

Frequently Asked Questions: Functional Capacity Evaluations in the Employment Testing Process

General Issues

Have FCEs been shown to reduce injuries and decrease medical costs?

This is unknown as no reports demonstrating this claim have been published in the peer-reviewed literature. Job-related physical ability tests *have* shown this result (Harbin & Olson, 2005; Keyserling et al., 1980a & b; Biering-Sorensen, 1984; Arnold, et al., 1982; Reimer et al., 1994; Chaffin, 1974; Randolph, 2000; Rayson, 2000; Toeppen-Sprigg, 2000), but these are job-specific physical ability tests and not FCEs.

Are FCEs an efficient form of employment testing?

It is unlikely. Several proprietary FCEs are based on the 20 factors from the *Dictionary of Occupational Titles*. Few jobs require all of these factors in significant degrees. Yet many FCEs test for all of these factors in large test batteries. Tests can be as long as two days in length and be very expensive. Research suggests that a more focused approach directed at the specific job is more effective (Gaudino, et al., 1999, 2001; Matheson, 2001).

We use a physical therapy clinic to conduct Functional Capacity Examinations for our company. Are they qualified to validate and perform these tests?

It depends. If they have additional education and experience in exercise science, industrial psychology, and tests and measurement they could be. Jahn, Cupon, & Steinbaugh (2004) have questioned the credentials and experience of the therapists who provide proprietary FCE services. He noted that they lack the education, experience, or training to access testing equipment and that employment testing is not closely related to the therapeutic arts, but rather to psychology, biomechanics, kinesiology, physics, physiology and exercise science.

Jahn and his colleagues also questioned whether therapist training and education acquainted them with the professional testing standards and the Civil Rights Act (as amended) and the Americans with Disabilities Act of 1990.

There is a serious question as to the education and training of physical therapists that decide to practice outside of their area of training. For example, the author conducted a review of a sample ten (10) of major undergraduate and graduate physical therapy programs. In every case, the course catalogs showed no courses in industrial psychology, educational psychology, human factors, job analysis methods, psychometrics, tests and measurement, statistics, or legal issues in employment testing (MED-TOX, 2013).

Validity & Reliability Issues

Isn't there a great deal of published information that shows that FCEs are reliable and valid tests?

The major comprehensive reviews of all proprietary FCEs have concluded that both reliability and validity in the vast majority of FCEs is unproven (Mahmud, et al. 2010, Gouttebauge, et al., 2009, Pransky, et al., 2004; King, et al., 1998; Innes & Striker 1999a & 1999b).

What is a reliable FCE?

Reliability research with FCEs typically means that two independent observers will rate the performance of a job applicant in a similar fashion. Correlation coefficients are typically used to make this assessment. Correlation coefficients are used but are probably inappropriate statistics for this use. For example, the following pairs of ratings (2,3; 3,4; and 5,6) have a perfect correlation of 1.0 but the ratings are not the same. Correlation coefficients, alone, do not prove that the ratings are accurate.

For example, Brouwer, et al., (2003) found rater the evaluators' estimates to vary 19.8 kg in a group with a mean performances 31-29.3 kg. These are remarkably high error rates. Brouwer attributed this variance to the testing procedure, differences in interpretation of the evaluators, measurement errors and random errors of the testing procedure.

An example of test reliability might be two consecutive scores by the same applicant on a Jamar Grip Strength device. The two scores will be very similar. Likewise, in a test of anaerobic power a person running the 100- yard dash on two consecutive tries would be expected to have very similar scores on each trial. These ability tests have extremely high documented reliability. These are examples of actual test reliability *and not* observer reliability.

Are there good examples of validation studies using FCEs?

Validation studies must be consistent with the principles and requirements of the professional standards (APA, SIOP) and the UGESP. Literature reviews of FCE practices (Gibson, et al., 1997; Newton, et al., 1993; Strong, et al., 1996; Velozo, 1993), and marketed evaluation systems (Innes & Straker, 1999a; 1999b; King, et al., 1998; Pranksy, et al., 2004) are unanimous in their conclusions: there is no empirical evidence to support one assessment approach or in system over another and no research to establish the predictive validity of these assessment practices.

