Quality of functional capacity evaluation tests: a clinimetric approach
Gouttebarge, V.

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5

Construct validity of Functional Capacity Evaluation lifting tests in construction workers on sick leave due to musculoskeletal disorders

Abstract

Objective
To assess the construct (discriminative and convergent) validity of five Ergo-Kit (EK) Functional Capacity Evaluation (FCE) lifting tests in construction workers on sick leave due to musculoskeletal disorders (MSDs).

Design
Cross sectional within-subject design.

Setting
Occupational health service for the construction industry.

Participants
Seventy-two construction workers (all men) on six week sick leave due to MSDs.

Interventions
Not applicable.

Main Outcome Measure
After being assessed on five EK FCE lifting tests, participants were asked to complete the Von Korff questionnaire (VK) on pain intensity and disability due to MSDs and the Instrument for Disability Risk (IDR) assessing the risk for work disability. Discriminative validity was evaluated by comparing the results of the EK FCE lifting test scores between the two groups of participants based on the IDR scores (high risk for work disability compared to low risk for work disability). Convergent validity was evaluated by assessing the associations between the results of the EK FCE lifting tests and VK self-reported pain intensity and disability due to MSDs.

Results
The hypothesized differences between both IDR risk groups on the five EK FCE lifting tests were found in the expected direction but were not statistically significant (one test
exhibited a trend). Pearson Correlation Coefficients showed a poor convergent validity between the scores of the VK questionnaire and the EK FCE lifting tests (-.29 ≤ r ≤ .05).

**Conclusions**

Poor construct validity of the five EK lifting tests was found: discriminative validity was not statistically established and convergent validity with self-reported pain intensity and disability was poor.
Introduction

Musculoskeletal disorders (MSDs) are a major burden on working populations, health systems and social care entities worldwide (1,2). In the construction industry, MSDs are the primary reason for long-term sickness absence and related work disability (3,4). They appear to be strongly associated with manual material handling, especially lifting (3,4). In 10 states in the United States, lifting in the construction industry was found to be responsible for 21% of workers’ compensation cases due to MSDs (5). In order to reduce sick leave and workers’ compensation cost due to MSDs and to facilitate and empower knowledge on return to work, it is imperative for occupational professionals to acquire more insight into the physical work-ability of an injured worker (e.g., lifting ability in construction workers). Occupational professionals do not possess many instruments for assessing physical work-ability, relying essentially on information provided by physical examinations and patient self-reports (6-8). Therefore, the use of a performance-based instrument especially designed to assess physical work-ability could provide occupational professionals with relevant information.

Functional Capacity Evaluation (FCE) methods aspire to be performance-based assessment methods designed to measure the current physical work-ability of workers with or without MSDs (9-11). The Ergo-Kit (EK) FCE method relies on a battery of standardized tests that reflect work-related activities (e.g., standing, walking) with particular attention to manual material handling activities (e.g., lifting) (12). In workers compensation claimants with low back injury, performance on an FCE floor-to-waist lifting test appeared to predict time to resolution of temporary total disability as well as information from the entire FCE protocol (13). As for any clinical instrument, before FCE tests can be legitimately applied in any context, their clinimetric properties (e.g., reproducibility, validity) should be assessed (14-16). A systematic literature review indicated that evidence regarding the reliability and validity of the EK FCE tests was lacking (17). The EK FCE lifting tests were subsequently found reproducible in participants without MSDs and in participants suffering from low back pain (LBP) (18,19). The question remains as to whether the EK FCE lifting tests are valid.
The assessment of validity is not straightforward because of its plural types and definitions. Three types of validity are generally considered: content, criterion-related (concurrent and predictive) and construct (convergent, divergent and discriminative) (16,20,21). The validation of an instrument, even if it seems an endless process (as it can be assessed in many different populations), can be fulfilled by seeking and finding evidence for each type of validity. Criterion-related validity and construct validity are seen as the most relevant types for functional assessments (20,21). Within construct validity, discriminative validity is said to be present when the evaluated test discriminates between groups that are expected to be different from each other (also known as the Known Groups Method) (16). Convergent validity is another common way to support construct validity and measures how the evaluated test relates to another instrument which is assumed to reflect an associated concept, outcomes of both tests or instruments being expected to correlate moderately with each other (20,21).

