Summary

In this thesis, research to measurement qualities of Functional Capacity Evaluations (FCEs) is described. FCE’s are performance based tests to measure capacity of activities in a standardized environment. FCEs are used to identify individual’s capacity in respect to their workload and are used in the field of rehabilitation, occupational and insurance medicine. In these fields, the model of work load and work capacity is frequently used to make decisions concerning the functional capacity of a patient. FCEs may play an important role to operationally define Work Capacity of patients in these cases. This thesis aims at improvement of measurement qualities of FCE. These measurement qualities can be divided in the following issues: safety, reliability, validity, practicality and utility. In order to meet these objectives, 7 Studies are performed which are described in chapters 2 to 8.

In Chapter 1, the objectives of this thesis are described. Additionally, a theoretical framework for this thesis is provided.

In Chapter 2, a study is described which focuses on content validity of FCE. Inconsistent terminology and the absence of a conceptual framework within the FCE literature are addressed. Worldwide, 22 experts participated in a Delphi study. Semi and full structured questionnaires are used in three rounds to gain consensus considering 39 statements concerning operational definitions and conceptual framework. Consensus of definitions is considered when 75% or more of all experts agree with a definition. Consensus is met in operational definitions in 10 out of 19 statements. With respect to conceptual framework, consensus is met in 9 out of 20 statements. Experts consented to use the International Classification of Functioning, Disability and Health (ICF) and to use the terms as specified within the ICF. No consensus is reached on a definition of Functional Capacity Evaluation. For future research, it is recommended that researchers use the ICF as a conceptual framework and to provide a definition of FCE.

The research described in Chapter 3 is described to design an FCE protocol for patients with Work Related Upper Limb Disorders (WRULD FCE) and to provide evidence for content validity. A review to epidemiological literature is conducted to identify physical risk factors for WRULD. The results indicate that physical risk factors are related to repetition, duration, working in awkward and static positions and forceful movements of the upper extremity and neck. An FCE is designed based on the risk factors identified. Eight tests are selected to cover all risk factors: the overhead lift, overhead work, repetitive reaching, handgrip strength, finger strength, wrist extension strength, fingertip dexterity, and a hand and forearm dexterity test. Content validity of this FCE was established by providing the rationale, specific objectives and operational definitions of the FCE.
In Chapter 4, a study is presented that researched the test-retest reliability of the WRULD FCE protocol described in Chapter 3 of this thesis. Thirty three healthy adults performed the WRULD FCE, consisting of 26 items, twice with a 10-day interval. Means, 95% confidence intervals, one-way random Intraclass Correlation Coefficients (ICCs) and 95% limits of agreement are calculated. The results show that 14 of 26 items have excellent reliability, 9 of 26 items have good reliability and 3 of 26 items have moderate reliability. Significant learning effects are present in the Purdue Pegboard Task and in the Complete Minnesota Dexterity Test. It is concluded that the WRULD FCE is reliable on test-retest reliability in healthy adults.

The study described in Chapter 5 is performed to quantify the intensity, duration, location and nature of the pain response following an FCE in healthy subjects and to compare this pain response with the pain response of patients with Chronic Low Back Pain (CLBP). A total of 197 healthy working subjects (102 men, 95 women) performed a 12-item FCE. Pain response is measured by a self-constructed Pain Response Questionnaire (PRQ). Descriptive statistics are used to describe the pain response following an FCE. The results are that 82% of all subjects reported a pain response following the FCE. The intensity of the pain response after 24 h post FCE is a median of 3.0 on a numeric rating scale (0-10). About 78% of all pain is reducible to muscle soreness. Pain is most often reported in the upper legs (51%), the lower back (38%), the shoulders (37%) and upper arms (36%). Symptoms decreased to pre-FCE levels in a mean of 3 days. The pain response of 2 subjects (1%) lasted for 3 weeks. The intensity and duration of the pain response of healthy subjects was not significantly different from the response of patients with CLBP. The conclusion of this study was that the pain response of 99% of all subjects was a normal musculoskeletal reaction and can be expected.

