Chapter 6

Repeatability and safety of the functional capacity evaluation – one-handed for individuals with upper limb reduction deficiency and amputation

Sietke G. Postema, Raoul M. Bongers, Andrea G. Cutti, Corry K. Van der Sluis, Michiel F. Reneman

Submitted
Abstract

Purpose To assess repeatability and safety of the FCE-OH, a functional capacity evaluation for one-handed individuals, consisting of 8 items.

Method The FCE-OH protocol was administered twice to 23 individuals with upper limb absence (87% male; median age 46 years; median 2 days between sessions). Test-retest reliability and agreement were assessed with the intraclass correlations coefficient (ICC) and limits of agreement (LoA), respectively. Reliability was considered acceptable when ICC-values were ≥.75. Widths of LoA of four tests were compared with those of healthy adults. Safety and pain response were assessed with a questionnaire.

Results After controlling for stability of construct, ICC-values ranged between .23 and .96, and widths of LoA ranged between 16% and 79%. Intertrial (learning) effects were present in three test items. No serious adverse reactions were reported. A pain response was reported by 30% of the participants.

Conclusion Good or excellent reliability was observed in five tests, while three items showed poor or moderate test-retest reliability. Interpretation of agreement was possible for four tests, of which three showed widths of LoA similar to those reported in healthy adults. Learning effects were present; therefore, interpretation at the individual level should be performed with care. Safety was confirmed.
Introduction

After an amputation of the upper limb, return to work is often an important goal of rehabilitation, as employment is generally beneficial for the individual. However, functional capabilities may have altered due to the amputation and prosthesis use. Also, individuals who are born with a transversal reduction deficiency of the upper limb may experience physical limitations due to one-handedness, which may influence their functional capacity. In order to assess the functional capacity of individuals with upper limb absence (ULA) in a standardized environment, a Functional Capacity Evaluation (FCE) for one-handed individuals was developed. This instrument can be used to guide decision making of rehabilitation professionals regarding suitable work, or to measure outcomes of vocational rehabilitation programs. Moreover, it helps assessing work limitations due to musculoskeletal complaints, which are a frequent problem in individuals with ULA. Test outcomes can be compared directly with workload or indirectly with reference values. The FCE for one-handed individuals (FCE-OH) contains items that are adapted from an FCE for individuals with work-related upper limb disorders (WRULD). While all tests of the FCE-WRULD were considered reliable in healthy adults, this cannot automatically be assumed for the tests of the FCE-OH in its target patient group. Because the FCE-OH contains substantial alterations with regard to the FCE-WRULD, and is developed for a different patient group, psychometric properties should be examined separately.

It has been demonstrated in healthy workers and patient groups with different chronic pain syndromes that FCE has not led to serious adverse effects, although, a temporary pain increase is common. The safety of FCE application and pain response during and after FCE application in patients with ULA has not been investigated. The aim of this study was to examine the repeatability and safety of the FCE-OH.

Method

Setting

Patients were recruited from, and FCE sessions were held at the Prosthetic Center of the Italian Workers’ Compensation Authority (INAIL) in Vigorso di Budrio, Italy.

Design

Test-retest design; two FCE sessions were held with an interval of minimally 24 hours apart. A questionnaire on demographics was answered directly after session 1, while 24 hours after session 1 the participant was asked to answer several questions about pain response. Stability of construct (e.g. the participant’s self-perceived physical and mental health status being unchanged) was assessed with a questionnaire prior to session 2. The
guideline for reporting reliability and agreement study (GRRAS) checklist was followed.\textsuperscript{113}

**Ethics**

Participants received written information about the study and gave written informed consent before commencement of the protocol. This study was approved by the local medical ethical committee (reference number: 0001341), and registered in the Netherlands Trial Register (NTR; reference number: 5807).

**Participants**

Potential participants were inpatients from INAIL, who stayed at the centre for several days for prosthetic fitting, repair or training. Inclusion criteria were: age 18 to 62 years; presence of an upper limb reduction deficiency or amputation at or proximal to the carpal level; normal function of the unaffected hand; all seven items of the Italian translation of the physical activity readiness questionnaire (PAR-Q)\textsuperscript{101,114} were answered negatively, or, when the latter was not the case, if participation was considered safe by a medical doctor. Prosthesis use was not necessary for participation. All patients who met the inclusion criteria were invited. Determination of sample size was based on sample sizes of previous studies on repeatability of FCEs (ranging from 18 to 50; median of 30 participants\textsuperscript{95,109–112,115,116}), and availability of participants during the allocated study period.