Through the assessment of several activities, FCEs strives to measure functional physical work-ability. As suggested by Vlaeyen et al., patients suffering from (chronic) musculoskeletal pain may fear pain and also work-related activities that are expected to cause pain (22,23). Even more, they could even fear movement and physical activities assumed to cause reinjury (22,23). Then, from this perspective, it is thoughtful to expect that fear of pain i.e. movement due to musculoskeletal pain could negatively affect performance testing on FCEs. With regard to the convergent validity of the EK FCE lifting tests, a hypothesis could be that EK FCE lifting outcomes correlate moderately with pain-related outcomes, a high pain-related level being expected to lead to a poor performance on the EK FCE lifting tests.

Because of the significance of manual material handling, especially lifting, in the construction industry, the incidence of MSDs in this sector and the need for evidence on the validity of the EK FCE lifting tests, the following two research questions were formulated:

1. What is the discriminative validity of the EK FCE lifting tests in constructions workers on sick leave due to MSDs?
(2) What is the convergent validity of the EK FCE lifting tests with self-reported pain intensity and disability in constructions workers on sick leave due to MSDs?

Methods

Participants and recruitment procedures
From a nationwide list of construction workers on sick leave obtained from a large occupational health and safety service in the construction sector, potential participants were contacted by phone by the first author. If interested to participate, detailed written information on the study procedure was sent and statements of informed consent were signed. Participants were included according to the following inclusion criteria: (1) performing physical work in the construction industry, (2) age between 18 and 55 years old, and (3) on sick leave for at least 6 weeks (± 1 week) due to MSDs. Participants were free to quit the study at any time. A power analysis was performed for both research questions ([a] 2-tailed T test with $\alpha = 0.05$ and power = 0.80; [b] confidence level of 0.95, correlation coefficient set at .50 and limit at .30), indicating that 50 subjects were required. Taking drop-outs into account, we strived to include 75 participants. This study was performed in accordance with the Helsinki Declaration and approved by the Medical Ethics Committee of the Academic Medical Center in Amsterdam, the Netherlands.

The Ergo-Kit lifting tests
From among the EK “Physical Agility” tests (manipulation, balance, strength and endurance), five EK FCE lifting tests (Figure 1), that are reliable in adults without MSDs and with low back pain (LBP) (18,19), were selected in this study: two isometric lifting tests, the Back-torso lift test (Btlit) and Shoulder lift test (Slit), and three dynamic lifting tests, the Carrying Lifting Strength Test (Clst), Lower Lifting Strength Test (Llst) and Upper Lifting Strength Test (Ulst). Table 1 presents descriptions and outcomes of the
EK FCE lifting tests. The assessment of these five EK FCE lifting tests takes approximately 30 minutes and was performed in accordance with the standardized procedures, as described in the EK handbook by certified raters in 27 possible locations (12).

Table 1: EK FCE lifting test descriptions and outcomes (12).

<table>
<thead>
<tr>
<th>EK FCE tests</th>
<th>Description</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-torso lift test (Btlt)</td>
<td>Use of a &quot;back and leg dynamometer&quot; fixed on a platform, a chain and a handle. Handle is set at patella height for Btlt and at elbow height for Slt. Maximal pulling during 4 s, 2 tries per test</td>
<td>Maximal isometric lift capacity (kg)</td>
</tr>
<tr>
<td>Shoulder lift test (Slt)</td>
<td></td>
<td>Maximal safe weight for lifting (kg)</td>
</tr>
<tr>
<td>Carrying lifting strength test (Clst)</td>
<td>Use of a stand with two vertically adjustable shelves, a box with different weights and a step (20 cm). Following standardized procedure, weight is added to the box (2.5, 5, 7.5 or 10 kg), depending on the subject’s coordination in the task, subject’s perception of the weight of the box, and subject complaints. 4-6 carries 5 m for Clst, 4-6 lifts from knuckle height to step for List and 4-6 lifts from knuckle to acromion height for Ulst</td>
<td>Maximal safe weight for lifting (kg)</td>
</tr>
<tr>
<td>Lower lifting strength test (List)</td>
<td></td>
<td>Maximal safe weight for lifting (kg)</td>
</tr>
<tr>
<td>Upper lifting strength test (Ulst)</td>
<td></td>
<td>Maximal safe weight for lifting (kg)</td>
</tr>
</tbody>
</table>

