and MHQ were evenly responsive.

Interpretability of the DASH and MHQ was shown by presentation of mean scores and SDs before and after treatment, the ability to discriminate between different groups, and patient global ratings of change. For the HFS and PEM, only one type of information was reported.

Applicability was high for all the questionnaires. However, ease of scoring was good for the DASH and MHQ, while the others were rated as very good on this criterion.

Overall, the DASH received the best ratings for its clinimetric quality with 15 positive scores, closely followed by the MHQ with 14 positive scores and no information found on one criterion.

Performance tests

Information on content validity was found only for the JTHT and the SHAP. Construct validity of the Moberg and Sollerman tests was not studied. Correlations between the BBT and GPT were high, while the FDT, JTHT, MHPT, and PPT correlated moderately. The correlation between JTHT and NKHDT was moderate to strong. The FDT, JTHT, and PPT also showed moderate correlations with ADL. Correlations for the Moberg test and two-point discrimination were low, which was not expected. The Sollerman test correlated strongly with disability rating tables, while the JTHT and NKHDT correlated moderately with patient-rated function (Table 4).

Internal consistency was inappropriate for the SHAP. For the PPT, factor analysis was conducted. Information on this criterion was not found or was not applicable for the other tests.

Intra- and interrater reliability were good for the BBT, FDT, and SHAP. Intrarater reliability was presented for the GPT, JTHT, PPT NHPT, and NKHDT and found to be good for the first and negative for the other three tests. For the JTHT and NHPT, a Pearson's correlation coefficient < 0.80 was reported, while for the NKHDT and PPT an intraclass correlation coefficient (ICC) < 0.70 was found. However, in another study the mean of three trials showed an ICC > 0.70 for the PPT.

Good interrater reliability was found for the Sollerman test, while the NKHDT showed an ICC < 0.7. The Moberg test showed a PCC < 0.80 and the NHPT had PCC > 0.80. The reliability of the NKHDT tended to be higher for the dominant hand than for the non-dominant hand, with ICC > 0.7 for some subtests.

Agreement was reported only for the GPT. Responsiveness was reported only for the Moberg test and the NKHDT. Both had adequate SRMs and ES, while for the Moberg test adequate ES was also reported. Two types of information on interpretability, mean scores, and SDs and the ability to discriminate between groups, were reported for the FDT, JTHT, SHAP, and Moberg test. Normative values and correlation with a functional measure were reported for the NKHDT. For the other tests only one type of information was found, mostly normative scores in healthy subjects.

Six out of nine tests received the highest possible rating for applicability. The JTHT was rated a little lower, because the examiner needs to study instructions and the mean time to perform was 15 min. This was up to 20 min for the NKHDT, the SHAP, and the Sollerman test.

Overall, the FDT received the highest ratings for clinimetric quality with ten positive scores and no information found on four criteria. This was followed by the BBT and GPT. The PPT and Moberg test scored high on applicability, but few studies were found on their clinimetric properties. Compared with the questionnaires, clinimetric quality was studied less extensively for the performance tests. In addition, five performance tests received negative ratings on reproducibility and validity, while none of the questionnaires received negative ratings on these criteria.

DISCUSSION

In this paper the content and clinimetric quality of measurement tools assessing activities and participation in patients with hand injuries were systematically reviewed. As a result five questionnaires and

ten performance tests were identified and their clinimetric quality was evaluated. Overall, the DASH was the most extensively studied tool and had the most supporting evidence for good clinimetric quality, with positive results for all criteria, closely followed by the MHQ. These are both questionnaires. The DASH also received the best ratings for its clinimetric properties in the study by Bot et al.,[17] who evaluated clinimetric properties of shoulder disability questionnaires. Of all the performance tests, the FDT received the best ratings. Applicability for daily use was good for most of the included tests. Information on clinimetric properties was often lacking, especially for the performance tests. Moreover, most of the included performance tests used time as the outcome measure and did not contain bilateral tasks. One can argue whether these tests really reflect activity levels in someone's daily life. Therefore, a number of these tools seem to lack face validity.

