Quality of functional capacity evaluation tests: a clinimetric approach

Gouttebarge, V.

Citation for published version (APA):
Summary
In the present thesis, clinimetric properties of Functional Capacity Evaluation (FCE) tests, purposed to measure physical work-ability, are evaluated. Clinimetrics deals with the design, administration and interpretation of instruments or tests meant to properly, usefully and accurately measure clinical and epidemiological variables. As stated in chapter 1, accurate measurements are a prerequisite for using any instrument or test for any purpose in any given context. Clinimetrics is covered by two major considerations: reproducibility and validity. Both of these considerations are fundamental for determining an instrument or test’s quality of measurement.

As musculoskeletal complaints (MSCs) and disorders (MSDs) are the primary reason for long-term sickness absence and work-related disability, properly and accurately assessing physical work-ability is an imperative matter in the contexts of rehabilitation, return to work and work disability. Professionals, working either in rehabilitation, occupational medicine or insurance medicine, do not possess many instruments to assess physical work-ability. Instead, they rely principally on the patient’s anamnesis and their physical examination. The use of complementary information, provided from a relevant performance-based instrument that is specifically developed to assess physical work-ability, could empower their judgement and decision making process. However, whether a judgement or decision can be based on an instrument outcome depends on the measurement quality of this instrument, i.e., its clinimetric properties (Chapter 1).

FCE methods are performance-based assessments that claim to measure the current physical work-ability of a person with or without MSCs. The Ergo-Kit (EK) FCE is one of the available FCEs in the Netherlands. Based on observation and testing criteria, the different EK FCE tests assess work-related activities such as sitting, reaching, bending, carrying and lifting. Whether the EK FCE tests could assist occupational professionals in their assessment of physical work-ability depends on their clinimetric properties, i.e., whether the EK FCE tests give reproducible and valid outcomes.
Summary

The following research questions are assessed in this thesis:

1. What is known in the international literature about the reliability of four FCE methods available in the Netherlands? (Chapter 2)
2. How reproducible are EK FCE tests in subjects without musculoskeletal complaints? (Chapter 3)
3. How reproducible are EK FCE tests in subjects suffering from musculoskeletal complaints? (Chapter 4)
4. What is known in the international literature about the validity of four FCE methods available in the Netherlands? (Chapter 2)
5. What is the construct validity of EK FCE tests in employees on sick leave due to MSDs? (Chapter 5)
6. What is the criterion-related validity of EK FCE tests in employees on sick leave due to MSDs? (Chapter 6)

In chapter 2, a systematic literature review is performed in order to gather existing information on the reliability and validity of four FCE methods available on the Dutch market: the Blankenship System (BS), the ERGOS Work Simulator (EWS), the Ergo-Kit (EK) and the Isernhagen Work System (IWS). Two independent reviewers applied the inclusion criteria to select all relevant articles retrieved from five databases (CINAHL, Medline, Embase, OSH-ROM and Picarta) and evaluated the methodological quality of all included articles. The search resulted in 77 potential relevant references, but only 12 papers were eligible for inclusion and were assessed for their methodological quality. The interrater reliability (Intraclass correlation coefficient ICC ≥ 0.95; agreement between raters ≥ 87%) and predictive validity of the IWS were evaluated and found to be good. No study was found that focused on the reliability of the EWS and EK while their concurrent validity was not demonstrated. No study concerning the reliability and validity of the BS was found. Our systematic literature review emphasizes the need for more studies on FCE clinimetrics.
Chapter 3 reports a study performed in adults without MSCs to evaluate the intra- and interrater reliability of seven EK FCE tests (Table 1): two isometric EK FCE lifting tests (Back-torso lift test and Shoulder lift test); two EK FCE manipulation tests (Forward manipulation test and Lower manipulation test crouching); and three dynamic EK FCE lifting tests (Carrying lifting strength test, Lower lifting strength test and Upper lifting strength test). Using a within-subject design, 27 subjects (15 males and 12 females) were assessed on the seven EK FCE tests at three different times (time interval of four days between each assessment), twice by rater R₁ and once by rater R₂. Intrarater reliability was high for the two isometric EK FCE lifting tests (ICC ≥ 0.85), moderate for the three dynamic EK FCE lifting tests (0.69 ≤ ICC ≤ 0.75), and low for the two manipulation EK FCE tests (ICC ≤ 0.46). Interrater reliability was moderate (8 day time interval; 0.67 ≤ ICC ≤ 0.90) to high (4 day time interval; ICC ≥ 0.85) for the two isometric and three dynamic EK FCE lifting tests, while it was low (8 day time interval; ICC ≤ 0.01) to moderate (4 day time interval; 0.74 ≤ ICC ≤ 0.78) for both manipulation EK FCE tests. The conclusion of this study is that the isometric and the dynamic EK FCE lifting tests have a moderate to high level of reliability in adults without MSCs, while the manipulation EK FCE tests have a low level of reliability.