The evaluation of assessment approaches articulated in training manuals (Blankenship, 1994; Isernhagen, 1992; Key, 1986; Lechner, Roth, & Straaton, 1991; Matheson, 1990) or professional programs is limited to observer reliability studies of selected instruments or measures.

Even those studies that are published have significant design flaws and/or are limited in scope and focus (Beaton, et al., 1995; Cooke, et al., 1994; Dusik, Menard & Cooke, 1993; Matheson, et al., 1995; Menard, et al., 1994; Newton, et al., 1993). Existing

studies have evaluated components of the assessment procedures rather than the overall assessment and a person's ability to perform a job at work.

Moreover, FCEs generally fail to predict whether a person is even capable of returning to any work at all (Gross & Battié, 2005). Predicting future performance on a specific job is likely a step too far. In another study, the FCE could not detect "clinically significant" changes in subjects' following participation in a work hardening program (Durand, et al., 2008).

To date, no FCE has been validated for any specific job and published in the peer-reviewed literature.

The physical therapist that validated our company's test did not provide a written validation study. Is this a problem?

Only if you are approached by a private attorney who specializes in employment law, a state FEP regulatory agency or the EEOC. Since Griggs (1971), all tests must be related for the job in question and consistent with business necessity. The failure of persons with real or perceived disabilities, females, and since March 30, 2005 (Smith v. City of Jackson) persons over age 40 all may have questions as to how the validity study was conducted which resulted in their exclusion from a particular job.

In one recent case, the "validation study" consisted of a physical therapist meeting with a supervisor in a parking lot for two hours. After the meeting, the therapist listed some physically demanding job tasks and made a set of recommendations for a FCE and suggested a local clinic for testing.

A long-term employee, upon his return to work after undergoing cancer treatment, failed the test. The employer was unable to document, or even explain, the relationship between the FCE and the job. The employer wisely chose to pay a large cash settlement before this case went to trial.

We supplement our FCE with work simulations. Isn't this a content-valid test?

Possibly. However, content-related tests can be very difficult to defend. Case law suggests that the job analysis must be extremely comprehensive and fully reflect the content of the job. Even with a comprehensive job analysis, work samples can have serious problems. First, if the test measures any skill that will subsequently be trained for, the test is an illegal employment practice. Second, work samples are subject to significant learning effects. Subjects who perform the same work sample on consecutive days can increase their performance by as much as 300%. This learning effect constitutes proof that the test is unstable and unreliable. Few work samples are ever examined for learning effects. This is a serious problem.

Work simulations typically require more test proctors, equipment, and materials from the job. They require more time to administer and can be difficult score. If a job is relatively complex involving several physical abilities, the test can be quite complex, time consuming, and expensive. Finally, few jobs require a worker be fully proficient

prior to a single day being spent on the job.

Job Analysis Issues

Our physical therapist uses the Dictionary of Occupational Titles (DOT) to characterize our jobs. Isn't this the best available tool?

The most recent job analyses that went into the Revised Fourth Edition of the DOT are dated 1990. As of 2012, 100 percent of the DOT definitions will be over 20-years old. A gradual change occurs in most jobs so that after a few years it is no longer realistic to expect that skills and abilities acquired in a job done then continue to apply. The DOT is obsolete was abandoned by the Department of Labor decades ago.

Thirty-three years ago the National Academy of Sciences reviewed the DOT in concluded that the worker functions, including the strength demands, SVP, and GED variables were not based in then-current vocational theory. This problem was not corrected in the last edition of the DOT.

The National Academy of Sciences wrote of these worker functions:

Rather, they are frozen in a now outmoded mold. Scales that more or less adequately reflected the state of the art of vocational trait measurement at mid-century are now outdated. This condition serves to underscore the urgency of adopting a new strategy in producing the DOT that includes as an intrinsic aspect continuous research and technical improvement of the document as a whole and of each of its components.