**Procedures**

All sessions were administered by the same tester, who was trained in the standardized FCE-OH procedures. Participants and tester were blinded for the test results of session 1 until the second session was completed. After a general introduction of the sessions, the participant was verbally instructed how to perform each test. Each test was demonstrated by the tester. The participant was also instructed on the four termination criteria: 1) the participant wished to stop one or all tests for whatever reason; 2) the tester deemed it unsafe to continue, 3) the participant’s heart rate was above 85\% of his or her age-related maximum [220-age] or 4) a set time limit or number of repetitions was reached. Delayed onset (muscle) soreness as a result of the FCE-OH was expected and the participants were informed accordingly before signing the informed consent form. To provide safety during the FCE-OH tests, heart rate was monitored continuously with a heart rate monitor.

Six tests were performed, of which two (repetitive reaching test and fingertip dexterity test) were performed with the unaffected limb and prosthetic limb separately, thus making a total of eight test items. The repetitive reaching test with the unaffected limb and prosthetic limb, the fingertip dexterity test with the unaffected limb and the prosthetic limb, and the handgrip strength test with the unaffected limb were each performed three times (referred to as three trials). The tests were performed in a set order: overhead lifting test with a receptacle, overhead lifting test with a 2.0kg weight (with the unaffected limb), overhead
working test, repetitive reaching test (alternating three trials with the unaffected limb and three trials with the prosthetic limb), fingertip dexterity test (alternating three trials with the unaffected limb and three trials with the prosthetic limb), and handgrip strength test (three trials with the unaffected limb). The overhead lifting test with a receptacle and the overhead working test were performed two-handed (unaffected and prosthesis hand), unless the participant had no prosthesis available or had a transhumeral amputation. In that case the test was performed with the unaffected limb only. The fingertip dexterity test could not be performed with a cosmetic prosthesis. Materials, objects and test procedures are presented in Appendix 1.

Pain response was assessed with self-reported questionnaires. After session 1 participants received an extended version of the pain response questionnaire (PRQ) and were asked to answer the first 3 questions 24 hours after finishing session 1. These 3 questions informed after: 1) whether pain was perceived and, if so, the type of pain (muscular pain, other pain, or a combination of these), 2) whether the participant perceived this pain as being directly caused by the FCE session, and 3) whether the patient had experienced any other physical reaction after the first FCE session. The remaining 14 questions were answered prior to session 2, and assessed stability of construct and presence of pain in the 12 hours prior to session 2. If pain was present, the location of pain and the severity of the pain (on an 11 point numeric rating scale from 0 (no pain) to 10 (worst imaginable pain)) was asked. To control for stability of construct of measurement, which is a prerequisite for test-retest reliability analyses, changes in mental and physical health (equal, better or worse) compared to the first session, and changes in medication use since the first measurement were recorded. Moreover, it was asked whether the participants had received prosthesis training between sessions, and whether changes were made to the prosthesis (e.g. repair) since session 1. In addition, a questionnaire about demographics was administered.

Data analysis

All test scores consisted of continuous variables. For the items of the repetitive reaching test and the fingertip dexterity test, and the hand grip strength test the average of three trials was calculated and used for further analyses. Skewness and kurtosis values divided by their standard error were used to assess distribution of normality of difference between test outcomes of session 1 and 2. If both outcomes were smaller than ±1.96 normal distribution was assumed, and a paired sample t-test was performed to analyze whether test results of session 1 were significantly different from results of session 2. When the difference was not normally distributed, a Wilcoxon signed-rank test was used, and the median and IQR were presented. Differences were considered statistically significant when \( p < .05 \).

