

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/8549146>

Practical Aspects of Functional Capacity Evaluations

Article in *Journal of Occupational Rehabilitation* · October 2004

DOI: 10.1023/B:JOOR.0000022763.61656.b1 · Source: PubMed

CITATIONS

73

READS

4,427

2 authors:



[Glenn Pransky](#)

University of Massachusetts Medical School

190 PUBLICATIONS 10,794 CITATIONS

[SEE PROFILE](#)



[Patrick G Dempsey](#)

National Institute for Occupational Safety and Health

125 PUBLICATIONS 2,938 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



[ErgoMine App View project](#)



[Ergonomics in Mining View project](#)

Practical Aspects of Functional Capacity Evaluations

Glenn S. Pransky^{1,3} and Patrick G. Dempsey²

Physicians, employers, insurers, and benefits adjudicators often rely upon functional capacity evaluations (FCEs) to determine musculoskeletal capacity to perform physical work, often with legal or occupational consequences. Despite their widespread application for several decades, a number of scientific, legal, and practical concerns persist. FCEs are based upon a theoretical model of comparing job demands to worker capabilities. Validity of FCE results is optimal with accurate job simulation and detailed, intensive assessments of specific work activities. When test criteria are unrelated to job performance, or subjective evaluation criteria are employed, the validity of results is questionable. Reliability within a subject over time may be adequate to support the use of serial FCE data collection to measure progress in worker rehabilitation. Evaluation of sincerity of effort, ability to perform complex or variable jobs, and prediction of injury based upon FCE data is problematic. More research, especially studies linking FCE results to occupational outcomes, is needed to better define the appropriate role for these evaluations in clinical and administrative settings.

KEY WORDS: disability evaluation; work capacity evaluation; physical fitness; work physiology.

INTRODUCTION

Physicians, employers, insurers, and benefits adjudicators often rely upon functional capacity evaluations (FCEs) to provide definitive answers in a variety of situations involving physical work. These evaluations are quite common—over half a million formal evaluations of impairment and ability to work are conducted each year within the U.S. workers' compensation system, many including FCEs (1). Results of these evaluations have significant implications for further rehabilitation efforts, employment, compensability determinations, and cash benefits. Despite their widespread application for several decades, a number of scientific, legal, and practical concerns persist. This paper will examine these issues, on the basis of available scientific evidence and practical experience, focusing on those FCEs designed to evaluate ability to perform physical work.

¹Liberty Mutual Research Institute for Safety, Center for Disability Research, Hopkinton, Massachusetts.

²Liberty Mutual Research Institute for Safety, Center for Safety Research, Hopkinton, Massachusetts.

³Correspondence should be directed to Glenn S. Pransky, Center for Disability Research, 71 Frankland Road, Hopkinton, Massachusetts 01748; e-mail: glenn.pransky@libertymutual.com.

Current Scope and Use of FCEs

FCEs include a wide range of activities (2). The simplest evaluations involve a series of standardized tasks with measured weights and distances, and a trained observer; these are available for upper extremity as well as back/lower extremity activities. Other approaches use machines to measure peak force, velocity, and range of motion in several planes, with isometric and isokinetic techniques, for arm/hand, back and lower extremity function. In these situations, workers are generally asked to exert a maximal effort. Job simulation, using tasks and equipment specific to a particular job, has recently become more popular, in part due to the Americans with Disabilities Act requirement that valid testing should be job-specific and focus on a comparison of capacity to actual job demands (3).

One of the earliest applications of FCEs was in the preplacement setting, to identify individuals at increased risk of injury in physically demanding jobs. Prior medical approaches, such as X-rays or lumbar range of motion, had failed to provide useful information about risk for future work injury (4). Using an isometric FCE testing protocol based on biomechanical similarity to strenuous job tasks, Chaffin *et al.* demonstrated that those hired who had marginal strength in comparison to job demands were three times more likely to have a back injury at work, compared with those who had the highest relative strength compared to job demands (5). Subsequently, there was a proliferation of isometric testing devices, followed by development of machines to evaluate dynamic strength during movement, for use in preplacement screening evaluations. One variation that developed was the use of FCEs on a periodic basis to certify continued ability to perform infrequent but physically demanding tasks, such as firefighting.