Table 1 EK FCE test descriptions

<table>
<thead>
<tr>
<th>EK FCE test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Isometric lifting tests:</strong></td>
<td></td>
</tr>
<tr>
<td>Back-torso lift test (Blt; kg)</td>
<td>Use of a ‘Back and leg dynamometer’ fixed on platform, a chain and a handle. Handle is set at patella height for Blt and at elbow height for Slt. Maximal pulling during 4 s, 2 tries per test.</td>
</tr>
<tr>
<td>Shoulder lift test (Slt; kg)</td>
<td></td>
</tr>
<tr>
<td><strong>Manipulation tests:</strong></td>
<td></td>
</tr>
<tr>
<td>Forward manipulation test, standing (Fmt; s)</td>
<td>Use of a ‘Dexterity Ring’ (DR) fixed on shelves that can be vertically adjusted on a stand. DR is set at chest height for Fmt and at patella height for Lmtc. Screwing of 9 nuts and bolts, and shifting them to another position on the DR while maintaining standing posture for Fmt and crouching posture for Lmtc.</td>
</tr>
<tr>
<td>Lower manipulation test, crouching (Lmtc; s)</td>
<td></td>
</tr>
<tr>
<td><strong>Dynamic lifting tests:</strong></td>
<td></td>
</tr>
<tr>
<td>Carrying lifting strength test (Clst; kg)</td>
<td>Use of a stand with two vertically adjustable shelves, a box with different weights and a step (20cm). Following a standardized procedure, weight is added to the box (2.5, 5, 7.5 or 10 kg) depending on the task’s coordination, subject’s perception of the weight of the box, and (possible) subject complaints. 4-6 carries 5 m for Clst, 4-6 lifts from knuckle height to step for Llst and 4-6 lifts from knuckle to acromion height for Ulst</td>
</tr>
<tr>
<td>Lower lifting strength test (Llst; kg)</td>
<td></td>
</tr>
<tr>
<td>Upper lifting strength test (Ulst; kg).</td>
<td></td>
</tr>
</tbody>
</table>

kg, kilograms; s, seconds; cm, centimeters
In chapter 4, a reproducibility study is performed in order to assess the reliability and agreement between the raters of five EK FCE lifting tests by using a within-subject design. Twenty-four physiotherapist-recruited subjects (10 males and 14 females) suffering from low back pain (LBP) were assessed on two isometric (Back-torso lift test and Shoulder lift test) and three dynamic (Carrying lifting strength test, Lower lifting strength test and Upper lifting strength test) EK FCE lifting tests on two occasions, $t_1$ and $t_2$, by two different raters. Agreement between raters was expressed with the Standard Error of Measurement (SEM) in order to get a cautious insight into their sensitivity to change. ICCs (reliability) of isometric and dynamic EK FCE lifting tests ranged from 0.84 to 0.97. SEM values (agreement) ranged from 1.9 to 8.6 kg, while mean test scores ranged from 17.0 to 65.9 kg. The conclusion of this study is that the isometric and dynamic EK FCE lifting tests have a good reliability in subjects with LBP, while the agreement between raters can be considered to be acceptable when compared to the mean test scores.

Chapter 5 describes a study evaluating the construct validity (convergent and discriminative) of five EK FCE lifting tests in construction workers on sick leave due to MSDs. After being assessed on two isometric (Back-torso lift test and Shoulder lift test) and three dynamic (Carrying lifting strength test, Lower lifting strength test and Upper lifting strength test) EK FCE lifting tests, 72 participants were asked to fill in the Von Korff questionnaire (VK) on pain intensity and disability due to MSDs as well as the Instrument for Disability Risk (IDR). The IDR is a construction sector-specific instrument that identifies workers at risk for work disability due to MSD after a two-year period. The discriminative validity was evaluated by using the Known Groups Method and comparing the results of the EK FCE lifting tests scores between two groups of participants based on the IDR scores (higher risk for future work disability compared to lower risk for future work disability based on the IDR). Convergent validity was evaluated by assessing the associations between the scores of the EK FCE lifting tests and the outcomes of the VK self-report pain intensity and disability due to MSDs. Hypothesized differences in the five EK FCE lifting tests between both IDR risk groups were found in the expected direction, but they were not statistically significant ($0.07 \leq p$).
Pearson Correlation Coefficients ($r$) showed a poor convergent validity between the outcomes of the VK questionnaire and the EK FCE lifting tests scores ($-0.29 \leq r \leq 0.05$). The conclusion of this study is that the construct validity of the five EK FCE lifting tests could not be supported in construction workers on sick leave due to MSDs.

