Intra- and Interrater Reliability of the Ergo-Kit Functional Capacity Evaluation Method in Adults Without Musculoskeletal Complaints

Vincent Gouttebarge, MSc, Haije Wind, MD, MSc P. Paul Kuijer, PhD, Judith K. Sluiter, PhD, Monique H. Frings-Dresen, PhD

ABSTRACT. Gouttebarge V, Wind H, Kuijer PP, Sluiter JK, Frings-Dresen MH. Intra- and interrater reliability of the Ergo-Kit functional capacity evaluation method in adults without musculoskeletal complaints Arch Phys Med Rehabil 2005; 86:2354-60.

Objective: To evaluate the intra- and interrater reliability of tests from the Ergo-Kit (EK) functional capacity evaluation method in adults without musculoskeletal complaints.

Design: Within-subjects design.

Setting: Academic medical center in the Netherlands.

Participants: Twenty-seven subjects without musculoskeletal complaints (15 men, 12 women).

Interventions: Not applicable.

Main Outcome Measures: Seven EK tests (2 isometric, 3 dynamic lifting, 2 manipulation tests) were each assessed 3 times (over 4 days), twice by 1 rater (R_1) and once by another rater (R_2). Intrarater reliability was calculated using the EK test scores assessed by R_1. Interrater reliability was calculated using the EK test scores assessed by both raters. Counterbalancing the rater order made possible the calculation of 2 interrater reliability levels (at time intervals of 4 and 8d). All reliability levels were expressed as intraclass correlation coefficients (ICCs).

Results: Intrarater and interrater reliability (8-d time interval) was high (ICC, >.80) for the isometric lifting tests, moderate (ICC range, .50–.80) for the dynamic lifting tests, and low (ICC, <.50) for the manipulation tests. The interrater reliability of the isometric and dynamic lifting tests (4-d time interval) was high (ICC, >.80), and it was moderate (ICC range, .50–.80) for both manipulation tests.

Conclusions: The isometric and dynamic lifting tests of the EK have a moderate to high level of reliability; the manipulation tests have a low level of reliability.

Key Words: Musculoskeletal system; Outcome assessment (health care); Rehabilitation; Reliability and validity.

THE MUSCULOSKELETAL SYSTEM has been identified as a major cause of work-related disorders and work loss throughout the world, particularly with regard to pain in the lower back, neck, and upper and lower limbs.[1-6] In the Netherlands, musculoskeletal disorders accounted for 6% of all health care costs in 1996. In addition, 36% of all people evaluated for work disability claims had either occupational disorders or injuries that were related to the musculoskeletal system.[7,8] In light of the enormous economic and financial consequences for society, it is imperative that the functional abilities of injured workers with musculoskeletal complaints are assessed.

Functional capacity evaluation (FCE) methods offer systematic, comprehensive, and multifaceted ways to measure current physical abilities of people who commonly have musculoskeletal complaints caused by work-related tasks.[9-13] In the Netherlands, the Ergo-Kit[a] (EK) FCE method is principally used in physiotherapy (PT) and rehabilitation centers to evaluate rehabilitation programs. It is also used in adjudicating work disability claims. The EK reports the functional physical capacity of injured workers objectively through the results of a battery of 55 standardized tests. It uses measurements and observations to evaluate the performance of subjects in tasks of both short and long duration.[14] Data reported during the tests includes blood pressure, heart frequency, load lifted, working height, working distance, manipulation velocity, coordination, degree of pain, and fatigue. This information is used to assess such work-related activities as reaching, lifting, carrying, and walking.[14]

Before FCE methods can be legitimately applied in rehabilitation centers, their psychometric properties should be defined.[15-17] This is particularly important in situations involving work disability claims because the test results can have major financial consequences for workers as well as both governmental and insurance entities. Innes and Straker[18,19] suggest that more research is needed in this area, and Gardener and McKennac[20] argue that the lack of documented reliability and validity diminishes the level of confidence in any FCE method. Although the EK is used in rehabilitation and PT settings, it information is lacking about its psychometric properties. A systematic literature review by Gouttebarge et al[21] failed to identify any studies on the reliability of the EK, even though this psychometric property is possibly the most important factor that determines the quality of a test measurement.[19,22,23] An assessment is considered reliable if its measurements are consistent, free from significant random error, and repeatable over time and across evaluators.[24,25] Of the various definitions of reliability, intra- and interrater reliability are most commonly associated with work-related assessment.[19,26] Intrarater reliability (also known as test-retest reliability, reproducibility, or repeatability) refers to the consistency or stability of test outcomes from 1 testing occasion to another, under the assumption that the characteristic being measured does not change over time.[19,27] Intrarater reliability is based on the estimation of

From the Coronel Institute for Occupational and Environmental Health, Academic Medical Center, AmCOGG: Amsterdam Center for Research into Health and Health Care, University of Amsterdam, Amsterdam, the Netherlands.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated.