Descriptives of the test results during the first and second trial, one-way random Interclass Correlation Coefficients (ICC) for single measures, and 95% confidence intervals (95% CI) of the ICC-values were computed. ICCs of \( \geq .90 \) were interpreted as excellent,
ICCs between .75 and .90 as good, ICCs between .50 and .75 as moderate, and ICCs of ≤.50 as poor. ICC-values of ≥.75 were considered acceptable. 95% Limits of agreement (LoA) were calculated ([mean difference ± 1.96 x standard deviation of mean difference]), which represent the size of difference between both measurements; approximately 95% of differences will lie between these LoA. Only changes outside the LoA should be considered as real change. In order to get a global impression of the width of the LoA, a ratio between the LoA and the mean score of session 1 and 2 was calculated ([1.96 x standard deviation of mean difference] / [mean session 1 and 2]) x 100%. When the difference between test outcomes of session 1 and 2 was not normally distributed, the value of 1.96 in the previous two calculations was replaced with the value 2. Interpretation of LoA is a clinical decision and not a statistical one. Widths of LoA of the overhead lifting test, overhead working test, fingertip dexterity test and handgrip strength test were compared with the widths of LoA found in healthy adults; differences of ≥10% were considered deviant. As it is unknown whether the LoAs found in healthy adults are acceptable, clinically relevant interpretation of (widths of) LoA is not possible.

Test-retest reliability explains the extent to which scores for patients who have not changed are the same for repeated measurement over time. Therefore two analyses were performed; first an analysis including all participants, and second an analysis including only the participants with stability of construct of measurement (e.g. stable functional capacity, in this study determined as unchanged physical and mental health status, and no changes in medical use as measured with the PRQ). Furthermore, in order to assess inter-trial variation a repeated measures one-way ANOVA was performed if a test consisting of multiple trials showed a significant difference in test results between session 1 and 2. If Mauchly’s test of sphericity was significant, the Greenhouse-Geisser estimate was reported. To examine differences between trials a post hoc Bonferroni analysis was performed.

All analyses were performed in IBM Statistics SPSS 22.

Results

Thirty-two individuals were invited to participate, of which 2 declined. Therefore, 30 individuals participated, of which 23 performed the FCE-OH protocol twice (completers), with a median time of 47.2 hours (IQR: 43.7 to 68.0) between sessions. Reasons for 7 participants not to perform session 2 (non-completers) were: logistic difficulty to schedule a second session due to time constraints (n=4), declining for unknown reason (n=2), and no show (n=1). Characteristics of the participants are presented in Table 1. Differences between completers and non-completers were small (less than 10% difference for each variable).
Table 1 Characteristics of the participants.

| Characteristics                                | All participants | Participants included in the test-retest analyses |
|------------------------------------------------|------------------|--------------------------------------------------|
| Number of participants                         | 30               | 23                                               |
| Gender: male/female (n)                        | 24/6             | 20/3                                             |
| Age (years) (median (IQR))                     | 46.2 (35.2; 54.4)| 46.0 (35.2; 55.5)                                |
| BMI (kg/m$^2$) (median (IQR))                  | 25.8 (23.4; 28.3)| 25.7 (23.7; 27.8)                                |
| Marital status (n (%))                         |                  |                                                  |
| Single                                         | 10 (33.3)        | 7 (30.4)                                        |
| Living together with partner                   | 19 (63.3)        | 16 (69.6)                                       |
| Divorced or widowed                            | 1 (3.3)          | 0 (0)                                           |
| Highest level of education (n (%))             |                  |                                                  |
| Primary school                                 | 1 (3.3)          | 1 (4.3)                                         |
| Middle or high school                          | 21 (70.0)        | 15 (65.2)                                       |
| College or university                          | 8 (26.7)         | 7 (30.4)                                        |
| Employed (n (%))                               | 14 (46.7)        | 12 (52.2)                                       |
| Level of amputation (n)                        |                  |                                                  |
| Transhumeral                                    | 5 (16.7)         | 3 (13.0)                                        |
| Transradial or wrist disarticulation           | 25 (83.3)        | 20 (87.0)                                       |
| Side of deficiency: left/right (n)              | 13/17            | 11/12                                           |
| Cause of deficiency (n)                        |                  |                                                  |
| Congenital                                     | 1 (3.3)          | 1 (4.3)                                         |
| Amputation: trauma                             | 29 (96.7)        | 22 (95.7)                                       |
| Amputation of dominant hand (n (%))             | 17 (63.0)        | 12 (60.0)                                       |
| Time since amputation (years) (median (IQR))   | 1.7 (1.4; 5.1)   | 1.7 (1.4; 5.1)                                 |
| At clinic for first fitting with prosthesis (n (%)) | 6 (20.0)         | 5 (21.7)                                        |
| Years of prosthesis use (median (IQR))         | 1.5 (1.0; 18.0)  | 1.5 (1.0; 18.8)                                |
| Prosthetic use during FCE-OH testing (n (%))   | 19 (63.3)        | 15 (65.2)                                       |
| Myoelectric traditional                        | 11 (57.9)        | 8 (53.3)                                        |
| Myoelectric with multiarticulating hand         | 1 (5.3)          | 1 (6.7)                                         |
| Body-powered                                    | 4 (21.1)         | 3 (20.0)                                        |
| (Prototype) cosmetic                            | 3 (15.8)         | 3 (20.0)                                        |