More studies were found and better clinimetric quality was reported for the questionnaires. However, it must be mentioned that quality scores of questionnaires and performance tests cannot be compared without caution. There are considerable differences between these two types of tools. Performance tests are objective tools that focus on capacity, that is, what a person can do in a standardized environment. Questionnaires are subjective tools that focus on performance, which is a person's perception of what he or she actually does in a real life environment.[38,39] According to these definitions, performance tests assess activities, while questionnaires can assess activities as well as participation. Questionnaires included in this review all contained items on both activities and participation. Because of these differences, the two types of tools should complement each other when measuring someone's activity and participation level.[39,40]

We found low correlations between impairments and measures of activity and participation, which is supported by other authors.[41,42] This emphasizes that the ICF domains are related, but separate, entities.

Optimal measurement tool selection depends highly on the situation, the purpose of assessment, and the type of hand injury. Aesthetics, an item belonging to the ICF-domain "personal factors," is considered an important item for hand-injured persons. The MHQ measures aesthetics, has the ability to assess the left and right hand separately, and contains questions relating specifically to the hand. Therefore, the MHQ may be preferred to the DASH in specific cases.[43] The DASH is suitable for any type of upper-extremity injury and is widely accepted, promoting the exchangeability of results in clinical evaluation and research.

Although there is no consensus on the best way to determine the quality of measurement tools, by using the rating system suggested by Terwee et al.,[16] it was possible to provide an overview of scores on different aspects of clinimetric quality. Terwee et al. mentioned that their criteria should be refined and complemented further by other investigators. In this review, we added criteria on applicability in order to assess the tool's usefulness in daily clinical practice. In addition, we refined some criteria to differentiate between "good" and "very good."

The authors found it difficult to rate construct validity and interpretability. Many studies failed to specify hypotheses concerning construct validity or the hypotheses were presented only in the result section. In these cases we choose to rate the criterion as "good," because the availability of this information was considered as "better" than no information at all.

There are limitations to this paper that must be considered. There are no established or recommended search strategies for locating outcome measures, and a large number of search terms were necessary for an exhaustive search. The search strategy might not be reproducible, despite our efforts to describe the process in detail.

A major limitation of this study is that there was only one reviewer. Therefore, bias might have occurred in the inclusion and rating process. Some relevant publications may have been missed, which might have affected the systematic nature of data extraction and synthesis so typical for systematic reviews. However, the reviewer's findings during the inclusion and data extraction process were discussed with the other authors, so that consensus on the interpretation and synthesis of the data was achieved.

Disease-specific tools were not included in this review, because the authors wanted to provide an overview of tools that are used for any type of hand injury, not just for one specific condition. Some excluded tools such as the AUSCAN osteoarthritis hand index, the Boston Carpal Tunnel Questionnaire, the Sequential Occupational Dexterity Test (SODA), and the Canadian Occupational Performance Measure have not been validated for hand injuries. For example, the SODA is one of the few performance tests that measures ADL, including bilateral tasks, but has been validated for rheumatoid arthritis only. Because several of these tools specifically measure activities and participation, validation is recommended for their use in the assessment of hand-injured persons. Tests that contain bilateral ADL tasks and generic tools that focus on activities and participation should be targeted in future research. In addition, how to define and rate difficult concepts such as construct validity and interpretability needs to be discussed and studied more.

CONCLUSIONS

This systematic review revealed a need for further research on the clinimetric properties of hand injury

measurement tools, especially on performance tests. Overall, the DASH was the most extensively studied tool with positive ratings for all criteria, closely followed by the MHQ. Of the performance tests, the FDT received the best ratings.

A number of measurement tools with strong clinimetric properties address activities and participation in hand-injured persons. Optimal measurement tool selection depends highly on the purpose of measurement and the type of hand injury.