**Discriminative validity**

To assess discriminative validity, two groups of participants for whom differences were expected, were formed post-hoc based on the Instrument for Disability Risk (IDR). This construction sector specific instrument identifies workers at risk for work disability due to MSD over a two-year period (24). The IDR has been developed in a Dutch setting and showed evidence of content validity (24). The IDR is determined from a questionnaire exploring four risk factors for work disability in the construction industry through nine questions addressing age, work ability, sickness absence and musculoskeletal complaints (see Appendix 1). The IDR is calculated with the scores obtained from nine questions and provides risk prognosis (expressed as a percentage) for work disability with high percentages denoting high risks for future work disability. Experts tend to set the IDR cut-off point at 38% for a high risk for work disability (24). Based on the IDR outcomes, participants were classified either in the group with a high risk for work disability or in the group with a low risk for work disability. Discriminative validity was assessed by comparing both groups with each other on their scores on the five EK FCE lifting tests, with the expectation that the high-risk group would score lower than the low-risk group.
Convergent validity
Convergent validity was evaluated by assessing the associations between the outcomes of the five EK FCE lifting tests and self-reported pain intensity and disability due to MSDs. In this study, an existing Dutch translation of the Von Korff questionnaire (VK) (25) was used to assess pain intensity and disability due to MSDs. The VK has been evaluated as reliable and valid (26,27). As the original version of the VK assesses pain and related disability of the low back in the past six months (25), the VK was adjusted for this study in order to fit the inclusion criteria of the present study. Therefore, MSDs pain intensity and related disability were asked with regards to the last week prevalence of MSDs and disability (see Appendix 2).

Pain intensity due to MSDs was assessed with three questions (scored on a scale from 0 to 10) about the subjects’ pain during performing activities of daily living. A 0 – 100 pain score (0 = no pain and 100 = the worst pain possible) was calculated based on the mean of these three questions (25). As the third question of the VK (see Appendix 2) focuses on the average pain intensity due to MSDs during a week, the associations between this third VK question alone and the outcomes of the five EK FCE lifting tests were also evaluated.

Disability due to MSDs was assessed with three questions about the subjects’ ability to perform activities and/or work (scored on a scale from 0 to 10). A 0 - 100 disability score (0 = not disabled and 100 = fully disabled) based on the mean of the disability questions multiplied by 10 was calculated (25). As the sixth question of the VK (see Appendix 2) focuses on MSD-related disability during work activities (while the other questions focus also on disability during social and daily living activities), the associations between this sixth VK question apart and the outcomes of the five EK FCE lifting tests were also evaluated.

Design and study procedures
A cross sectional study, using a within-subject design, was conducted to evaluate the discriminative and convergent validity of the five EK FCE lifting tests. According to
Dutch legislation, construction workers on sick leave have to visit an occupational health and safety service after six weeks of sick leave. There, after signing statements of informed consent, they were asked to fill in the VK and IDR questionnaires. Next, they were assessed on the five EK FCE lifting tests by certified raters at the nearest Ergo-Kit test center. The time interval between both assessments (EK FCE lifting tests and questionnaires) was targeted to be as short as possible (< 10 days).

**Data analyses**

Based on the IDR percentage scores of risk for work disability from the whole study sample and the original cut-off point established by experts, two groups expected to be different from each other were formed post-hoc: one group with a high risk for work disability and one group with a low risk for work disability. Descriptive statistics (i.e., means, frequencies, standard deviations and ranges) of the whole study sample and of both groups were calculated for age, gender, height, body weight, and for the outcomes of the five EK FCE lifting tests and the VK pain intensity and disability.

Discriminative validity between the high-risk group and the low-risk group was assessed using independent sample t-tests (16). A p-value below 0.05 was considered statistically significant.

Convergent validity was evaluated by correlating the EK FCE lifting tests scores and the VK pain intensity and disability outcomes using a Pearson correlation coefficient (16,21). For convergent validity, correlations greater than 0.60 were considered good, between 0.30 and 0.60 were moderate, and less than 0.30 were poor (16,21). All data analyses were performed using the statistical analysis software SPSS 13.0 for Windows.
Results

Participants
Over 400 potential subjects were contacted by phone by the first author and asked whether they were interested to participate voluntarily in the study. Seventy-two construction workers were willing to participate in this study. The main reason for not participating was that the most construction workers contacted by phone expected to return to work within a few days, becoming then ineligible for inclusion in our study. Some participants were simply not motivated or willing to participate. To guarantee that the time interval between both assessments (EK FCE lifting tests and questionnaires) was as short as possible, participants who did not return their questionnaires within 3 days (after their assessment on the EK FCE lifting test) were again contacted by phone.