The principles of measuring ability to perform a job were extended to postinjury populations. These evaluations were designed to determine work capacity in relation to a specific job or class of jobs, as well as level of consistent effort and cooperation. The results of these evaluations are frequently used to direct treatment and rehabilitation efforts, and in legal proceedings, to determine work capacity and eligibility for indemnity benefits.

THEORETICAL BASIS OF FCEs

A brief review of the theoretical basis of FCEs is important to understand how the design of FCEs relates to their intended purposes, and the scientific evaluations of these tests. The concept of matching job/workplace demands to the capabilities and limitations of a worker is a fundamental assumption underlying FCE application. Figure 1 illustrates one model that describes the relationship between job demands and worker capabilities, including the components of each construct (6). Job demands include the workers' physical and organizational surroundings. The capabilities and limitations of the worker are expressed in terms of what have been called the "limiting subsystem"—the human aspect corresponding to the predominant mismatch between job requirements and worker performance (7). Ideally, adjustments (at a worker or job level) can lead to a favorable ratio of worker capacity to job demands, and subsequently to safe and productive work that is sustained for long periods (health), as illustrated in Fig. 2 (8).

One of the most important aspects of an FCE is that the measurement of capacity is specific to the demands posed by the job. Given that most capacity measures, whether they

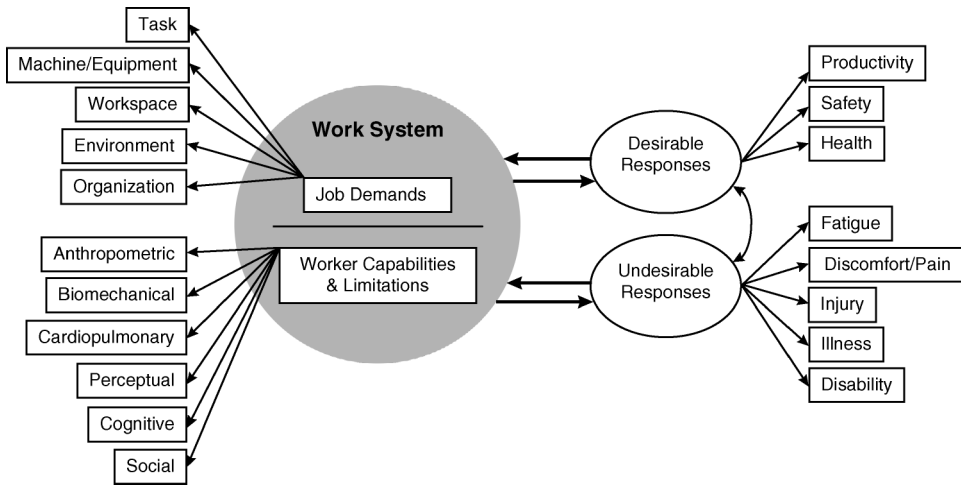


Fig. 1. Conceptual model of ergonomics practice (adapted from Dempsey *et al.* (6)).

be measures of aerobic capacity or muscle strength, are highly task-specific, this implies that the test may need to consider multiple tasks within a job. Thus, before any FCE is administered, a job analysis is required. This is critical from the standpoint of both scientific and legal perspectives. The task-specific nature of human capacity is further complicated by the fact that capacity can change because of injuries, aging, and other influences.