REFERENCES

1. Muenzen PM, Kasch MC, Greenberg S, Fullenwider L, Taylor PA, Dimick MP. A new practice analysis of hand therapy. J Hand Ther. 2002 Jul;15:215–25.
2. Roth LP, Dimick MP, Kasch MC, Fullenwider L, Mullins P. Practice analysis of hand therapy. J Hand Ther. 1996 Jul;9: 203–12.
3. Brockow T, Cieza A, Kuhlow H, et al. Identifying the concepts contained in outcome measures of clinical trials on musculoskeletal disorders and chronic widespread pain using the International Classification of Functioning, Disability and Health as a reference. J Rehabil Med. (44 Suppl)2004 Jul:30–6.
4. Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the international classification of functioning, disability and health. J Rehabil Med. 2002 Sept;34:205–10.
5. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. Disabil Rehabil. 2005 Mar 18;27:315–40.
6. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF participation. Disabil Rehabil. 2005 May 6;27:507–28.
7. Chan J, Spencer J. Adaptation to hand injury: an evolving experience. Am J Occup Ther. 2004 Mar;58:128–39.
8. Keller RB, Rudicel SA, Liang MH. Outcomes research in orthopaedics. Instr Course Lect. 1994;43:599–611.
9. Hand Therapy Certification Commission. Definition of hand therapy and scope of practice of certified hand therapists; 2002.
10. MacDermid JC, Fess EE, Bell-Krotoski J, Cannon NM, Evans RB, Walsh W, et al. A research agenda for hand therapy. J Hand Ther. 2002 Jan;15:3–15.
11. Gummesson C, Atroshi I, Ekdahl C. The quality of reporting and outcome measures in randomized clinical trials related to upper-extremity disorders. J Hand Surg [Am]. 2004 Jul;29: 727–34.
12. Jerosch-Herold C, Leite JC, Song F. A systematic review of outcomes assessed in randomized controlled trials of surgical interventions for carpal tunnel syndrome using the International Classification of Functioning, Disability and Health (ICF) as a reference tool. BMC Musculoskelet Disord. 2006; 7:96.
13. Macey AC, Burke FD, Abbott K, Barton NJ, Bradbury E, Bradley A, et al. Outcomes of hand surgery British Society for Surgery of the HandJ Hand Surg [Br]. 1995 Dec;20:841–55.
14. Bindra RR, Dias JJ, Heras-Palau C, Amadio PC, Chung KC, Burke FD. Assessing outcome after hand surgery: the current state. J Hand Surg [Br]. 2003 Aug;28:289–94.
15. World Health Organisation. 2002. International classification of functioning, disability and health. Geneva, Switzerland; 2002.
16. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60:34–42.
17. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis. 2004 Apr;63:335–41.
18. Auger C, Demers L, Swaine B. Making sense of pragmatic criteria for the selection of geriatric rehabilitation measurement tools. Arch Gerontol Geriatr. 2006 Jul;43:65–83.
19. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess. 1998;2:i-74.
20. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. Disabil Rehabil. 2005 Mar 18;27:315–40.
21. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) The Upper Extremity Collaborative Group (UECG)Am J Ind Med. 1996 Jun;29:602–8.
22. Matheson LN, Kaskutas VK, Mada D. Development and construct validation of the Hand Function Sort. J Occup Rehabil. 2001 Jun;11:75–86.
23. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg [Am]. 1998 Jul;23:575–87.
24. MacDermid JC. Development of a scale for patient rating of wrist pain and disability. J Hand Ther. 1996 Apr;9:178–83.
25. MacDermid JC, Tottenham V. Responsiveness of the disability of the arm, shoulder, and hand (DASH) and patient-rated wrist/hand evaluation (PRWHE) in evaluating change after hand therapy. J Hand Ther. 2004 Jan;17:18–23.
26. Aaron DH, Jansen CW. Development of the Functional Dexterity Test (FDT): construction, validity, reliability, and normative data. J Hand Ther. 2003 Jan;16:12–21.
27. Sharma S, Schumacher HR, McLellan AT. Evaluation of the Jebsen hand function test for use in patients with rheumatoid arthritis. Arthritis Care Res. 1994 Mar;7:16–9.
28. Turgeon TR, MacDermid JC, Roth JH. Reliability of the NK dexterity test. J Hand Ther. 1999 Jan;12:7–15.
29. Tiftin J. Purdue Pegboard emaniner's manual. Chicago: Science Research Associates., 1987.
30. Light CM, Chappell PH, Kyberd PJ. Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: normative data, reliability, and validity. Arch Phys Med Rehabil. 2002 Jun;83:776–83.
31. Desrosiers J, Bravo G, Hebert R, Dutil E, Mercier L. Validation of the box and block test as a measure of dexterity of elderly people: reliability, validity, and norms studies 26. Arch Phys Med Rehabil. 1994 Jul;75:751–5.
32. Mathiowetz V, Katz JN, Volland G. Adult norms for the nine hole peg test of finger dexterity. Occup Ther J Res. 1985;5: 25–37.
33. Schmidt SL, Oliveira RM, Rocha FR, breu-Villaca Y. Influences of handedness and gender on the grooved pegboard test. Brain Cogn. 2000 Dec;44:445–54.
34. Sollerman C, Ejeskar A. Sollerman hand function test. A standardised method and its use in tetraplegic patients. Scand J Plast Reconstr Surg Hand Surg. 1995 Jun;29:167–76.
35. Ng CL, Ho DD, Chow SP. The Moberg pickup test: results of testing with a standard protocol. J Hand Ther. 1999 Oct;12: 309–12.
36. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 Jun;30:473–83.
37. Ware JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34:220–33.
38. Ustun TB, Chatterji S, Bickenbach J, Kostanjsek N, Schneider M. The International Classification of Functioning, Disability and Health: a new tool for understanding disability and health. Disabil Rehabil. 2003 Jun 3;25:565–71.
39. Wittink H. Functional capacity testing in patients with chronic pain. Clin J Pain. 2005 May;21:197–9.
40. Wittink H, Rogers W, Sukiennik A, Carr DB. Physical functioning: self-report and performance measures are related but distinct. Spine. 2003 Oct 15;28:2407–13.
41. Roddey TS, Cook KF, O'Malley KJ, Gartsman GM. The relationship among strength and mobility measures and self-report outcome scores in persons after rotator cuff repair