$^1$ Not included in the calculation of the percentage were two individuals who were ambidextrous before the amputation and one individual with a congenital reduction deficiency; $^2$ Not included in the analysis was one individual with a congenital reduction deficiency; $^3$ Individuals who visited the clinic for the first fitting with a prosthesis; therefore these individuals were not experienced with prosthesis use and did not have a prosthesis available; $^4$ Individuals who visited the clinic for the first fitting with a prosthesis were not included in the analysis; $^5$ One missing value.

Stability of construct

Six individuals had prosthesis training between FCE-OH session 1 and 2, all but one participant used the same prosthesis during both sessions, and two participants had changes to the prosthesis. In total, six individuals mentioned changed health status at session 2: mental health was better (n=3), or worse (n=1), or physical health was better (n=2), or worse (n=1). No participants had alterations in medication use between sessions. One participant showed a difference in overhead lifting capacity of 16kg. During session 2, issues with the prosthesis led to substantial difficulties in lifting performance, as observed
by the tester and confirmed by the participant. Therefore, the results of this participant were omitted from the analyses of the overhead lifting test.

Repeatability

Both primary (with all participants; Table 2) and secondary (participants with changed health status excluded; Table 3) analyses showed acceptable reliability (ICC-values of ≥.75) for five out of eight items of the FCE-OH, and one item close to the .75 threshold of acceptable reliability. Differences between primary and secondary analyses were small and did not influence interpretation of acceptability. Secondary analyses revealed widths of LoA ranging between 16% and 79%. Differences of the widths of LoA of the overhead lifting test, fingertip dexterity test, and handgrip strength test observed in this study and in healthy adults (23%, 14%, and 20%, respectively) were ≤10%, and thus considered similar. The width of LoA of the overhead working test was wider in this study (79%) compared to the width of LoA in healthy adults (41%).

Participants performed significantly better during the second session on the repetitive overhead lifting test with the 2.0kg weight, the repetitive reaching test (both with the unaffected limb and the prosthetic limb), and on the fingertip dexterity test with the unaffected hand (Table 2 and 3). Analyses of inter-trial variation are presented in Table 4. As the first trial of the repetitive reaching test was performed slower compared to all following trials an extra analysis was performed, omitting this trial. The ICC-value showed an evident increase (to .69, 95% CI: .32 to .88), however, the .75 level of acceptable reliability was still not reached. The width of LoA changed minimally.

Safety

No tests were terminated due to surpassing 85% of the age-related maximum heart rate. During the first session the overhead lifting test was thrice terminated by the tester, as it was deemed unsafe to continue (generally due to too much bodily swing while lifting, sometimes in combination with difficult grip with the prosthesis). During the second session this occurred only once. No serious adverse reactions occurred, but one individual reported a bruise on the unaffected forearm one day after the first session. This adverse reaction was most likely caused by pressure of the lower rim of the container on the forearm, while lifting the container with one hand during the overhead lifting test. After this event the container was padded with foam, and no such incident occurred again. Eight (30%) participants reported a physical response 24 hours (pain or other) after the first FCE-OH session, which was partly or completely caused by the test procedure (Figure 1). Five of these eight individuals performed both sessions; in all five participants the pain was still present at the start of session two (median pain grade: 4, range: 2 to 5). Three of the eight individuals with a pain response did not perform session two; one of these three individuals declined further participation due to the pain response.
Table 2 Test results of two FCE sessions, and limits of agreement and interclass correlation coefficients between these test results (all participants; n=9-23).