Reprint requests to Vincent Gouttebarge, MSc, Coronel Institute for Occupational and Environmental Heath, Academic Medical Center/University of Amsterdam, PO Box 22700, 1100 DE Amsterdam, the Netherlands, e-mail: *v.gouttebarge@amc.uva.nl*.

0003-9993/05/8612-9904$30.00/0
doi:10.1016/j.apmr.2005.06.004

Table 1: Subject Characteristics

Subjects	n	Age (y) Mean ± SD	Age (y) Range	Height (cm) Mean ± SD	Height (cm) Range	Weight (kg) Mean ± SD	Weight (kg) Range
Men	15	41±16	22–64	176±7	164–193	76±11	58–90
Women	12	40±15	25–59	164±6	157–176	60±7	48–69
18–30y	15	27±2	22–30	171±9	157–193	66±11	48–86
45–65y	12	57±5	46–64	170±9	158–185	72±13	52–90
Total	27	40±15	22–64	171±9	157–193	69±12	48–90

Abbreviation: SD, standard deviation.

variance due to subjects and to the measurement of error.[24] In FCE, test-retest reliability is essentially the same as intrarater reliability because the rater's skill affects the accuracy of the test assessment.[17] Interrater reliability refers to the consistency of measures or scores made by raters, testers, or examiners on the same phenomenon.[19,27] It is based on the estimation of variances due to subjects, raters, and the measurement of error.[24]

Reliability studies should address populations that are relevant for the implementation of tests or instruments in the field. The EK has been used to assess subjects with and without musculoskeletal complaints (eg, at the end of their rehabilitation programs). Furthermore, it is unreasonable to assume that a test that is not reliable for healthy subjects could be reliable for subjects with musculoskeletal complaints. Our purpose in this study was to evaluate the intra- and interrater reliability of EK tests in adults without musculoskeletal complaints.

METHODS

Participants

Twenty-seven adults (15 men, 12 women) without musculoskeletal complaints participated in this study. All were employed either part time or full time in a variety of professions. Their mean age was 40±16 years (range, 22–64y), mean height was 171±9cm (range, 157–193cm), and mean body weight was 69±12kg (range 48–90kg). To be eligible to participate, subjects were required to meet the following inclusion criteria: (1) no acute (at time of the assessment) or chronic (no more than 3 complaints regarding the same body area in the past 3mo) musculoskeletal complaints,[28] and (2) age between 18 and 30, or between 45 and 65 years. (The second requirement was specified because people who are evaluated for work disability claims are usually either younger than 30 or older than 45[7]). To verify the inclusion criteria, each subject completed a checklist prepared for that purpose. Prior to enrollment, and after receiving verbal and written information on the study and procedures, subjects signed statements of informed consent. Subjects were free to quit the study at any time. Table 1 lists the subjects general characteristics.

Raters

A list of all the 24 raters who were certified for EK assessment in the Netherlands was obtained from the provider of this FCE method. All had completed the same training program, which consists of 4 days of instruction in the method and 12 hours of practice. Because the test assessments in this study were to be done in Amsterdam, 2 certified raters (R_1, R_2) from the Amsterdam area were selected for practical reasons. Both had between 3 and 4 years of experience in performing the EK assessments. We offered raters a small financial compensation and their expenses could be reimbursed.

Ergo-Kit Tests: Selection, Description, and Outcomes

From among the EK "physical agility" tests (manipulation, balance, strength, and endurance tests),[14] predictors were selected for 3 types of complaints: back, upper extremity, and lower extremity (table 2). Table 3 presents descriptions and outcomes of the EK tests. Standardized procedures were performed as described in the EK handbook.[14] Figures 1, 2, and 3 illustrate 3 of the 7 tests. The provider of the EK did not financially support this study.

Procedures

We used a within-subjects design to assess intra- and interrater reliability. Each subject was assessed on the EK at 3 different times (t1, t2, t3) twice by rater R_1 and once by rater R_2. Both raters were blinded to the other's test results, and rater R_1 was blinded to his prior test results during the second test assessment.

A time interval of 4±1 days was used between t1 and t2 and between t2 and t3, and a time interval of 8±2 days was used between t1 and t3. We assumed that the subjects' health status would be relatively stable between tests. Each subject was assessed at the same time of day as in the original test to avoid any effects related to a change in the time.[29]

As specified in the EK protocol,[7] the 7 tests were administrated in the following order: the isometric lifting tests (back-torso lift test [BTLT]; shoulder lift test [SLT]); the manipulation tests (forward manipulation test [FMT]; lower

Table 2: EK Test Predictors for Lower-Extremity, Back, and Upper-Extremity Complaints

Complaints	Lower Extremity	Back	Upper Extremity
Isometric strength	Back-torso lift test (BTLT)	Back-torso lift test (BTLT)	Shoulder lift test (SLT)
Manipulation ability	Lower manipulation test crouching (LMTC)	Forward manipulation test (FMT)	Forward manipulation test (FMT)
Dynamic strength	Carrying lifting strength test (CLST)	Lower lifting strength test (LLST)	Upper lifting strength test (ULST)