The aim of Chapter 6 was to study whether personal characteristics and performance measures are predictive for onset, intensity and duration of Delayed Onset Muscle Soreness (DOMS) after FCE. The same dataset is used as the research described in Chapter 5. Five groups of predictors are tested in a multiple regression analysis model: personal variables, self reported activity, self reported health, perceived effort during the test and objective outcomes of the test. Twenty-three independent variables are selected and tested with a backward regression analysis. The onset of DOMS can be explained for 7% by the variables gender and the work index of the Baecke questionnaire. Variance of intensity of DOMS can be explained for 13% by the variables age, gender and VO₂max. Variance in duration of DOMS can be explained for 8% by the variables gender and reported emotional role limitations. Onset, intensity and duration of DOMS remained unpredictable for 87% or more. It is concluded that the intensity and duration of self reported DOMS can only minimally be predicted from the candidate predictors used in this study.
The objective of the study, presented in Chapter 7, was to establish normative values for FCE of healthy working subjects. Subjects working in over 180 professions performed an FCE consisting of 12 work-related tests. Subjects are classified into categories based on physical demands according to the Dictionary of Occupational Titles (DOT). Means, ranges, standard deviations and percentiles are calculated and regression analyses for outcome of the 12 tests were performed. The results of this study are that normative FCE values of 701 healthy working subjects (448 male, 253 female) are established for 4 physical demand categories. These normative values enable comparison of patients’ performances to these values. If the patients’ performances exceed the lowest scores in his/her corresponding demand category, the patients’ capacity is very likely to be sufficient to defy the work load. Further research is needed to test validity of the normative references with respect to work place assessments and return to work recommendations.

In Chapter 8, the lifting low test of the WorkWell FCE is compared to the Progressive Isoinertial Lifting Evaluation (PILE) to study whether both lifting assessment tests can be used interchangeably in patients with Chronic Low Back Pain (CLBP) and to explore whether psychosocial variables can explain possible differences. Fifty three patients with CLBP are tested twice in a counter balanced design. Pearson Correlation Coefficient of $r > 0.75$ and non-significant differences on two-tailed t-tests were considered as good comparability. The Pearson Correlation Coefficient was 0.75 ($p < 0.01$) but there appeared a significant difference of 6.0 kg. The difference between the PILE and the WWS appeared unrelated to psychological variables. It can be concluded that the PILE and the WWS FCE cannot be used interchangeably.

In Chapter 9, the main results of this thesis are discussed. The contribution to the measurement qualities of FCE is described and recommendations with regards to future research are given. The WorkWell FCE protocol which was used in the studies presented in this thesis was found safe for clinical testing when all safety procedures are followed. It was concluded that a pain response after FCE was a normal reaction of the musculoskeletal system after intensive exercise. The pain response in healthy subjects was similar to the pain response of patients with CLBP. It was recommended to clinicians to inform patients that a temporarily increase of pain may be expected and that this is no sign of pathology.

Evidence for content validity of FCE is confirmed in Chapter 2 and 3. Consensus is reached among experts in a considerable amount of definitions and theoretical framework. The WRULD FCE is found to be content valid and test-retest reliable. Further research is necessary to confirm other kinds of validity. The WRULD FCE protocol can be a worthy addition for clinicians to screen capacity of patients suffering from WRULD.
Concurrent validity between FCEs appears to be low and in Chapter 8 it is concluded that FCEs cannot be used interchangeably. It appears difficult to describe other kinds of measurement qualities of FCE. Data appear difficult to be interpreted, because no gold standard for FCE is available. The International Classification of Functioning, Disability and Health (ICF), offers a good conceptual framework to within classify terminology used in FCE research. Consensus in experts is of major importance to correctly interpret data. It is recommended to perform further research with respect to operational definitions of the term Functional Capacity Evaluation.

Normative references of the FCE contribute to a better interpretation for clinicians concerning patients’ Functional Capacity. These values enable comparison of work capacity to these normative values of subjects performing in the same category of work load. It is recommended to use these normative values indicative because further research with regards to validity is necessary.