surgery: impairment measures are not enough. J Shoulder Elbow Surg. 2005 Jan;14(1 Suppl S):95S–8S.
42. van Oosterom FJ, Ettema AM, Mulder PG, Hovius SE. Impairment and disability after severe hand injuries with multiple phalangeal fractures. J Hand Surg [Am]. 2007 Jan;32: 91–5.
43. Kotsis SV, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and the Disabilities of the Arm, Shoulder and Hand questionnaire in carpal tunnel surgery. J Hand Surg [Am]. 2005 Jan;30:81–6.
44. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. J Hand Ther. 2001 Apr;14:128–46.
45. Wong JY, Fung BK, Chu MM, Chan RK. The use of Disabilities of the Arm, Shoulder, and Hand Questionnaire in rehabilitation after acute traumatic hand injuries. J Hand Ther. 2007 Jan;20:49–55.
46. Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. BMC Musculoskelet Disord. 2003 Jun 16;4:11.
47. McClure PW, Michener L. Measures of adult shoulder function. Arthritis Rheum. 2003;49:50–8.
48. Soohoo NF, McDonald AP, Seiler JG III, McGillivary GR. Evaluation of the construct validity of the DASH questionnaire by correlation to the SF-36. J Hand Surg [Am]. 2002 May;27: 537–41.
49. Marshall JR, Kingston SP. Comparative validity and responsiveness of the hand function sort in evaluating change in disability and impairment in hand therapy. J Hand Ther [Am]. 2006;19:452–3.
50. Chung KC, Hamill JB, Walters MR, Hayward RA. The Michigan Hand Outcomes Questionnaire (MHQ): assessment of responsiveness to clinical change. Ann Plast Surg. 1999 Jun;42: 619–22.
51. Kotsis SC, Lau FH, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and physical measurements in outcome studies of distal radius fracture treatment. J Hand Surg [Am]. 2007 Jan;32:84–90.
52. Forward DP, Sithole JS, Davis TR. The internal consistency and validity of the patient evaluation measure for outcomes assessment in distal radius fractures. J Hand Surg Eur. 2007 Jun;32: 262–7.
53. Hobby JL, Watts C, Elliot D. Validity and responsiveness of the patient evaluation measure as an outcome measure for carpal tunnel syndrome. J Hand Surg [Br]. 2005 Aug;30:350–4.
54. Dias JJ, Bhowal B, Wildin CJ, Thompson JR. Assessing the outcome of disorders of the hand. Is the patient evaluation measure reliable, valid, responsive and without bias? J Bone Joint Surg Br. 2001 Mar;83:235–40.
55. Sharma R, Dias JJ. Validity and reliability of three generic outcome measures for hand disorders. J Hand Surg [Br]. 2000 Dec; 25:593–600.
56. MacDermid JC, Turgeon TR, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. Osteoarthr Cartil. 1998;12:577–86.
57. Carmeli E, Bar-Yossef T, Ariav C, Levy R, Lieberman DG. Perceptual-motor coordination in persons with mild intellectual disability. Disabil Rehabil. 2007;1:1–7.
58. Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the box and block test of manual dexterity. Am J Occup Ther. 1985 Jun;39:386–91.
59. Bryden PJ, Roy EA. A new method of administering the Grooved Pegboard Test: performance as a function of handedness and sex. Brain Cogn. 2005 Aug;58:258–68.
60. Ruff RM, Parker SB. Gender- and age-specific changes in motor speed and eye-hand coordination in adults: normative values for the Finger Tapping and Grooved Pegboard Tests. Percept Mot Skills. 1993 Jun;76(3 Pt 2):1219–30.
61. Strenge H, Niederberger U, Seelhorst U. Correlation between tests of attention and performance on grooved and Purdue pegboards in normal subjects. Percept Mot Skills. 2002 Oct; 95:507–14.
62. Stern EB. Stability of the Jebsen-Taylor hand function test across three test sessions. Am J Occup Ther. 1992 July;46:647–9.
