Upper lifting performance of healthy young adults in functional capacity evaluations

Gerrits, EHJ; Reneman, MF

Published in:
Journal of Occupational Rehabilitation

Publication date:
2003

Journal of Occupational Rehabilitation, Vol. 13, No. 4, December 2003 (© 2003)

Upper Lifting Performance of Healthy Young Adults in Functional Capacity Evaluations: A Comparison of Two Protocols

S. IJmker,[1,3] **E. H. J. Gerrits,**[1] **and M. F. Reneman**[2]

The objectives of this study were to explore the concurrent validity of test results of upper lifting tasks of the Ergo-Kit FCE and the Isernhagen Work Systems (IWS) FCE. Seventy-one healthy young adults performed 5 upper lifting tests with at least 5 min of rest in between. The lifting tests included 3 standard protocols and 2 modified protocols. Three criteria for concurrent validity were established: 1) Pearson correlation higher than .75, 2) nonsignificant two-tailed t test, and 3) mean difference smaller than 5 kg. The results showed that none of the criteria were met for the standard protocols. For the modified protocols criteria 2 and 3 were not met. Individual differences larger than 10 kg were found for both standard and modified protocols. It was concluded that the standard protocols for upper lifting tasks of the Ergo-Kit FCE and the IWS FCE do not meet the criteria for concurrent validity and can, therefore, not be used interchangeably.

KEY WORDS: functional capacity evaluation; lifting; validity; disability determination.

INTRODUCTION

When determining work-related disability for purposes of compensation, it is necessary to select instruments that will realistically and accurately reflect an individual's functional ability in a work context (1). In the area of manual materials handling, physical overuse (physical task demands exceeding worker's physical capacity) has been postulated as a hypothesis for work-related disability (2,3). Functional capacity evaluations (FCEs) were developed to evaluate the work-related physical capacity of a person (4) and should thus be able to detect physical overuse. FCEs could, therefore, be used in the determination of work-related disability.

There is currently no FCE available that has published evidence in all areas of reliability and validity (5,6). In studies that have been recently published however, increasing evidence is available that portions of FCEs are reliable. Specifically, lifting tasks have shown good

[1]Institute for Movement Sciences, University of Groningen, Groningen, The Netherlands.
[2]University Hospital Groningen, Center for Rehabilitation, Haren, The Netherlands.
[3]Correspondence should be directed to S. IJmker, Institute for Research in Extramural Medicine, 'Vrije Universiteit' Medical Centre, Department of Social Medicine, Van der Boechorststraat 7, NL-1081 BT Amsterdam, The Netherlands; e-mail: s.ijmker.emgo@med.vu.nl.

1053-0487/03/1200-0297/0 © 2003 Plenum Publishing Corporation

reliability (7–9). A number of validity studies concerning FCEs have been published. Evidence of a moderate relationship between upper lifting measurements on an FCE (ERGOS work simulator) and clinical assessment of upper lifting strength has been produced (10). Two studies determined the predictive validity of FCEs. The results of a study of Fishbain *et al.* (11) indicated that the test results of patients with chronic pain during an FCE could not predict employment level. When patients passed specific tests without excessive pain level, they had a 75% chance of being employed 30 months after treatment. A second study concluded that overhead lifting made no significant contribution in predicting whether a patient returned to work or not (12). Overhead lifting was, however, significantly related to the level of return to work (full-time versus part-time work and modified versus not modified work).

It is currently unknown whether different FCEs can be used interchangeably. When different FCEs produce similar results, then generalization of research results may be possible, thereby creating a larger body of knowledge concerning the validity of FCEs. To explore the concurrent validity of FCEs, the test results of upper lifting tasks (i.e., lifting above waist level) of two commercially available FCEs were compared. In this study the ULS (Upper Lifting Strength) of the Ergo-Kit FCE[4] and the WOL (Waist-to-Overhead Lift) of the Isernhagen Work Systems FCE[5] (IWS FCE) were used. Both FCEs have sufficient construct validity, because they are largely based on the Dictionary of Occupational Titles (6). Regarding the content validity however, different operationalizations (i.e., test protocols) of the construct upper lifting are used, thereby challenging concurrent validity.