Likewise, the working condition and physical demand variables are essentially useless in the modern economy. They obviously were designed with unskilled factory and physical laboring jobs mainly in mind. As a consequence, they appear not to capture adequately the full range of variability in the working conditions and physical demands of jobs, omitting, for example, distinctions between machine-paced and worker-paced jobs, routine versus non-routine jobs, etc. (NAS, 1980, p. 168).

One of the chief problems with the DOT and the job analysis approach it uses is that it is simply inadequate to study the physical demands and working conditions of jobs. A single strength factor is used to cover all of the different types of human strength. The duration scales (0-33%, 33-67%, 67-100% of the time) are too crude to accurately measure either working conditions or physical demands.

What does the EEOC say about the DOT job analysis approach?

In 1992, the U.S. Department of Justice warned employers about using the DOT approach to job analysis in the Technical Assistance Manual to the Americans with Disabilities Act when it wrote:

Some job analysis methods ask current employees and their supervisors to rate the importance of general characteristics necessary to perform a job, such as "strength," "endurance," or "intelligence," without linking these characteristics to specific job functions or specific tasks that are part of a function. Such general information may not identify, for example, whether upper body or lower body "strength" is required, or whether muscular endurance or cardiovascular "endurance" is needed to perform a particular job function. Such information, by itself, would not be sufficient to determine whether an individual who has particular limitations can perform an essential function with or without an accommodation (DOJ, 1992, p. II-20).

In effect, DOJ recommended that a task analysis be performed that demonstrated a linkage for each task to a specific ability such as muscular strength, muscular power, muscular endurance or any other ability. A simple broad categorical rating of a job for "strength" was considered insufficient for purposes of job analysis under the ADA.

Cutoff Scores

In our test, applicants who fall below the normative data for the US population fail. Can normative data be used to set passing levels?

Normative data in FCEs is used in many different ways. For example, grip strength a female applicant's score might be compared to females of the same age. A male taking the test would be compared to the normative data for other males of his age. Passing levels may be set at some level below the normative value for average for each group.

A problem arises when normative data is applied in the employment context. For example, the mean score for a 20-29 year old man on the grip strength test is 125 lbs. This means that a male who scores below 125 lbs. is below the 50% cutoff level. For a female of the same age, a score of 76 lbs. would place her in the top half of her age group. Thus, a male could fail the test with a score of 125 lbs. and a female could pass with a score of 76 lbs.

This type of scoring procedure demonstrates that the test bears no relation to the job, but only to the values on a normative data table.

Legal Issues

My vendor says the FCE complies with Civil Rights law. Is this sufficient documentation?

The UGESP specifically warn employers that claims by test vendors or statements marketing brochures or other advertising material will carry no weight in the event of a challenged employment practice. This is especially important given that:

The College on Forensic Sciences (CFS) has identified that most FCE administrators are not sufficiently grounded in science, case law and forensic issues. Examples may include misquoting standard journal articles and texts, making false statements, providing "junk science" opinions and interpretation, and deliberately omitting important facts and knowledge (Jahn, et al., 2004).

Could my pre-offer test be illegal under the guidelines published by EEOC?

It's possible. To the extent that the answer is yes to any of the following questions, the FCE may violate the ADA as a prohibited pre-employment inquiry or medical examination.

According to the EEOC, an FCE may be a prohibited pre-employment inquiry if:

- Is it administered by a health care professional or someone trained by a health care professional?
- Are the results interpreted by a health care professional or someone trained by a health care professional?
- Is it designed to reveal impairment or physical or mental health?
- Is the employer trying to determine the applicant's physical or mental health or impairments?
- Is it invasive (for example, does it require the drawing of blood, urine or breath)?
- Does it measure an applicant's performance of a task, or does it measure the applicant's physiological responses to performing the task?
- Is it normally given in a medical setting (for example, a health care professional's office)?
- Is medical equipment used?

In many cases, a combination of factors will be relevant in figuring out whether a procedure or test is a medical examination. In some cases, one factor may be enough to determine that a procedure or test is medical.