| Test description                                      | N  | Mean or median session 1 | SD or IQR session 1 | Mean or median session 2 | SD or IQR session 2 | Mean difference | SD of mean difference | 95% CI of mean difference | p   | LoA | Ratio of LoA (%) | ICC   | 95% CI of ICC | Interpretation |
|-------------------------------------------------------|----|--------------------------|---------------------|--------------------------|---------------------|------------------|-----------------------|----------------------------|------|-----|----------------|-------|--------------|----------------|
| Overhead lifting test with a container\(^1\)          | 22 \(^4\) | 8.0 | 5.8; 12.0 | 8.0 | 6.0; 10.5 | 0.2 | 1.6 | -0.5; 0.9 | .581 | -3.1; 3.4 | 40.3 | .89 | .77; .95 | Good |
| Overhead lifting test with a 2.0kg weight\(^2\)       | 23  | 34.2 | 7.4 | 32.1 | 6.8 | 2.1 | 0.6 | 3.6 | .014\(^*\) | -5.2; 9.4 | 22.0 | .83 | .64; .92 | Good |
| Overhead working test\(^2\)                           | 23  | 216.0 | 127.0; 93.0 | 185.0 | 112.0; 61.0 | 37.8 | 77.0 | 6.3; 69.3 | .052 | -116.2; 191.8 | 70.6 | .74 | .48; .88 | Moderate |
| Repetitive reaching test with unaffected limb\(^2\)   | 22 \(^5\) | 51.0 | 47.2; 53.3 | 46.3 | 40.2; 50.4 | 6.5 | 5.6 | 4.2; 8.9 | <.001** | -4.7; 17.7 | 23.0 | .59 | .24; .81 | Moderate |
| Repetitive reaching test with prosthesis\(^2\)        | 13 \(^8\) | 57.7 | 46.2; 64.2 | 55.0 | 44.2; 60.3 | 5.1 | 5.4 | 2.1; 8.0 | .003** | -5.7; 15.8 | 19.1 | .88 | .65; .96 | Good |
| Fingertip dexterity test with unaffected hand\(^3\)   | 22 \(^7\) | 12.8 | 2.3 | 13.4 | 2.0 | -0.6 | 1.1 | -1.1; -0.1 | .022* | -2.9; 1.6 | 17.6 | .83 | .63; .92 | Good |
| Fingertip dexterity test with prosthesis\(^3\)        | 9 \(^8\) | 3.7 | 1.2 | 4.5 | 1.1 | -0.7 | 1.2 | -1.5; 0.0 | .098 | -3.1; 1.6 | 56.6 | .38 | -.30; .81 | Poor |
| Hand grip strength test\(^2\)                         | 23  | 43.6 | 12.3 | 44.6 | 11.8 | -0.9 | 4.4 | 2.7; 0.8 | .310 | -9.5; 7.6 | 19.4 | .93 | .85; .97 | Excellent |

---

1 Kilograms; 2 second; 3 number of pins placed; 4 The participant who lifted a difference of 16kg was omitted from analysis; 5 One participant performed only one trial during session 1; 6 Eight participants performed all tests without prosthesis, as they had a transhumeral amputation (n = 3), or did not have a prosthesis available (n = 5). Two of the remaining participants did not perform this test due to sweating of the stump, which would cause twitching of the skin in the socket; 7 One participant declined this test during session 2; 8 Eight participants did not use a prosthesis during testing, two participants did not perform this test due to sweating of the stump, and three participants used a cosmetic prosthesis and thus could not perform this test; * The difference is statistically significant at p<.05; ** The difference is statistically significant at p<.01.
Table 3 Test results of two FCE sessions, and limits of agreement and interclass correlation coefficients between these test results, of individuals with a stable construct (six individuals with changed self-perceived physical and mental health status excluded; n=6-17).