Participants were assessed on the EK FCE lifting tests in 15 different locations in the Netherlands. Table 2 presents the participants characteristics. Among the participants, carpenters were the occupation most represented (36%). Within the whole population, upper extremity MSDs accounted for 17% of the main diagnosis, lower extremity for 28%, back for 31%, and the remaining proportion of 24% concerned a combination of MSDs. The participants’ mean self-reported pain intensity and disability scores were 59 and 61, respectively.

Table 2: Means, standard deviations and ranges of age, height, bodyweight and outcomes of the Ergo-Kit FCE lifting tests and Von Korff questionnaire for the whole group of participants.

<table>
<thead>
<tr>
<th></th>
<th>Total group (N = 72)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Age (years)</td>
<td>41</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>182</td>
</tr>
<tr>
<td>Bodyweight (kg)</td>
<td>87</td>
</tr>
<tr>
<td>Back-torso lift test (kg)</td>
<td>85.9</td>
</tr>
<tr>
<td>Shoulder lift test (kg)</td>
<td>45.0</td>
</tr>
<tr>
<td>Carrying lifting strength test (kg)</td>
<td>35.9</td>
</tr>
<tr>
<td>Lower lifting strength test (kg)</td>
<td>32.9</td>
</tr>
<tr>
<td>Upper lifting strength test (kg)</td>
<td>21.6</td>
</tr>
<tr>
<td>Von Korff pain intensity (0-100)</td>
<td>59.3</td>
</tr>
<tr>
<td>Von Korff related disability (0-100)</td>
<td>60.8</td>
</tr>
<tr>
<td>Von Korff question 3 (0-10)</td>
<td>5.7</td>
</tr>
<tr>
<td>Von Korff question 6 (0-10)</td>
<td>6.1</td>
</tr>
</tbody>
</table>

N, number of subjects; SD, standard deviation; cm, centimeter; kg, kilogram
Discriminative validity
In the present study, in order to form groups that should be expected to be even more different from each other, the following two extreme groups were formed: one group with a high risk for work disability ($n = 14; 0 \leq IDR \leq 35$), and one group with a low risk for work disability ($n = 40; 40 \leq IDR \leq 100$).

Table 3 presents the characteristics for both groups in terms of age, height and weight, as well as the mean, standard deviation and range in scores of the EK FCE lifting tests. High risk for work disability resulted in lower outcomes for four of the five EK FCE lifting tests. However, independent sample t-tests did not show statistical significant differences between both groups. For both isometric lifting tests (Back-torso lift test and Shoulder lift test), both IDR groups performed nearly the same ($t = 0.27/p = 0.79$ and $t = -0.28/p = 0.78$, respectively), while more substantial differences in scores on the three dynamic lifting tests could be observed, particularly on the lower lifting strength test ($7.4$ kg; $t = 1.82/p = 0.07$).

Convergent validity
The mean, standard deviation and range in scores of the VK pain intensity and disability and in scores of the VK questions three and six are presented in Table 2. Pearson correlation coefficients showed poor associations between the scores of the EK FCE lifting tests and the outcomes on the VK questionnaire (Table 4). The upper value of $r = -0.29$ for the association between the Carrying lifting strength test and the VK question three on pain intensity was found. In addition, analyses allowed for exploration of the internal validity of the two isometric and three dynamic EK FCE lifting tests: correlations were moderate to high (Table 4).
Table 3: Mean, standard deviation and range of age, height, weight and outcomes of the Ergo-Kit FCE lifting tests for the high-risk and low-risk groups for work disability.

<table>
<thead>
<tr>
<th></th>
<th>High risk (N = 40)</th>
<th>Low risk (N = 14)</th>
<th>t*</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>46</td>
<td>35</td>
<td>-3.79</td>
<td>0.00</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>182</td>
<td>183</td>
<td>0.40</td>
<td>0.69</td>
</tr>
<tr>
<td>Bodyweight (kg)</td>
<td>87</td>
<td>85</td>
<td>-0.58</td>
<td>0.56</td>
</tr>
<tr>
<td>Back-torso lift test (kg)</td>
<td>85.5</td>
<td>88.7</td>
<td>0.27</td>
<td>0.79</td>
</tr>
<tr>
<td>Shoulder lift test (kg)</td>
<td>43.7</td>
<td>42.1</td>
<td>-0.28</td>
<td>0.78</td>
</tr>
<tr>
<td>Carrying lifting strength test (kg)</td>
<td>35.1</td>
<td>41.1</td>
<td>1.41</td>
<td>0.16</td>
</tr>
<tr>
<td>Lower lifting strength test (kg)</td>
<td>32.6</td>
<td>40.0</td>
<td>1.62</td>
<td>0.07</td>
</tr>
<tr>
<td>Upper lifting strength test (kg)</td>
<td>21.9</td>
<td>25.2</td>
<td>1.09</td>
<td>0.28</td>
</tr>
</tbody>
</table>