Although the concept of functional capacity relative to specific job demands is fairly straightforward, actual evaluation of functional capacity is a technically challenging process that often occurs within a complex legal and medical context. Because of the dynamic and complex nature of most job demands, as well as the dynamic nature of capacity due to morbidity, functional capacity is necessarily dynamic. This potential for variation presents a challenge to another conceptual basis for the use and interpretation of FCE results, that of scientific certainty. FCEs are often regarded as capable of providing data that is definitive in both measurements of capability as well as sincerity of effort, with accurate projections

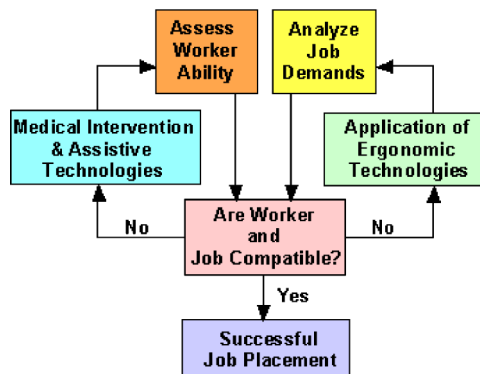


Fig. 2. A model of ergonomic evaluation rehabilitation, and person–job matching (from Armstrong *et al.* (8)).

to actual ability to return to specific jobs. However, as discussed below, the validity of FCE results and associated conclusions about sincerity of effort and work capacity are questionable in most situations, and thus present important limitations to application of FCE results (9).

VALIDITY

Validity primarily relates to the relationship between the FCE result and return to work or safe continuation of work as the primary outcome of concern; this is usually referred to as criterion or predictive validity. One of the most comprehensive investigations of RTW criterion validity was conducted by Dusik *et al.* (10). They compared FCE results using a standardized testing protocol for job simulation, and then followed RTW outcomes of a functional rehabilitation program postdischarge. The FCE appeared to be less accurate than the job simulation, and consistently underestimated actual ability to do the job, unless the job required only simple, repetitive motions that were similar to those encountered in the FCE protocol. Christian *et al.* reached a somewhat different conclusion, in a study of persons judged to be employable after a formal work capacity assessment related to indemnity compensation benefits in New Zealand (11). Of those judged employable but not working at follow-up (57% of the 141 participants in the study), some had repeat or reopened claims, possibly indicating a return to work at jobs that placed them at risk for further injury. Others have also observed similar findings, where limitations documented in the evaluation setting do not correlate with ability to return to work; these discrepancies appear to be most problematic with static tasks, less so with dynamic tasks or job simulation (12,13). Of all commercially available FCE protocols reviewed recently by Innes and Straker, only the Physical Work Performance Evaluation had adequate documentation of validity, for a narrow range of jobs (14).

Validity problems are due in part to both poor characterization of job demands, and inaccurate measurement of a worker's actual performance capability in relation to these demands. FCEs are generally based on an engineering—stress/strain model, where the physical job demands are compared with the measured work capacity of the individual tested to determine whether the “actual capacity” meets the “requirements” of the job. Barring a thorough job analysis, the job requirements are often extrapolated from the job title, and broad classifications of work requirements typically associated with the job, derived from the U.S. Department of Labor's O*NET database, or its predecessor, the Dictionary of Occupational Titles. Both systems provide rankings for each job group of the typical relative levels of demands along several dimensions (15).

For example, the DOT classifies “frequency of activity” into three categories: occasional (1–33% of the time), frequent (34–66% of the time), and constant (67–100% of the time). Job categories by lifting requirements are classified as shown in Table I.

O*NET provides rankings for relative level of job demands across hundreds of dimensions, for several thousand occupational titles. Dimensions in O*NET that are relevant to FCEs and physical job demands include trunk strength, stamina, handling objects, and dynamic flexibility (16). Both systems were developed primarily as vocational counseling aids. The intent of the DOT was to categorize the physical requirements for each generic occupational title and provide a standard method to analyze and classify jobs. These systems

Table I. Dictionary of Occupational Titles Definition of Job Categories by Lifting Requirement (in Pounds)

DOR job category	Frequency of lifting (% of time)	Maximum occasional lifting (pounds)
Sedentary		<10
Light	10	20
Medium	20	50
Heavy	50	100
Very heavy	100	> 100

were not designed to analyze people and their performance. The estimates of job demands are inexact generalizations that have not been validated through empiric testing. For example, information on the method, mechanism, or quantity or variation in required lifting is absent in DOT. Thus, the likelihood is high that results based upon generalizations from DOT or O*NET will be inaccurate, unless very large discrepancies between performance and job requirements are noted. Innes and Straker (17) concluded that the absence of a formal job evaluation in over 60% of the FCEs they reviewed constituted an important threat to validity.