| Test | n | Mean or median session 1 | SD or IQR session 1 | Mean or median session 2 | SD or IQR session 2 | Mean difference | SD of mean difference | 95% CI of mean difference | p | LoA | Ratio of LoA (%) | 95% CI of ICC | Interpretation |
|------|---|--------------------------|---------------------|--------------------------|---------------------|-----------------|--------------------|------------------------|----|-----|----------------|--------------|---------------|
| Overhead lifting test with a container | 17 | 8.0 | 6.5; 12.0 | 8.0 | 7.0; 11.0 | 0.2 | 1.2 | -0.3; 0.8 | .414 | -2.2; 2.6 | 29.4 | .94 | .84; .98 | Excellent |
| Overhead lifting test with a 2.0kg weight | 17 | 34.5 | 8.0 | 31.2 | 7.0 | 2.8 | 3.8 | 1.0; 4.6 | .007** | -4.6; 10.3 | 22.8 | .82 | .57; .93 | Good |
| Overhead working test | 17 | 205.0 | 117.5; 313.5 | 171.0 | 112.0; 278.0 | 40.6 | 83.0 | 1.1; 80.1 | .136 | -125.5; 206.6 | 78.5 | .73 | .40; .89 | Moderate |
| Repetitive reaching test with unaffected limb | 16 | 50.2 | 44.2; 52.8 | 45.3 | 39.2; 50.0 | 7.0 | 5.7 | 4.2; 9.8 | <.001** | -4.5; 18.5 | 24.2 | .57 | .13; .82 | Moderate |
| Repetitive reaching test with prosthesis | 9 | 53.0 | 43.0; 64.2 | 47.3 | 40.2; 59.7 | 5.7 | 6.3 | 1.6; 9.9 | .017* | -6.9; 18.4 | 23.4 | .87 | .57; .97 | Good |
| Fingertip dexterity test with unaffected hand | 17 | 12.8 | 2.1 | 13.4 | 2.0 | -0.6 | 1.1 | -1.1; 0.0 | .049* | -2.8; 1.6 | 17.1 | .82 | .57; .93 | Good |
| Fingertip dexterity test with prosthesis | 6 | 3.5 | 1.0 | 4.6 | 1.3 | -1.1 | 1.2 | -2.1; -0.1 | .073 | -3.5; 1.3 | 58.2 | .23 | -.58; .84 | Poor |
| Hand grip strength test | 17 | 41.2 | 11.1 | 41.7 | 11.8 | -0.6 | 3.3 | -2.1; 1.0 | .498 | -7.0; 5.9 | 15.6 | .96 | .95; .99 | Excellent |

1 Kilograms; 2 seconds; 3 number of pins placed; * The difference is statistically significant at p<.05; ** The difference is statistically significant at p<.01.
Table 4 Intertrial effects of the repetitive reaching test with the dominant hand, the repetitive reaching test with the prosthetic hand and the fingertip dexterity test with the dominant hand.

| FCE-OH test item                          | n   | Session 1 | Session 2 | F-value (df) | p    |
|-------------------------------------------|-----|-----------|-----------|--------------|------|
|                                           |     | Trial 1   | Trial 2   | Trial 3      | Trial 4 | Trial 5 | Trial 6 |           |      |
| Repetitive reaching test with the dominant hand | 16  | 54.8±8.9 | 50.3±10.2 | 47.9±11.7 | 45.8±7.3 | 43.8±8.9 | 42.5±7.8 | 19.2 (2.3) | <.001**|
|                                           |     | 2** 3** 4** 5** 6** | 50.3±10.2 | 47.9±11.7 | **      | **      | **      |           |      |
| Repetitive reaching test with the prosthetic hand | 9   | 61.1±17.9 | 56.8±18.6 | 53.4±20.1 | 52.2±14.1 | 52.0±13.5 | 49.9±13.6 | 6.6 (2.0)  | .008** |
|                                           |     | 5** 6*   | 56.8±18.6 | 53.4±20.1 | **      | **      | **      |           |      |
| Fingertip dexterity test with the dominant hand | 17  | 11.6±1.6 | 13.4±2.4 | 13.3±2.8 | 12.5±2.4 | 13.8±2.2 | 13.8±2.0 | 8.2 (5)  | <.001** |
|                                           |     | 2** 3** 5** 6** | 13.4±2.4 | 13.3±2.8 | **      | **      | **      |           |      |

For both items of the repetitive reaching test the Mauchly’s test of sphericity was statistically significant and therefore the Greenhouse-Geisser estimate was reported. Statistically significant differences between trials are reported in superscript (with numbers referring to trials). Data is presented as mean±SD. Trial 1 to 3 were performed during session 1, and trial 4 to 6 during session 2. Abbreviations: df: degrees of freedom.