N, number of subjects; SD, standard deviation; cm, centimeter; kg, kilogram; * comparison between the groups with a high and a low risk for work disability

Table 4: Correlation between the scores of Ergo-Kit FCE lifting tests and the outcomes on the Von Korff questionnaire (N = 72).

<table>
<thead>
<tr>
<th>Variables</th>
<th>bttl</th>
<th>sltt</th>
<th>cltt</th>
<th>llst</th>
<th>ulst</th>
<th>vkpi</th>
<th>vkrd</th>
<th>vkq3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Back-torso lift test</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Shoulder lift test</td>
<td>.55*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Carrying lifting strength test</td>
<td>.29*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Lower lifting strength test</td>
<td>.28*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Upper lifting strength test</td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Von Korff pain intensity</td>
<td>-.16</td>
<td>-.16</td>
<td>-.27*</td>
<td>-.23*</td>
<td>-.16</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Von Korff related disability</td>
<td>-.26*</td>
<td>-.17</td>
<td>-.15</td>
<td>-.11</td>
<td>.01</td>
<td>.58*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8. Von Korff question 3</td>
<td>-.26*</td>
<td>-.27*</td>
<td>-.29*</td>
<td>-.26*</td>
<td>-.16</td>
<td>.90*</td>
<td>.52*</td>
<td>1</td>
</tr>
<tr>
<td>9. Von Korff question 6</td>
<td>-.19</td>
<td>-.11</td>
<td>-.13</td>
<td>-.07</td>
<td>.05</td>
<td>.45*</td>
<td>.88*</td>
<td>.40*</td>
</tr>
</tbody>
</table>

N, number of subjects; * p < .05
Discussion

The purpose of this study was to evaluate the discriminative and convergent validity of five EK FCE lifting tests in constructions workers on sick leave due to MSDs. A group of construction workers with a high risk for work disability and a group of construction workers with a low risk for work disability were formed. None of the five EK FCE lifting tests could significantly discriminate between both groups. However, differences between both IDR groups on four out the five EK FCE lifting tests were found in the expected direction. Convergent validity was found to be poor between each EK FCE lifting test and self-reported pain intensity and disability. In this study, therefore, construct validity of the five EK FCE lifting tests could not be confirmed.

A strength of the present study concerns the population being used. As stated earlier, construction workers are heavily exposed to manual material handling, especially lifting, and MSDs are the major cause for sick leave and disability in the construction industry. Therefore, the use of construction workers on sick leave due to MSDs in the validation process of the EK FCE lifting tests is appropriate. In addition, the use of the IDR is also relevant. The Dutch construction industry has developed this specific instrument to identify workers at risk for work disability due to MSDs over a two-year period. Within the nine questions of the IDR, physical work ability is specifically addressed, indirectly reflecting the respondent’s lifting ability, as this one is the most important parameters within occupations in the construction industry. Furthermore, as age is one of the four risk factors being taken into account in the IDR risk prognosis for work disability, it is explicable that both IDR groups significantly differ in age (10 years; p < 0.01): the group with a high risk for work disability is almost 10 years older than the low-risk group.

From the results of this study, discriminative validity of the five EK lifting tests was not statistically established as no significant differences in EK lifting test scores were found between the two IDR groups. The expectation that two groups of construction workers could be formed with the IDR and would score statistically different on the EK FCE lifting tests, was not met. Furthermore, the expectation that both IDR groups would be
represented by an equal number of participants was not met also, with a larger number of construction workers in the low-risk group, explaining perhaps that statistical differences were not found.

However, the observed differences did occur in the expected direction; the high-risk group scoring lower on four out of the five EK FCE lifting tests than the low-risk group. Even more, the mean differences between both IDR groups in test scores for the three dynamic lifting tests were substantial: 6, 7 and 3 kg for the Carrying lifting strength test, Lower lifting strength test \( t = 1.82/p = 0.07 \) and Upper lifting strength test, respectively.