The test protocols differ in a number of ways: the type of box and handles, vertical height of lower and upper shelf, weight increments, number of lifts per set, and test termination criteria. The potential effect of these differences on lifting capacity is in part underlined by published guidelines on the safe weight limits of lifting tasks (13, 14). One of the most universally used guidelines is the NIOSH lifting equation which states that the following task factors influence the safe weight limit of a lifting task: vertical and horizontal starting position, vertical and horizontal displacement, frequency and duration of lifting, asymmetry of the trunk during lifting, and the type of handles used for the lifting task (13). The safe weight limits were developed in relation to the lifting capacity of workers. Lifting capacity was defined by biomechanical (compressive force at the lumbar spine), physiological (aerobic capacity), and psychophysiological criteria (perceived workload). It should be noted that the NIOSH lifting equation and other guidelines might be of limited use in the application to upper lifting tasks, because these guidelines rely solely on the load placed on the lower back and do not take into account the load placed on the upper extremities in the evaluation of lifting tasks.

To study the effect of differences in operationalization on the concurrent validity of the standard protocols for peak lifting performance, modified protocols have been created in which the test termination criteria and the number of lifts per set have been equalized for both tests. Specifically, the following questions should be answered in this study:

1. Are the standard protocols for peak lifting performance concurrent valid?
2. Does the concurrent validity improve when the test termination criteria for peak lifting performance are equalized?
3. Does the concurrent validity improve when the numbers of lifts for peak lifting performance are equalized?

[4]Ergo Control BV, Buurserstraat 214-216, 7544 RG Enschede, The Netherlands.
[5]Isernhagen Work Systems, 1015 E. Superior Street, Duluth, MN 55802.

METHODS

Subjects

A total of 72 healthy young adults volunteered to participate in this study. One subject did not perform the WOL and was excluded for data analysis, leaving a study population of 71 subjects (35 males and 36 females). Mean age of the males was 23 years (range, 19–28 years), mean height was 1.86 m (range, 1.67–2.10 m), and mean weight was 79 kg (range, 59–113 kg). Mean age of the females was 22 years (range, 19–28), mean height was 1.74 m (range, 1.61–1.89 m), and mean weight was 70 kg (range, 56–101 kg). Most of the subjects were students. Sixty-seven subjects participated in athletic activities (mean 5 h per week, range, 0.5–11 h per week). All subjects signed informed consents prior to testing, declaring to be healthy and able to perform the lifting tests at their own risk.

Materials

A lab situation was established with the materials needed for the lifting tests. Materials needed for both tests include a measurement frame with two adjustable shelves (2.5 cm increments), a stopwatch and a heart rate monitor. The unique materials used for each test are presented in Table I. The box used for the WOL was modified; for this test a commercially available box was used.

The Ergo-Kit FCE uses a Rating of Physical Difficulty (RPD) scale to estimate the subject's perception of workload and body part discomfort during the lifting test. This scale was derived from an original version (15) and translated into Dutch. The scale is similar to Borg's CR-10 scale (16). The company that owns the rights of the Ergo-Kit modified some of the verbal expressions. In contrast to the original scale, an 11-point scale ranging from 0 ("nothing at all") to 10 ("maximum") was created (17).

Procedures

General

Body weight and height were measured prior to the lifting tests. All the subjects then subsequently performed five isoinertial upper lifting tests: two standard protocols, two modified protocols for peak upper lifting performance, and one submaximal upper lifting endurance (ULE) test. In Table I the standard protocols and three modified protocols are presented. Additionally, in the case of the ULS standard, subjects were not allowed to stop at RPD 7, but had to continue lifting until one of the test termination criteria of the WOL was met. Tests were terminated by the evaluator when subjects lacked full control over the movements of the box, when subjects displayed a lack of balance, and when subjects used other body parts than the arms to move the box: strong impulses from the trunk and use of the chest to put the box on the upper shelf were not allowed. Subjects always began with a test of the modified protocols, followed by the tests of the standard protocols and they ended with the remaining test of the modified protocols. ULS and ULE were always coupled as two following tests. Subjects were randomly assigned to one out of four testing orders. All subjects had to use the handles of the boxes during lifting.