1 Seconds; 2 number of pins placed; * The difference is statistically significant at p<.05; ** The difference is statistically significant at p<.01.
Figure 1 Pain response 24 hours after FCE-OH session 1, and at the start of session 2.

Abbreviations: MSC: musculoskeletal complaints.

Discussion

Repeatability

Five of the eight items of the FCE-OH showed acceptable reliability. For the repetitive reaching test with the unaffected limb and the fingertip dexterity test with the prosthesis test-retest reliability was not acceptable. The overhead working test was close to reliable. However, the width of LoA of this test was much wider compared to the width of LoA in healthy adults. Three other tests showed similar widths of LoA, while for four remaining tests interpretation of agreement was not possible.

The overhead working test showed a large width of LoA, meaning large within individual differences between sessions. The long duration of the test enhances the chance that an individual performs notably different when repetitively performing the test.
However, this test showed considerably smaller widths of LoA when performed by healthy adults\(^9\) and patients with whiplash associated disorders\(^1\) (41\% and 49\%, respectively, versus 71\% in this study), as well as a higher ICC-values (.90 and .83, respectively). The overhead working test is known to show variable ICC-values, ranging from .36 in patients with low back pain\(^1\) to .90 in healthy adults.\(^9\) Possibly, participants did not completely recover between sessions, and results of this endurance test, were affected by muscle fatigue. Furthermore, it could be that reminiscence of fatigue and possibly pain decreased motivation.

In comparison with the overhead working test, the repetitive reaching test with the unaffected limb was performed significantly better during session 2, which may be caused by learning effects. The higher bound of LoA showed that a test had to be performed at least 19 seconds faster, to be considered as a real change.\(^1\) The first trial of session 1 was performed significantly slower than all following trials, and therefore it was hypothesized that removing this trial from analysis would improve reliability measures. The ICC-value increased evidently when this trial was removed; however, the .75 level of acceptable reliability was not reached. The test showed modest variability between subjects, which can substantially decrease reliability, as reliability demonstrates how well persons can be distinguished from each other, despite measurement error.\(^1\) In the FCE-OH, the repetitive reaching test has been substantially altered and therefore reliability measures cannot be compared with existing literature.

For the fingertip dexterity test with the prosthesis the low ICC-value may very well be mediated by the low number of participants. No definitive conclusions should be drawn until the reliability of this test is ratified in a larger sample.

Secondary analyses were performed, including only individuals with predefined stability of construct. ICC-values are ratio measures of the between-subject variance and the total variance, the latter including within-subject variance (measurement error).\(^1\) Excluding individuals without stability of construct decreased within-subject variance, but possibly also the between-subject variance, and definitely the number of participants, which may explain why differences between results of the primary and secondary analyses are small, and did not change interpretation of reliability. Significant difference of test results between session 1 and 2, and inter-trial variation was present in several tests. Nevertheless, when within-participant differences are smaller than between-participant differences, acceptable reliability coefficients are possible.\(^1\) However, it is important to be aware of these effects when assessing a patient.

Safety

When necessary precautions are taken, the FCE-OH seems to be safe in use, since no serious adverse events occurred, and heart rates of all participants fell within acceptable ranges. One patient experienced a bruise one day after FCE-testing, which may be classified as an adverse reaction, which is defined as “any untoward and unintended response to an investigational medical product”.\(^1\) Furthermore, several individuals
experienced a pain response after the first FCE-OH session. Pain was mostly denoted as muscular pain and located in the shoulder of the affected limb, which may be caused by the generally passive use of this limb in daily life. Moreover, exacerbation of MSC may occur. The percentage of individuals with a pain response was much lower than found by Soer et al.\(^99\) (30\% versus 82\%, respectively). Reasons for this finding are still speculative and beyond the scope of this article.

**Strengths & weaknesses**

While the COSMIN guideline recommends 50 participants, the results of this study, and of other FCE reliability studies with a similar number of participants,\(^{110,111,115}\) show that a substantially smaller sample can be sufficient to establish reliability. However, the smaller sample may have provoked large CIs of ICC-values, making it necessary to frame clinical interpretations at the individual level with care. Most individuals eligible and available for the study were willing to participate; however, completion of both sessions was not always possible and mostly related to time constraints. It is unknown whether this caused any bias.