From a methodological perspective, the Known Group Method appears to be the appropriate method for assessing discriminative validity. In this validation approach, a hypothesized construct \( (i.e., \text{lifting capacity measured with two isometric and three dynamic EK FCE lifting tests}) \) is measured in two groups of people who are expected to differ from each other on the attribute that is the focus of the instrument being evaluated. Within such an assessment, the selection criteria of both groups are crucial. Based on the IDR outcomes, the lifting capacity of a group of construction workers on sick leave with a high risk for work disability was compared to the low-risk group. The expectation was that the high-risk group would score less than the low-risk group, and thus, the five EK FCE lifting tests would discriminate between both types of construction workers. We did find substantial differences in EK FCE lifting test scores between both IDR groups, but differences were not statistically significant, except for a trend toward the Lower lifting strength test, leading to the conclusion that perhaps the use of two, more different groups would have been more suitable. For instance, the comparison of a group of construction workers on sick leave due to MSDs and a group of working construction workers would have been a good possibility to use with the Known Group Method, without losing any relevance for the construction sector and providing occupational professionals with useful data on the EK FCE lifting test scores of disabled versus non-disabled construction workers.
However, we did try to use the IDR outcomes optimally by excluding the middle group of construction workers with a IDR score between 35% and 40% to form two extreme groups that could meet our expectations. Furthermore, post-hoc scatter plots were drawn to explore whether two different groups of construction workers could be formed for each of the EK FCE lifting tests by using different IDR cut-off points. The visual interpretation of these plots did not give any reason to believe that a different cut-off point per EK FCE lifting test would be better used in the present validation process.

Self-reported measures of pain, disability or physical work-ability are often used in clinical, rehabilitation or occupational settings because they are practical, inexpensive and only moderately time consuming. Measuring disability or physical work-ability can also be done through performance-based testing, such as FCE methods. Several studies suggested that moderate to strong correlations were found between self-reported measures of pain, disability or functioning, and FCE performance, especially FCE lifting tests (28-33), but others authors did not find associations between performance-based and self-reported measures of pain or disability (34-36). However, these studies were performed in different contexts or populations than the present study.

Convergent validity between the five EK FCE lifting tests and self-reported pain intensity and disability, based on the VK questionnaire, was not established in construction workers on sick leave due to MSDs. Even if some associations between each EK FCE lifting test and self-reported pain intensity and disability were statistically significant, no correlation coefficient met our criteria for good (greater than 0.60) or moderate (between 0.30 and 0.60) level of convergent validity. Therefore, a possible explanation for our results could be that measuring physical work-ability or disability through self-reported questionnaires or through performance-based testing appear to be two different approaches that did not allow to find a moderate level of correlation between the outcomes on the VK and EK FCE lifting tests. The self-reported VK disability questions reflect what someone reports he/she can or cannot perform, while the EK FCE lifting tests measure what someone performs. It is unclear whether functional self-efficacy beliefs of patients suffering from MSDs influence FCE
performance (37,38) but in general, people perform more or better than what they expect they can do. In our study, despite sick leave due to MSDs, pain and perhaps an associated low level of functional self-efficacy belief, construction workers were still able to perform well on the EK FCE lifting tests when compared with the normative scores available in the EK user guide (12). Especially for the three dynamic EK FCE lifting tests, the mean scores of the constructions workers would be interpreted as high on a five-point scale (from extremely low to extremely high) (12). On the other hand, when compared to other worker groups with heavy physical work demands (e.g., fire fighters), construction workers on sick leave due to MSDs scored lower (+/- 30 kg for the Btlt; +/- 15 kg for the Slt, and +/- 14 kg for both Llst en Ulst) on the EK FCE lifting tests (39,40). The expectation that pain due to MSDs would interfere negatively with the performance of the construction workers on the EK FCE lifting tests appears to be ungrounded. As suggested in other studies (41,42), construction workers were able to cope with their pain and to score at a high level on the EK FCE lifting tests.