Example: An employer requires applicants to lift a thirty-pound box and carry it twenty feet. This is not a medical examination; it is just a test of whether the applicant can perform this task. But, if the employer takes the applicant's blood pressure or heart rate after the lifting and carrying, the test would be a medical examination because it is measuring the applicant's physiological response to lifting and carrying, as opposed to the applicant's ability to lift and carry.

Below are some commonly asked questions about the ADA's restrictions on pre-offer

medical examinations.

May an employer require applicants to take physical agility tests?

Yes. A physical agility test, in which an applicant demonstrates the ability to perform actual or simulated job tasks, is not a medical examination under the ADA.

Example: A police department tests police officer applicants' ability to run through an obstacle course designed to simulate a suspect chase in an urban setting. This is not a medical examination.

May an employer require applicants to take physical fitness tests?

Yes. A physical fitness test, in which an applicant's performance of physical tasks -- such as running or lifting -- is measured, is not a medical examination. However, if an employer measures an applicant's physiological or biological responses to performance, the test would be medical.

Example: A messenger service tests applicants' ability to run one mile in 15 minutes. At the end of the run, the employer takes the applicants' blood pressure and heart rate. Measuring the applicant's physiological responses makes this a medical examination.

If we give the FCE post-offer instead of pre-offer, doesn't this remove our liability should the test be found discriminatory?

Because most FCEs are given post-offer, they more typically resemble pre-placement medical evaluations. Because the ADA states that medical examinations do not need to be job-related or consistent with business necessity, the use of these tests has flourished. It is the information and the use of the information derived from the FCE that must be job-related and consistent with business necessity. Therefore under disability law, it is the employer (who makes the final hiring decision based upon FCE information), who must justify the job-relatedness requirements of the ADA with regard to the FCE tests.

With regard to gender discrimination, any practice may be challenged at any point in the selection process. Since virtually all tests of strength have adverse impact against female applicants, FCEs cannot escape strict review in the event of a legal challenge.

Therefore, FCEs need documentation that the test is reliable, valid for the job in question, predictive of important work behaviors required on the job, and consistent with business necessity.

References

American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME) Joint Committee. (1999). *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association.

American College of Sports Medicine. (1991). *Guidelines for Exercise Testing and Prescription, 4th ed.* Philadelphia, PA: Lea & Febiger; 1991.

Arnold, J.D., Rauschenberger, J.M., Souble, W.G. & Guion, R.M. (1982). Validation and utility of strength test for selecting steelworkers. *J App Psych*, 67, 588-604.

Beaton, D.E., O'Driscoll, S.W., & Richards, R. (1995). Grip strength testing using the BTE work simulator and the Jamar dynamometer: A comparative study. *J Hand Surg*, 20A, 293-298.

Biering-Sorensen, F. (1984). Physical measurement as risk indicators for low back trouble over a one-year period. *Spine*, 11, 252-256.

Blankenship, K.L. (1994). *The Blankenship System Functional Capacity Evaluation: The Procedure Manual*. Macon, Georgia: The Blankenship Corporation.

Brouwer, S., Reneman, M.F., Dijkstra, P.U., Groothoff, J.W., Schellekens, J.M.H. & Geken, L.N.H. (2003). Test-retest reliability of the Isernhagen Work Systems functional capacity evaluation in patients with chronic low back pain. *J Occ Rehabil*, 13, 4, 207-217.

Campion, M.A. (1983). Personnel selection for physically demanding jobs: review and recommendations. *Pers Psych*, 36, 527-551.

Chaffin, D.B. (1974). Human strength capacity and low back pain. *J Occ Med*, 16, 248-254.

Chaffin, D.B., Herrin, G.D., & Keyserling, W.M. (1978). Pre-employment strength testing. *J Occ Med*, 67, 403-408.

Civil Rights Act of 1991, Public Law 102-166 - Nov. 21, 1991.

Cooke, C., Dusik, L., & Menard, M. (1994). Relationship of performance on the ERGOS Work Simulator to illness behaviour in a Worker's Compensation population with low back versus limb injury. *J Occ Med*, 36, 757-762.