The interval between both sessions should be long enough to avoid recall bias and fatigue, but short enough to avoid changes in health status, causing genuine difference in performance. Following practical considerations, the interval in this study was variable, with a median of approximately two days. This is a shorter time interval compared to most studies, which had time intervals of one to two-three weeks,\(^{95,110–112,116}\) but similar to the time interval in the study of Gross and Battie, who had two to four days between sessions.\(^{115}\) A short interval may cause recall bias of test results, leading to higher ICC-values; but simultaneously may lead to lower ICC-values as the interval might not allow for full recovery. With exception of the overhead working test, we don’t expect the short interval to have played a role, as participants typically performed equal or better during session 2. Reliability measures of the overhead working test are preferably replicated in a study with a larger interval between sessions.

In this study widths or LoA are compared with healthy adults.\(^{95}\) Some tests of the FCE-OH were substantially altered, and therefore could not be compared. Interpretation of LoA is a clinical decision,\(^{120}\) and a possible way to interpret them is by using the minimal clinically important change. However, the FCE-OH being new, the minimal clinically important change still is to be established. Therefore, further considerations on LoA will follow.
Conclusion

Good or excellent test-retest reliability was observed in five tests, while the remaining three tests showed poor or moderate test-retest reliability. Comparison of agreement was possible for four tests, of which three showed similar agreement. The FCE-OH was considered safe in use when the right precautions are taken. Large CIs of the ICC-values and LoA, as well as learning effects, make it necessary to frame clinical interpretations at the individual level with care.
## Appendices

### Appendix 1 Materials, objects and procedure of the six FCE-OH tests.

| Test                          | Objective                                         | Materials                                                                 | Procedure                                                                 | Measurement outcome |
|-------------------------------|---------------------------------------------------|---------------------------------------------------------------------------|---------------------------------------------------------------------------|---------------------|
| Overhead lifting with a container | Functional strength of shoulder and arm musculature | Plastic receptacle (40x30x26cm), a wall mounted system with adjustable shelves and weights of 1.0, 2.0 and 4.0kg | 5 lifts from table to crown height v.v. within 90 seconds in standing position. Increments of weight (2.0kg for females, 4.0kg for males) until maximum amount of weight is reached. Prosthesis is used when available. | Maximum amount of lifted weight (kg) |
| Overhead lifting with a 2kg weight | Functional strength of shoulder and arm musculature | A wall mounted system with adjustable shelves and a weight of 2.0kg | 20 lifts from table to crown height v.v. in standing position, with the unaffected hand. | Time needed for 20 lifts v.v. (sec) |
| Overhead working               | Static holding time of shoulder and neck musculature | Aluminium plate adjustable in height with 5 holes, bolts and nuts, and one cuff-weight of 1.0kg | Standing with hands at crown height, manipulating nuts and bolts, and wearing cuff weights around the unaffected wrist. Prosthesis is used when available. | Time (sec) which static position can be held |
| Repetitive reaching            | Fast repetitive movements of the upper extremity    | Two click-systems                                                        | Sitting with clicking systems on wing span, touching clicking system at table height 30 times from left to right v.v. with unaffected or prosthesis hand. | Time (sec) 30 movements (left to right v.v.) are made |
| Fingertip dexterity            | Fingertip dexterity                                | Purdue Pegboard (model #32020, Lafayette IN.)                             | Sitting in front of the pegboard, placing as many pins as possible with unaffected or prosthesis hand in a 30 second trial.* | Number of pins |
| Handgrip strength              | Isometric hand grip strength                       | Hand dynamometer (G100 Biometrics LTD E-LINK)                             | Sitting with the shoulder adducted and neutrally rotated, elbow flexed at approx. 90°, and the forearm and wrist in neutral position, pinching a hydraulic handgrip dynamometer in position 2 for approx. 3 seconds. | Pinched kg |

NOTE: Participants with an above elbow ULA perform all tests without prosthesis.

* When the test is performed with the prosthesis, the participant is allowed to place the pins in the prosthesis hand with the unaffected hand.