63. Labi ML, Gresham GE, Rathey UK. Hand function in osteoarthritis. Arch Phys Med Rehabil. 1982 Sept;63:438–40.
64. Amirjani N, Ashworth NL, Gordon T, Edwards DC, Chan KM. Normative values and the effects of age, gender, and handedness on the Moberg Pick-Up Test. Muscle Nerve. 2007 Jun;35: 788–92.
65. Wong KH, Coert JH, Robinson PH, Meek MF. Comparison of assessment tools to score recovery of function after repair of traumatic lesions of the median nerve. Scand J Plast Reconstr Surg Hand Surg. 2006;40:219–24.
66. Grice KO, Vogel KA, Le V, Mitchell A, Muniz S, Vollmer MA. Adult norms for a commercially available nine hole peg test for finger dexterity. Am J Occup Ther. 2003;57:570–3.
67. MacDermid JC, Mulè M. Concurrent validity of the NK hand dexterity test. Physiother Res Int. 2001;6:83–93.
68. Amadio PC, Silverstein MD, Ilstrup DM, Scleck CD, Jensen LM. Outcome after colles fracture: the relative responsiveness of three questionnaires and physical examination measures. J Hand Surg. 1996;21A:781–7.
69. Shahar RB, Kizony R, Nota A. Validity of the Purdue Pegboard Test in assessing patients after traumatic hand injury. Work. 1998;11:315–20.
70. Buddenberg LA, Davis C. Test-retest reliability of the Purdue Pegboard Test. Am J Occup Ther. 2000 Sept;54(5):555–8.
71. Mathiowetz V, Rogers SL, Dowe-Keval M, Donahoe L, Rennells C. The Purdue Pegboard: norms for 14- to 19-year-olds. Am J Occup Ther. 1986 Mar;40:174–9.
72. Adams CI. A comparison of dominant and non-dominant hand function in both right- and left-handed individuals using the South Hampton Hand Assessment Procedure (SHAP). J Hand Ther [Br]. 2003;8:4–10.
73. Rosen B, Dahlin LB, Lundborg G. Assessment of functional outcome after nerve repair in a longitudinal cohort. Scand J Plast Reconstr Surg Hand Surg. 2000 Mar;34:71–8.
74. Backman C, Mackie H, Harris J. Arthritis hand function test: development of a standardized assessment tool. Occup Ther J Res. 1991;11:245–55.
75. Bear-Lehman J, Abreu BC. Evaluating the hand: issues in reliability and validity. Phys Ther. 1989 Dec;69:1025–33.
76. Greenseid DZ, McCormack RM. Functional hand testing: a profile evaluation. Plast Reconstr Surg. 1968 Dec;42:567–71.
77. Lafayette Instrument Company. The complete Minnesota Dexterity Test Examiner's Manual. Lafayette, Indiana. 1991.
78. Carr JH, Shepherd RB, Nordholm L, Lynne D. Investigation of a new motor assessment scale for stroke patients. Phys Ther. 1985 Feb;65:175–80.
79. O'Neill G. The development of a standardised assessment of hand function. Br J Occup Ther. 1995;58:477–80.
80. van LW, van't Pad BP, Bakker J, Terwindt S, Franssen M, van RP. Sequential occupational dexterity assessment (SODA): a new test to measure hand disability. J Hand Ther. 1996 Jan;9: 27–32.
81. Rossberg JI, Eiring O, Friis S. Work environment and job satisfaction. A psychometric evaluation of the Working Environment Scale-10. Soc Psychiatry Psychiatr Epidemiol. 2004 Jul; 39:576–80.
82. Bellamy N, Campbell J, Haraoui B, Gerecz-Simon E, Buchbinder R, Hobby K, et al. Clinimetric properties of the AUSCAN Osteoarthritis Hand Index: an evaluation of reliability, validity and responsiveness. Osteoarthr Cartil. 2002 Nov;10: 863–9.
83. Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. J Bone Joint Surg Am. 1993 Nov;75:1585–92.
84. Suurmeijer TP, Doeglas DM, Moum T, Briancon S, Krol B, Sanderman R, et al. The Groningen activity restriction scale for measuring disability: its utility in international comparisons. Am J Public Health. 1994 Aug;84:1270–3.