Durand, M.J., Brassard, B., Hong, Q.N., Lemaire, J. & Loisel, P. (2008). Responsiveness of the physical work performance evaluation, a functional capacity evaluation, in patients with low back pain. *J Occup Rehabil*.18(1), 58-67

Dusik, L., Menard, M., & Cooke, C. (1993). Concurrent validity of the ERGOS Work Simulator versus conventional functional capacity techniques in a worker's compensation population. *J Occ Med*, 35, 759-767.

EEOC. Uniform Guidelines on Employee Selection Procedures. *Federal Register*, 43, Aug 25, 1978, 38290-38315.

Fleishman, E.A. (1964). *The Structure and Measurement of Physical Fitness*. Engelwood Cliffs, NJ: Prentice Hall.

Gaudino E., Mael, F. & Matheson, L. (1999). *A Literature Review: Functional Assessment and Related Construct Taxonomies*. Washington, DC: American Institute for Research.

Gaudino E, Matheson L., & Mael, F. (2001). Development of the functional assessment taxonomy. *J Occup Rehabil*, 11,155-175.

Gibson, L., & Strong, J. (1997). A review of functional capacity evaluation practice. *Work*, 9, 3-11.

Gouttebauge, V., Kuijjer, P.P., Wind, H., van Duivenbooden, C., Sluiter, J.K. & Frings-Dresen, M.H. (2009). Criterion-related validity of functional capacity evaluation lifting tests on future work disability risk and return to work in the construction industry. *Occup Environ Med*, 66(10), 657-63

Gross, D.P. & Battié, M.C. (2005). Functional capacity evaluation performance does not predict sustained return to work in claimants with chronic back pain. *J Occup Rehabil*, 15(3), 285-94.

Harbin, G. & Olson, J. (2005). Post-offer pre-placement testing in industry. *Am J Ind Med*, 47, 296-307.

Herrin, G.D., Jaraiedi, M. & Anderson, C.K. (1986). Prediction of overexertion injuries using biomechanical and psychophysical models. *Am Indust Hyg Assoc J*, 47, 322-330.

Hogan, J. (1991). The structure of physical performance in occupational tasks. *J Applied Psych*, 76, 495-507.

Innes E. & Straker, L. (1999a) Validity of work-related assessments. *Work*, 13, 125-152.

Innes E. & Straker, L. (1999b) Reliability of work-related assessments. *Work*, 13, 107-124.

Jackson, A.S., Osburn, H.G. & Laughery, K.R. (1984). Validity of an isometric strength test for predicting performance in physically demanding jobs. *Proceedings of the Human Factors Society 28th Annual Meeting*, 28, 452-454.

Jahn, W.T., Cupon, L.N. & Steinbaugh, J.H. (2004). Functional and work capacity evaluation issues. *J Chi Med*, 1, 1-5.

Key, G.L. (1986). *The Key Method of Functional Capacities assessment: Key Method User's Guide*. 1010 Park Avenue, Minneapolis, Minnesota.

Keyserling, W.M., Herrin, G.D. & Chaffin, D.B. (1980a). Establishing an industrial strength testing program. *Am Ind Hyg Assoc J*, 41, 730-736.

Keyserling, W.M., Herrin, G.D. & Chaffin, D.B. (1980b). Isometric strength testing as a means of controlling medical incidents on strenuous jobs. *J Occ Med*, 22, 332-336.

King, P.M., Tuckwell, N., & Barrett, T.E. (1998). A critical review of functional capacity evaluations. *Phys Ther*, 78, 852-866.

Lechner, D.E., Bradbury, S.F. & Bradley, L. (1998). Detecting sincerity of effort: a summary of methods and approaches. *Phys Ther*, 78, 8.

Lechner, D.E., Roth, D.L., & Straaton, K.V. (1991). Functional capacity evaluation in work disability. *Work*, 1, 37-47.

Liles, D.H., Deivanayagam, S., & Ayoub, M.M. (1984). A job severity index of the evaluation and control of lifting injury. *Hum Fact*, 26, 683-693.