In **chapter 6**, a study evaluates the criterion-related validity of the five EK FCE lifting tests in construction workers on sick leave due to MSDs. Criterion-related validity was established through a prospective longitudinal within-subject design by evaluating the concurrent and predictive validity between two isometric (Back-torso lift test and Shoulder lift test) and three dynamic (Carrying lifting strength test, Lower lifting strength test and Upper lifting strength test) EK FCE lifting tests, and a reference test (the Instrument for Disability Risk (IDR)). The IDR is a construction sector-specific instrument that identifies workers at risk for work disability due to MSD after a two-year period. Furthermore, durable return to work (RTW)--the number of days on sick leave until full RTW for a period of at least four weeks--was recorded. This allowed the evaluation of the predictive validity of the EK FCE lifting tests on durable RTW. Six weeks (baseline), six months and one year after the first sick leave day ($t_0$, $t_1$ and $t_2$, respectively), participants were assessed on the five EK FCE lifting tests and asked to complete the IDR. Concurrent validity was evaluated by assessing the associations between the scores of the five EK FCE lifting tests and the IDR outcomes at $t_0$, $t_1$ and $t_2$. Predictive validity was evaluated by assessing the associations between the scores of the five EK FCE lifting tests at baseline and the IDR outcomes six months and one year later. The concurrent validity with future work disability risk at three time points during one year was found to be poor for the two isometric EK FCE lifting tests ($-0.15 \leq r \leq 0.04$) and moderate for the three dynamic EK FCE lifting tests at $t_1$ and/or $t_2$ ($-0.47 \leq r \leq -0.31$). Only one dynamic EK FCE lifting test, the Carrying lifting strength test, showed a moderate level of predictive validity on future work disability risk, especially at $t_1$ ($r =-0.39$; Area under the curve AUC=0.72). Furthermore, the predictive validity of the five EK FCE lifting tests on the number of days on sick leave until full durable RTW was weak ($1.00 \leq \text{Hazard ratio HR} \leq 1.05$). Overall, criterion-related validity was found
Summary

to be poor for the two isometric EK lifting tests and to be moderate for the three dynamic lifting tests, especially for the Carrying lifting strength test.

Chapter 7 reports the main findings of the five studies in this thesis. Furthermore, methodological and procedural considerations are given while the selection of EK FCE tests is discussed, evoking other issues that are important in evaluating the quality of FCEs. The clinimetrics model illustrated in the general introduction (Chapter 1) is also discussed, and a general conclusion is given with regard to the different research questions of this thesis:

1. our literature review up to 2004 showed that more studies were required to demonstrate the reproducibility of FCE tests;
2. the reproducibility of the EK FCE manipulation tests is poor in adults without MSCs and good for the isometric and dynamic EK FCE lifting tests;
3. the reproducibility of the isometric and dynamic EK FCE lifting tests is good in subjects suffering from LBP;
4. our literature review up to 2004 showed that more studies were required to demonstrate the validity of FCE tests;
5. the construct validity of the EK FCE lifting tests with self-reported pain and disability is not supported in construction workers on sick leave due to MSDs;
6. in the construction industry, the criterion-related validity with future work disability risk is poor for the isometric EK FCE lifting tests, and it is only moderate for the dynamic EK FCE lifting tests, with an upper value for the Carrying lifting strength test.

Furthermore, some implications for practice are given, questioning the possible use of the Carrying lifting strength test. This test, as it presents only a moderate level of validity, may not be solely used by occupational professionals working in health and safety services for jobs exposed to manual material handling. Finally, suggestions are made with regard to possible future research. For example, it will be important to determine whether information coming from the Carrying lifting strength test, in combination with, for instance, information provided by anamnesis, clinical examination
Summary

and self-reported questionnaires, could have an added value for the judgment and decision making process of occupational professionals in their assessment of physical work-(dis)ability.