Table I. Materials and Procedures for Test Protocols of Upper Lifting Strength (ULS) and Waist-to-Overhead Lift (WOL)

	ULS	WOL
Materials		
Box size	39 × 29 × 11.4 cm	43 × 35 × 23 cm (modified, see text)
Handles	Cylindrical handles, length = 17 cm, circumference = 8 cm, pivot angle = 10° from horizontal	Cutout handles, length = 11 cm, wide = 5 cm, horizontal placement
Weights	Weight bags of 2.5 and 5 kg	Metal weights of 2 and 4 kg
Rating scale	Rating of Physical Difficulty (RPD) Scale	None
Procedures		
Starting height	Knuckle height (standing)	Waist height (elbows 90° flexed)
Ending height	Shoulder height	Crown height (top of the head)
Repetitions per set	1	5
Starting weight	5 kg (box + weight)	4 kg (box + weight)
Weight increments	2.5 or 5 kg (see text)	Manifold of 2 kg
Criteria weight increments	– Evaluator's observation of body mechanics – RPD – Unlimited number of sets	– Evaluator's observation of body mechanics – Maximum reached in 3–6 sets
Test continuation criterion	1 min of rest	Heart rate ≤ 70% of age-related maximum (modified, see text)
Test termination criteria	– Evaluator: unsafe body mechanics – Subject has reached maximum – Subject wishes to stop – RPD ≥ 7 ("very hard")	– Evaluator: unsafe body mechanics – Subject has reached maximum – Subject wishes to stop – Heart rate > 85% of age related maximum
Test result interpretation	Safe amount of weight up to 33 lifts per 8 h working day	Safe amount of weight for 1–5 % of an 8 h working day
Modified protocols	– ULS with test termination criteria of the WOL (ULS modified 1) – ULS with five repetitions per set and test termination criteria of the WOL (ULS modified 5)	– WOL with one repetition per set (WOL modified 1)

ULS Standard

General procedures are presented in Table I. After each lift heart rate, RPD, and observations relating to body mechanics were recorded. A new set with incremented weight began after 1 min of rest. The weight increments during the test were based on RPD score (see Materials section). When the RPD score during the ULS was lower than 3 ("moderate") and the subject used safe body mechanics (determined by evaluator), the weight increment was 5 or 2.5 kg (determined by evaluator). If the RPD was 3 or higher, the weight increment was 2.5 kg. The weight remained the same when unsafe body mechanics were used and RPD was lower than 7. When unsafe body mechanics were used during the following set, the test was terminated. When safe body mechanics were used, the test continued and

the criteria for weight increment were applied. Preliminary data showed good test–retest reliability in a worker's compensation population with a Pearson correlation of .92 (18).

WOL Standard

During the WOL the evaluator observed the lifting behavior of the subject. Based on these observations and the heart rate after each set, the weight increments were determined. The test was continued when heart rate was below 70% of age-related maximum (19). The maximum amount of weight had to be reached in 3–6 sets (20). A number of studies demonstrated good test–retest reliability of the WOL with ICC values of at least .75 and $r = .90$ (9,10,19,21) and good intra- and intertester reliability to determine level of lift effort with Cohen's kappas ranging from .62 to .88, ICC = .96, and 91–93% agreement (9,22,23).

Data Analysis

Descriptives, Pearson product–moment correlations, and dependent *t* tests (two-tailed) were used to analyze the data. To explore the difference between the test results for individual cases, the method proposed by Altmann and Bland (24) was used. First the limits of agreements between the two tests were calculated (i.e., the absolute values of the difference between which 95% of the study population scored) and second the 95% confidence intervals of these limits were calculated. Data regarding the ULE were not analyzed, because they were not subject of this study.

Data Interpretation

The criteria for concurrent validity were: 1) Pearson correlation higher than .75 (6), 2) nonsignificant two-tailed *t* test, and 3) mean difference smaller than 5-kg. The 5 kg criterion was based on the Dutch disability legislation in which the level of disability in the case of short-term lifting depends on categories with increments of 5 kg (25). Alpha level was set at .05.