**Conclusion**

Poor construct validity of the five EK FCE lifting tests was found in construction workers on sick leave due to MSDs. Discriminative validity of the five EK FCE lifting tests was not statistically established between a group with a high risk for work disability and a group with a low-risk group, except for a trend toward the Lower lifting strength test. However, differences in EK FCE lifting tests scores between both IDR groups were found to occur in the expected direction in four of the five lifting tests. Convergent validity of the five EK FCE lifting tests with self-reported pain intensity and disability was poor. Further studies are needed to assess other validity aspects of the EK FCE lifting tests, especially criterion-related validity.
Acknowledgment

The authors would like to thank to Marco van de Velde (‘ArboDuo’) and Cor van Duivenbooden (‘Stichting Arbouw’) for their support in this study. We are also grateful to all certified raters who assessed the EK FCE lifting tests and to all participants.

Funding

This study was financially supported by ‘Stichting Arbouw’ and ‘Stichting Instituut GAK’.
Appendix 1: Instrument for Disability Risk (IDR) (24).

1. How would you rate your current work ability compared to the lifetime best, where 0 is ‘not able to work’ and 10 is ‘best work ability ever’? (0-10 scale)

2. How would you rate your current work ability with regard to the physical work demands of your job? (5-points Likert scale)

3. How would you rate your current work ability with regard to the psychological work demands of your job? (5-points Likert scale)

4. Give in the following list of 51 diseases the number of current diseases you suffer diagnosed by a physician and diagnosed by your self. (number of diseases)

5. Give your estimation of work impairment due to diseases. (1-6 scale)

6. How many days were you on sick leave during the past year? (1-5 scale)

7. From your own perspective, do you think you will be working in your own job in two years? (3-points Likert scale)

8a. Did you lately enjoy your daily life? (5-points Likert scale)

8b. Have you been lately active and fit? (5-points Likert scale)

8c. Have you had lately trust in the future? (5-points Likert scale)

9a. Do you have regular neck stiffness or pain? (yes/no)

9b. Do you have regular stiffness or pain in the upper extremity? (yes/no)

9c. Do you have regular back stiffness or pain? (yes/no)

9d. Do you have regular stiffness or pain in the lower extremity? (yes/no)
Appendix 2: Von Korff (25) pain intensity (1 to 3) and disability (4 to 6) questions.

1. How would you rate your musculoskeletal pain on a 0-10 scale at the present time, that is right now, where 0 is ‘no pain’ and 10 is ‘pain as bad as could be’?
   
<table>
<thead>
<tr>
<th>No pain</th>
<th>Pain as bad could be</th>
</tr>
</thead>
<tbody>
<tr>
<td>0  1  2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>

2. In the past week, how intense was your worst pain rated on a 0-10 scale where 0 is ‘no pain’ and 10 is ‘pain as bad as could be’?
   
<table>
<thead>
<tr>
<th>No pain</th>
<th>Pain as bad could be</th>
</tr>
</thead>
<tbody>
<tr>
<td>0  1  2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>

3. In the past week, on the average, how intense was your pain rated on a 0-10 scale where 0 is ‘no pain’ and 10 is ‘pain as bad as could be’? (That is, your usual pain at times you were experiencing pain.)
   
<table>
<thead>
<tr>
<th>No pain</th>
<th>Pain as bad could be</th>
</tr>
</thead>
<tbody>
<tr>
<td>0  1  2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>

4. In the past week, how much has musculoskeletal pain interfered with your daily activities rated on a 0-10 scale where 0 is ‘no interference’ and 10 is ‘unable to carry on any activities’?
   
<table>
<thead>
<tr>
<th>No interference</th>
<th>Unable to carry on any activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>0  1  2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>

5. In the past week, how much has musculoskeletal pain changed your ability to take part in recreational, social and family activities where 0 is ‘no change’ and 10 is ‘extreme change’?
   
<table>
<thead>
<tr>
<th>No change</th>
<th>Extreme change</th>
</tr>
</thead>
<tbody>
<tr>
<td>0  1  2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>

6. In the past week, how much has musculoskeletal pain changed your ability to work (including housework) where 0 is ‘no change’ and 10 is ‘extreme change’?
   
<table>
<thead>
<tr>
<th>No change</th>
<th>Extreme change</th>
</tr>
</thead>
<tbody>
<tr>
<td>0  1  2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>
References

10. Strong S. Functional capacity evaluation: the good, the bad and the ugly. OT Now. 2002;Jan/Feb:5-9
40. Sluiter JK, Frings-Dresen MHW. What do we know about ageing at work? Evidence-based fitness for duty and health in fire fighters. Ergonomics. 2007;50:1897-1913