Thus, a formal job assessment is desirable for those FCEs intended to measure ability to work at a specific job. Several assessment systems are available, designed to interface with FCE protocols. However, accurate assessment of job demands can be challenging. Job modifications (both formal and informal), and complex tasks that can be performed in a variety of ways are important threats to validity (3,18). Workers frequently alter how a job is performed or implement informal accommodations in order to perform a job despite physical limitations. Discussion with the examinee regarding job requirements may be helpful, but workers may not always be able to provide reliable data about physical job demands (19). Standard job descriptions from employers can be equally inaccurate. When an FCE is being performed to assess ability to perform a broad class of jobs, a high degree of job-specific validity may not be required; however, evaluators should note that results could easily be misleading. For example, the authors have observed multiple employees within a facility that have a job title such as “material handler” or similarly vague title, but that have very different job demands in terms of the loads handled and the frequency of lifting. Thus, the validity of an FCE across workers in the same job title could vary.

Implementation and execution of an FCE protocol that accurately simulates job tasks and adequately tests limiting subsystems is difficult. The standardized tasks offered in most FCEs do not correspond to actual work demands, except for those jobs where tasks are few, simple, and regularly repeated. FCE protocols often extrapolate from tasks that are stereotypical, or performed at near-maximal levels for a short period of time, to predict ability to sustain job activities for a full workday and workweek. Extrapolation from maximal ability to perform occasional lifting to expected ability to perform frequent lifting on the job is a frequent practice that lacks a firm scientific basis (20). This is where the limiting subsystem becomes important, as the capacity to perform low-frequency, high load lifts taxes the musculoskeletal system, whereas highly repetitive tasks bring the cardiopulmonary system into consideration.

Performance on FCE tasks are often compared with population or coworker norms, as actual job force requirements (such as the amount of torque generation at the joints required for lifting a given load) are not often estimated during this process. One study noted that preplacement judgments based on very general, non-work-related tests that often measure aerobic capacity, with comparison against normative data, leads to discrimination against those who have higher body mass indices, even though they are capable of performing the requirements of the job (21). Isotonic or isokinetic tasks do not represent agility, coordination, or constrained postures required at work (20). Evaluator judgment can also be a source of low validity, especially when the test criteria have little relationship to job performance. Several studies have demonstrated that attitudes and prejudices of the examiner can also affect interpretation of results and outcomes (22,23). Observations of changes in body mechanics when lifting loads of increasing weight has been proposed as a criterion for maximal acceptable load. This kinesio-physical approach mandates that lifting methods be judged within safe guidelines, and evaluators extrapolate from safe body mechanics in test situations (24), despite absence of scientific support for its validity (25). Both overestimation as well as underestimation of actual ability can occur as a result (26). Abdel-Moty *et al.* concluded, and we concur, that general testing is of little value to measure actual capacity to perform a particular job, and that job-specific testing with direct linkage to job tasks is needed for valid results (15).

FCEs based on job simulation test only the physical components of the job, and fail to simulate the environmental (hot, cold, vibration) or psychological components (time pressure, working in isolation) (27). Thus, validation is difficult in some situations without strong evidence for job performance linkage around physical tasks (28). Even if physical job demands are accurately measured and appropriately simulated in the FCE setting, actual return to work is a function of not only physical demands and capacity, but also skill, motivation, workplace, and psychosocial factors. Using this standard, the validation of a particular FCE method is impossible without taking into account all the other factors that may affect a successful return to work (29).