RESULTS

General

Test results of all protocols are presented in Table II. The mean differences between 1 and 5 repetitions of the four testing orders were within 1.3 kg of the grand mean difference. This indicated that testing order did not have a relevant impact on results.

Peak Lifting Performance

Pearson correlation between ULS standard and WOL standard was .72 ($p = .000$) and the difference between the mean test results was significant ($t = 12.1$, $p = .000$). The mean difference between the test protocols (ULS standard − WOL standard) was 6.2 kg. The

Table II. Test Results (kg) of Study Population ($N = 71$)

Test protocol[a]	Mean	*SD*	Minimum	Maximum
ULS standard	25.4	6.1	15.0	40.0
ULS modified 1	30.2	7.8	20.0	50.0
ULS modified 5	24.4	6.2	12.5	42.5
WOL standard	19.2	5.3	8.0	30.0
WOL modified 1	24.8	6.7	14.0	38.0

[a] ULS standard: 1 repetition; ULS modified: 1 repetition, test termination criteria of WOL; ULS modified 5: 5 repetitions, test termination criteria of WOL; WOL standard: 5 repetitions; WOL modified 1: 1 repetition.

limits of agreement were −2.4 and 14.9 kg. The 95% confidence intervals of these limits were −4.3 to −0.7 kg for the lower limit and 13.1–16.7 kg for the upper limit.

Effect of Test Termination Criteria

When the test termination criteria of the ULS were equalized to the criteria of the WOL, the correlation between ULS and WOL was .85 ($p = .000$) and the difference between the mean test results was significant ($t = 21.8$, $p = .000$). The mean difference between the test results (ULS modified − WOL standard) was 11.0 kg. The limits of agreement were 2.5 and 19.5 kg. The 95% confidence intervals of these limits were 0.7–4.2 kg for the lower limit and 17.8–24.3 kg for the upper limit. Compared to the standard protocols, equalizing the test termination criteria lead to a stronger relationship, but a larger difference between the test results.

Effect of Number of Repetitions

When the number of repetitions per set was 1 repetition for both tests (and test termination criteria of the WOL were applied), the correlation between the test results was .90 ($p = .000$) and the difference between the mean test results was significant ($t = 13.2$, $p = .000$). The mean difference between the test results (ULS modified − WOL modified 1) was 5.4 kg. The limits of agreement were −1.5 and 12.3 kg. The 95% confidence intervals of these limits were −2.9 to −0.1 kg for the lower limit and 10.9–13.7 kg for the upper limit. Compared to the standard protocols, equalizing the number of repetitions per set to 1 repetition and the test termination criteria lead to a stronger relationship and a slightly smaller difference between the test results.

When the number of repetitions per set was 5 repetitions for both tests (and test termination criteria of the WOL were applied), the correlation was .77 ($p = .000$) and the difference between the mean test results was significant ($t = 11.0$, $p = .000$). The mean difference between the test results (ULS modified 5 − WOL standard) was 5.2 kg. The limits of agreement were −2.8 and 13.2 kg. The 95% confidence intervals of these limits were −4.4 to −1.1 kg for the lower limit and 11.6–14.9 kg for the upper limit. Compared to the standard protocols, equalizing the number of repetitions per set to five repetitions and the test termination criteria lead to a slightly stronger relationship and a slightly smaller difference between the test results.

DISCUSSION

This study was designed to explore the concurrent validity of the test results of upper lifting tasks of two commercially available FCEs. Three criteria were used to establish concurrent validity. For peak upper lifting performance none of the criteria were met, indicating an absence of sufficient concurrent validity between the standard protocols of the Ergo-Kit FCE and the IWS FCE. The data demonstrated individual differences up to 14.9 kg between the test results of the ULS and the WOL. Equalizing the test termination criteria created even larger differences between the test results of ULS and WOL, although the relationship between the test results was stronger. An explanation of this phenomenon is that the protocols measured the underlying construct more uniform, whereas larger systematic differences were created between the protocols. Correlation statistics measure the strength of a relation between two variables, not the agreement between them (24). The number of repetitions per set did explain a minor part of the difference between test results of ULS and WOL. When the number of repetitions per set was equalized to one repetition, the relationship between results became stronger, but the other two criteria for concurrent validity were still not met. Individual test results demonstrated differences up to 12.3 kg. Equalizing the number of repetitions per set to five for both protocols did not produce a stronger agreement between the test results of the protocols.