85. Fries JF. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980;23:137–45.
86. Overend TJ, Wuori-Fearn JL, Kramer JF, MacDermid JC. Reliability of a patient-rated forearm evaluation questionnaire for patients with lateral epicondylitis. J Hand Ther. 1999 Jan;12: 31–7.
87. Desrosiers J, Hebert R, Dutil E, Bravo G. Development and reliability of an upper extremity function test for the elderly: the TEMPA. Can J Occup Ther. 1993;60:9–16.
88. Pransky G, Feuerstein M, Himmelstein J, Katz JN, Vickers-Lahti M. Measuring functional outcomes in work-related upper extremity disorders. Development and validation of the upper extremity function scale. J Occup Environ Med. 1997 Dec;39:1195–202.
89. Chan J, Spencer J. Contrasting perspectives on pain following hand injury. J Hand Ther. 2005 Oct;18:429–35.
90. Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian occupational performance measure: an outcome measure for occupational therapy. Can J Occup Ther. 1990 Apr;57:82–7.
91. Malec JF. Goal attainment scaling in rehabilitation. Neuropsychol Rehabil. 1999;9:253–75.
92. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. New York: Oxford University Press, 2003.
93. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials. 1991 Aug;12(4 Suppl):142S–58S.

Appendix

CRITERIA FOR CLINIMETRIC QUALITY RATING

Validity

Validity is the degree to which an instrument measures what it is supposed to measure. The measurement tools were rated for content and construct validity. Internal consistency, which is a measure of homogeneity of a (sub)scale, was also rated.

Content validity

Content validity is the extent to which the domain of interest is comprehensively sampled by the items of the measurement tool. The items must reflect areas that are important to patients with hand injuries.

++ = patients and experts were involved during item selection and item reduction.
+ = only patients or only experts were involved during item selection and item reduction.
− = no patient and expert involvement.? = no information found on content validity.

Construct validity

Construct validity is the extent to which scores on the measurement tool relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning concepts that are being measured.

+ = specific hypotheses were formulated and ≥75% of the results confirm these hypotheses.
− = specific hypotheses were formulated but <75% of the results confirm these hypotheses or doubtful method used.
? = no information found on construct validity.

Internal consistency

Internal consistency is the extent to which items in a (sub)scale are correlated, thus measuring the same concept. This is important only for measurement tools that intend to measure a single concept by using multiple items. For measurement tools in which the items are different aspects of a complex clinical phenomenon that do not have to be correlated, internal consistency is not relevant.

++ = factor analysis *and* Cronbach's alpha were applied for each dimension separately *and* scored between 0.70 and 0.90.
+ = Cronbach's alpha was reported only for the total of the questionnaire and alpha scored between 0.70 and 0.90 or only factor analysis was reported.
− = no factor analysis and Cronbach's alpha were applied or Cronbach's alpha scored <0.70.
? = no information found on internal consistency.