Mahmud, N., Schonstein, E., Schaafsma, F., Lehtola, M.M., & Fassier, J.B., et al. (2010). Functional capacity evaluations for the prevention of occupational re-injuries in injured workers. *Cochrane Database Syst Rev.*, 7(7), 72-90.

Matheson L., Kaskutas V., & McCowan, S., et al., (2001). Development of a database of functional assessment measures related to work disability. *J Occup Rehabil*, 11, 177-199.

Matheson, L.N., Mooney, V., & Grant, J.E. (1995). A test to measure lift capacity of physically impaired adults. Part 1: Development and reliability testing. *Spine*, 20, 2119-2129.

Matheson, L. (1990). *Industrial Rehabilitation Resource Book. Performance Assessment and Capacity Testing*. Rancho Santa Margarita, CA

Menard, M.R., Cooke, C., Locke, S.R., Beach, G.N., & Butler, T.B. (1994). Pattern of performance in workers with low back pain during a comprehensive motor performance evaluation. *Spine*, 19, 1359-1366.

MED-TOX (2013): *Curriculum Review of Universities with Doctorate Degree Programs in Physical Therapy*: Western University, CA; University of Southern California, CA; Loma Linda University, CA; University of California at San Francisco, CA; Ithica College, NY; Chatham University, PA; University of Kentucky, KY; University of Toledo, OH; Husson University, ME; College of Health & Rehabilitation, MA.

National Academy of Sciences (1980). *Work, Jobs, and Occupations: A Critical Review of the Dictionary of Occupational Titles*. National Academy Press. <<http://www.nap.edu>>.

Newton, M., & Waddell, G. (1993). Trunk strength testing with iso-machines. Part 1: Review of a decade of scientific evidence. *Spine*, 18, 801-811.

Newton, M., Thow, M., Somerville, D., Henderson, I., & Waddell, G. (1993). Trunk strength testing with iso-machines. Part 2: experimental evaluation of the Cybex II Back Testing System in normal subjects and patients with chronic low back pain. *Spine*, 18(7), 812-824.

Pransky, G.S. & Dempsey, P.G. (2004). Practical aspects of functional capacity evaluations. *J Occup Rehabil*, 14, 217-229.

Randolph, D.C. (2000). Use of functional employment testing to facilitate safe job placement. *Occup Med*, 15, 81-122.

Rayson, M.P. (2000). Fitness for work: the need for conducting job analysis. *Occup Med*, 50, 434-436.

Reimer, D.S., Halbrook, B.D., & Edreyfuss, P.H. (1994). A novel approach to pre-employment worker fitness evaluations in a material-handling industry. *Spine*, 19, 2026-2032.

Society of Industrial and Organizational Psychology (2003). *Principles for the Validation and Use of Personnel Selection Procedures*. 4th Edition. College Park, MD.

Strong, S. & Westmoreland, M. (1996). *Determining Claimant Effort and Maximum Voluntary Effort Testing*. RECDAC Report for the Institute for Work and Health and the Ontario Insurance Commission. Hamilton, ON: Work Function Unit, School of Rehabilitation Science, McMaster University.

Toeppen-Sprigg, B. (2000). Importance of a job analysis with functional capacity matching in medical case management. *Work*, 15, 133-137.

U.S. Department of Justice, Civil Rights Division, Office of the Americans With Disabilities Act. (1992). *The Americans With Disabilities Act: Title II Technical Assistance Manual*. Washington, DC.

U.S. Department of Labor, Employment and Training Administration. (1991). *Dictionary of Occupational Titles*. Washington, DC: US Government Printing Office.

Velozo, C.A. (1993). Work evaluations: Critique of the state of the art of functional assessment of work. *Am J of Occ Ther*, 47, 203-209.

Waddell, G., McCulloch, J.A., Kummel, E. & Verner, R.M. (1980). Nonorganic physical signs in low back pain. *Spine*, 5, 117-125.

Waddell, G., & Main, C.J. (1984). Assessment of severity of low back disorder. *Spine*, 9, 204- 208.