The test termination criteria and the number of repetitions per set could not explain the differences between the test results of the standard protocols for peak upper lifting performance. Testing order did not have a relevant impact on the differences between the test results. Two factors were not controlled in this study: the height of the upper shelf and the placement of the handles. Lifting to crown height during the WOL, as opposed to lifting to shoulder height during the ULS, forced subjects to greater flexion of the shoulders (approximately 90 and 45° respectively), thereby decreasing muscle length and limiting force generating capacity of the muscle filaments. In line with this, it has been demonstrated that the Maximal Voluntary Contraction (MVC) at an shoulder angle of 90° of flexion was 17% of the MVC at a neutral shoulder angle whereas the MVC for 60 and 30° of flexion were 24 and 50% of the neutral MVC (26).

An additional handicap in lifting to crown height, as opposed to lifting to shoulder height, was the greater amount of ulnar deviation of the wrists needed during lifting. Ulnar deviation greater than 20° has been associated with worker complaints and limited performance (27). The handles of the box used for the WOL were placed horizontal as opposed to the tilted placement of the handles of the box used for the ULS. Consequently, subjects would need more ulnar deviation during the WOL, even when lifting height would be the same. Further research is needed to verify the importance of lifting height, handle placement, and their interaction for upper lifting performance.

The strengths of this study are the large sample size and the use of modified protocols to explain (part of) the differences between the test results of the FCEs. A limitation of the study is the limited clinical relevance, because the study population consisted of healthy young adults, while FCEs are most often used for people with disabilities. Post hoc analysis showed that all subjects under study lifted at least 15 kg during the ULS while 14 females (39% of all females) and 1 male (3% of all males) lifted less than 15 kg during the WOL. Dutch legislation labels a lifting capacity of at least 15 kg with a frequency of 5 lifts per hour as “not disabled” for lifting (28). Theoretically these findings would implicate a disability status for 21% of the subjects under study when using the WOL, as opposed to 0% when

using the ULS. This implication for the study population should be interpreted carefully, because the effect of fatigue on the absolute values of the test results is unknown and *lifting height* is not defined in Dutch legislation. Additionally, the rationale or research supporting this criterion of 15 kg could not be traced. Further research is warranted to determine whether the differences found in this study can be generalized to a population of patients with limited lifting capacity and who are claiming compensation.

In The Netherlands physicians perform the determination of lifting capacity in the case of compensation claims. Their judgments are often based on an interview and a physical examination. Despite the lack of concurrent validity of the protocols, further research should address the question whether the input of FCEs improves the reliability and validity of the judgments of physicians in disability determination. Future study should contrast the differences in lifting capacity on the basis of the results of different FCEs to the differences based on interviews and physical examinations.

In conclusion, the test results for peak lifting performance of the Ergo-Kit FCE and the IWS FCE did not meet the criteria for concurrent validity. Consequently, the upper lifting tasks of the FCEs under study cannot be used interchangeably.

ACKNOWLEDGEMENTS

The authors thank Jan Plat of Ergo Control BV for test protocol training and for using the Ergo-Kit FCE, and Wietkse Kuijer for helping with data collection and data entry.