Reproducibility

Reproducibility concerns the degree to which repeated measures in stable persons provide similar results. For the questionnaires this was assessed by rating test−retest reliability and agreement. For the performance tests inter- and intrarater reliability and agreement were used.

Reliability

Reliability is the extent to which persons can be distinguished from each other despite measurements errors. The ICC or weighted Kappa (Kw) for ordinal measures are considered as adequate measures for reliability.[16,92] Application of Pearson's correlation coefficient (R) was rated as doubtful, as it neglects systematic errors if present.[17,93]

+ = ICC or Kw was used and ICC/Kw ≥ 0.7.
+/− = doubtful method was used; for example $R \geq 0.80$.
− = ICC or Kw was used and scored <0.7 or $R < 0.80$.
? = no information found on reliability.

Agreement

Agreement is the extent to which scores on repeated measures are close to each other (absolute measurement error). This is a relatively new approach and there is little consensus on adequate cutoff points for agreement. Therefore, a positive rating was given when an adequate method for agreement was used, and authors provided convincing arguments that agreement was acceptable.+ = LoA, Kappa coefficient, or standard error measurement (SEM) were presented and convincing arguments that agreement was acceptable.− = inadequate agreement or doubtful method used.? = no information found on agreement.

Responsiveness

Responsiveness is the ability of a measurement tool to detect clinically important changes over time in the concept being measured. There is no single agreed method to assess responsiveness. In this review, calculations of ES, SRM, and comparison of change scores with an external standard were considered as adequate measures for responsiveness. Furthermore, if information was found on the ability to distinguish MIC from measurement error, responsiveness was considered as very good. Adequate measures to determine this were relating the smallest detectable change (SDC) or LoA to the MIC, Guyatt's responsiveness ratio (RR) or the area under the receiver operating characteristics (ROC) curve.

++ = MIC > SDC or MIC outside the LoA or RR > 1.96 or AUC ≥ 0.70.
+ = ES > 0.8 or SRM or comparison with external standard (t-test; $p < 0.05$).
− = MIC ≤ SDC or MIC inside the LoA or RR ≤ 1.96 or AUC < 0.70 or ES < 0.8.
? = no information found on responsiveness.

Floor and ceiling effects

Floor and ceiling effects are considered to be present if more than 15% of respondents achieved the lowest or highest possible score, which means that extremes are missing in the lower and upper end of the scale or score.

+ = ≤15% of respondents achieved the highest or lowest possible scores.
− = >15% of respondents achieved the highest or lowest possible score.
? = no information found on floor and ceiling effects.

Interpretability

Interpretability is the degree to which one can assign qualitative meaning to quantitative scores. The following types of information can aid in interpreting scores: 1) means and SD of scores before and after treatment; 2) comparative data on the distribution of scores in relevant subgroups; 3) relationships of scores with well-known functional measures or clinical diagnosis; 4) association between changes in scores and patients' global ratings of their experienced magnitude of change.[17]

+ = two types of information as mentioned above were presented.
− = one type of information mentioned above was presented or doubtful method was used.
? = no information found on interpretability.

Applicability

A measurement tool should be suitable for daily clinical use and/or for research purposes. Important items concerning applicability are examiner burden and respondent burden, which we choose to rate as described below:

Examiner burden consists of two subscores:

A. the necessity for instructions or schooling

++ = no schooling needed.
+ = the examiner needs to study one page of instructions.
− = the examiner needs to study more than one page or needs special schooling.
? = no information found on instructions or schooling.

B. Ease of scoring

++ = easy; the items were simply summed.
+ = moderate; a Visual analogue scale (VAS) or a simple formula was used.
− = difficult; VAS and a simple formula or a difficult formula was used.
? = no information found on scoring.

Respondent burden consists of two subscores:

A. Readability (only applicable for questionnaires)

++ = patients tested the readability in a pilot study with good results.
+ = patients tested the readability with poor results.
− = patients were not involved.? = no information found on readability.

B. Time needed to complete questionnaire or performance test

++ = can be completed within 10 min.
+ = takes 10−15 min to complete.
− = takes over 15 min to complete.
? = no information found on time needed.