There is better evidence that changes in FCE performance over time can represent actual, meaningful improvement in function (30–32). Whether small changes over time are important, and how much change is sufficient for RTW, is hard to determine (33). Improved isokinetic measures did not predict RTW in one longitudinal study (31). Poor or absent documentation of inter- and intrarater reliability for most FCE approaches leaves clinicians without a scientific basis for evaluating whether changes over time represent actual improvement or measurement error (34,35). Most studies of reliability have failed to carefully link results to actual ability to sustain work, instead focusing on body mechanics or other intermediate measures (36). Innes and Straker concluded that reliability was sufficient for medicolegal purposes in only a few tests. Furthermore, the practice of obtaining repeated functional measurements during the course of physical rehabilitation may represent an unnecessary expense that is not required to achieve optimal outcomes (37).

The FCE represents findings from a single point in time, and it is not practical to perform a reevaluation every time a change in function or work demands occurs (Fig. 2) (8). Thus, there is little justification to conduct formal FCEs with patients who are early in their recovery, when physical capacity and pain tolerance is changing, or when the full range of available job accommodations has not been explored. Some type of preliminary

assessment of functional capacity might be useful, however, at this stage, to set goals and establish a baseline from which further progress can be measured.

Using the results of the FCE to establish an arbitrary weight-lifting limit does not make sense from a biomechanical standpoint, as the maximum safe load is a function of several factors besides the weight of the object. Ideally, the FCE might provide information about the postures and situations that are most problematic for an individual recovering from an episode of low back disability, and thus provide suggestions for effective work modifications.

Although FCEs are often promoted as a method of “objectively” identifying conscious attempts to reduce effort, no conclusive scientific proof of discrimination ability across a range of injured subjects is available (38). One study reported high sensitivity and specificity of tests used to determine sincerity of effort, but only in subjects who were instructed to provide a very significant (50%) reduction of maximal force, without specific information as to what was being measured to determine sincerity (39). Other studies have demonstrated that subjects can reproducibly perform at voluntarily reduced strength levels (40). Little evidence exists for a threshold of coefficient of variation that is unacceptable; suggested levels range from 5 to 29% (41).

The typical variations in pain and function that accompany chronic low back pain may well account for observed variability in performance even in persons who are consistently providing a maximal tolerated effort. Reliability can be poor because of variations in pain, position, self-limitation to avoid injury, equipment function, testing protocols, subject comprehension, or ability to follow specific directions (42). Poor performance can be due to failure to understand the degree of effort required, anxiety related to the test situation, depression, pain, fear-avoidance, unconscious or conscious illness behavior or exaggeration, or malingering (43). Training and acclimation can also affect reliability over time. Significant reactivity (learning effect) has been demonstrated in LBP patients with an isokinetic protocol, resulting in variations of 17–28% (44). Patients may have reasonable fears about overexertion leading to subsequent reinjury. Thus significant variability may occur for many reasons besides insincere effort (45,46).

Waddell’s signs are often used to detect signs of exaggeration or voluntarily reduced effort, even though they were never tested in this capacity. These signs are more appropriately used to detect symptoms that occur without a specific organic basis, and thus positive results may be most indicative of those who will benefit from psychological intervention. Whether or not the presence of these signs is directly related to proven insincerity of effort in subsequent FCE testing is unknown (47). However, kinesiophobia was not linked to decreased FCE performance in one study, suggesting that certain FCE results may not be overly influenced by fear of pain (48).

Conversely, consistent self-limitation can be interpreted as a “valid” result, and overexertion (effort in a range that is unsafe for the individual) is also a possibility. Pain can greatly hinder FCE performance, so testing may actually provide a measure of pain tolerance, not peak functional capacity (49). Thus changes over time may reflect changed psychological or behavioral factors affecting pain tolerance, not muscle strength (50). Hazard compared several indices of subject effort, including isokinetic force/distance curve patterns, peak force variations, blood pressure, and heart rates. He concluded that even the best physiologic measures and force curve analysis are not as reliable as an expert observer in detecting voluntary self-limitation. Although most FCE reports are restricted to

“objective” findings, inclusion of self-reported data from subjects could help to explain apparent inconsistencies.