REFERENCES

1. Innes E, Straker L. A clinician's guide toe work-related assessments: 3. Administration and interpretation problems. *Work* 1998; 11: 207–219.
2. Ayoub MM, Selan JL, Liles DH. An ergonomics approach for the design of manual materials-handling tasks. *Hum Factors* 1983; 25(5): 507–515.
3. Liles DH, Deivanayagam S, Ayoub MM, Mahajan P. A Job Severity Index for the evaluation and control of fifting injury. *Hum Factors* 1984; 26(6): 683–693.
4. Isernhagen SJ. Functional capacity evaluation: Rationale, procedure, utility of the kinesiophysical approach. *J Occup Rehabil* 1992; 2(3): 157–168.
5. Innes E, Straker L. Reliability of work-related assessments. *Work* 1999; 13(2): 105–122.
6. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13(2): 125–152.
7. Lechner DE, Jackson JR, Roth DL, Straaton DV. Reliability and validity of a newly developed test of physical work performance. *J Occup Med* 1994; 36: 997–1004.
8. Gross DP, Battié MC. Reliability of safe maximum lifting determinations of a functional capacity evaluation. *Phys Ther* 2002; 82(4): 364–371.
9. Reneman MF, Dijkstra PU, Westmaas M, Goëken LNH. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Med Rehabil* 2002; 12(4): 269–275.
10. Dusik LA, Menard MR, Cooke C, Beach GN. Concurrent validity of the ERGOS work simulator versus conventional functional capacity evaluation techniques in a worker's compensation population. *J Occup Med* 1993; 35(8): 759–767.
11. Fisbain DA, Cutler RB, Rosomoff H, Khalil T, Abdel-Moty E, Steele-Rosomoff R. Validity of the Dictionary of Occupational Titles residual functional capacity battery. *Clin J Pain* 1999; 15(2): 102–110.
12. Matheson LM, Isernhagen SJ, Hart DL. Relationship among lifting ability, grip force and return to work. *Phys Ther* 2002; 82: 249–256.
13. Waters TR, Putz-Anderson V, Garg A, Pine LJ. Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics* 1993; 36(7): 749–776.
14. Mital A, Nicholson AS and Ayoub MM. *A guide to manual materials handling, 2nd ed.* London: Taylor & Francis, 1997.

15. *Ergo-Kit handleiding 2.0* (in Dutch). Bussum: PGB, 1997.
16. *Handleiding Ergo-Kit* (in Dutch). Enschede: Ergo Control, 2000.
17. Borg G. *Borg's perceived exertion and pain scales*. Champaign, IL: Human Kinetics, 1998.
18. Schellekens J. *Beweging als warming-up voor werkhervatting: tussenrapportage* (in Dutch). Amsterdam: Lisv, 2001. Unpublished data.
19. Matheson NL, Mooney V, Grant JE, Affleck M, Hall H, Melles T, Lichter RL, McIntosh G. A test to measure lift capacity of physically impaired adults. Part 1-Development and reliability testing. *Spine* 1995; 20(19): 2119–2129.
20. Isernhagen Work Systems. *Functional capacity evaluation manual*. 1st edn. Duluth, MN, 1997.
21. Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens, JW, Goëken LNH. Test–retest reliability of the Isernhagen work systems functional capacity evaluation in patients with chronic low back pain. *J Occup Rehabil* 2003; 13: 207–218.
22. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgments of level of lift effort in a kinesophysical functional capacity evaluation. *Work* 1999; 12: 145–150.
23. Reneman MF, Jaegers SMHJ, Westmaas M, Goëken LNH. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *Work* 2002; 18: 23–27.
24. Bland JM, Altman, DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 8: 307–310.
25. *Functionele mogelijkheden en voorwaarden voor het verrichten van arbeid. Instrument voor verzekeringsarts en arbeidsdeskundige* (in Dutch). Amsterdam: Lisv, 2001. (www.lisv.nl/uwv/content/wetten/cbbs/FMLdecember2001.PDF)
26. Kim JY, Chung MK, Park JS. Measurement of physical work capacity during arm and shoulder lifting at various shoulder flexion and ad/abduction angles. *Hum Factors Ergon Manufact* 2003; 13(2): 153–163.
27. Drury CG, Pizatella T. Hand placement in manual materials handling. *Hum Factors* 1983; 25(5): 551–562.
28. *CBBS Normaalwaarden* (in Dutch). Amsteram: Lisv, 2001. (www.lisv.nl/uwv/content/wetten/cbbs/Normaal waarden.PDF)