In essence, it is difficult to determine whether limitations are based on what the applicant cannot do versus what they will not do. It is important to distinguish between validity as a scientific concept (primarily external validity) and attempts to measure sincerity of effort (the latter term is preferred) often evaluated through measures of reliability. Mislabeling as insincere leads to misdiagnosis, improper treatment, increased litigation, and increased cost of care (41). For practical purposes, FCEs appear to be effective in detecting submaximal efforts only when variation is high and the lack of full effort is obvious.

Thus, it is not surprising that in an analysis of demographic and FCE data in 539 subjects, only a small amount of variance in relation to RTW outcome was explained by FCE results. There was considerable overlap in performance between those who did and who did not RTW, and only one FCE test (floor to waist lift) was necessary to achieve this level of discrimination. Demographic and injury-related factors (gender and length of disability) were more predictive of RTW (51). Ruan *et al.* compared FCEs that consisted of a simple submaximal series of functional tests (static and dynamic lifting tests) with a more extensive evaluation, along with psychological testing, for chronic back pain patients and healthy controls (52). The main predictors of functional status were psychological measures, and the added information from the physical testing was minimal. They concluded that more extensive testing was unnecessary and likely not valid. Although accuracy of RTW prediction may be acceptable for a group of persons, the level of accuracy for an individual may be low and unacceptable. Passive tests of impairment are available (such as evaluations according to the AMA Guides to Evaluation of Permanent Impairment), but these results do not correlate with return to work, except in cases of severe impairment (53).

Clinical examination and mechanical methods of strength testing may often be comparable in terms of the accuracy of information they yield about a subject's work performance in relation to their capacity (54,55). In many instances, a thorough clinical evaluation that includes a review of functional activities of daily living may be sufficient to determine readiness to return to work. Some investigators have argued that, absent an accurate work simulation, questionnaires have greater validity and sensitivity to important change in work capacity than “objective” evaluations of functional capacity (56).

PREPLACEMENT TESTING AND LEGAL CONSIDERATIONS

In the preplacement setting, FCEs may have a role in injury prevention, but this has been demonstrated only in jobs involving a high level of physical demands, where FCE tasks were similar to actual job demands (5). Preplacement screening using FCEs raises concerns about practicality and cost, unless the risk of injury is very high. Since the majority of disability in industry due to low back pain is associated with the few cases that become chronically disabled, an ideal test must predict not only those who will be more likely to develop an injury, but also develop chronic disability. Currently, none of the FCE tests in common use have been shown to predict both the occurrence of LBP and discern between the majority that will have a quick recovery versus those who will go on to prolonged disability (57). Thus, inappropriate selection and discrimination can easily occur, as well as incurring significant expense for FCEs on all prospective employees without equivalent benefits in terms of avoidance of injury-related expenses. Furthermore, if prospective workers are given

a heavy lifting load, described as a job requirement, most will try to lift it regardless of capability, with potential risk of resulting injury (58).

Laws to prevent discrimination against a class of job applicants require that such tests represent a valid simulation of the job. The Uniform Guidelines on Employee Selection Procedures (1978) (29 Code of Federal Regulations, Chapter XIV, Part 1607) were designed to provide a “framework for assessing the proper use of tests and other selection procedures.” These guidelines apply to all employee-selection procedures. More recent legislation, the Americans with Disabilities Act (ADA), applies specifically to individuals with disabilities. In the case of preplacement testing, both regulations must be considered when selecting a particular FCE and applying the results.

The Uniform Guidelines on Employee Selection Procedures specify requirements for the validity of selection test(s), including criterion-related, content, and construct validity. Any test must have documented evidence for at least one of these types of validity. Such evidence can come from experimental data showing that a test is predictive of or significantly correlated with elements of job performance (criterion-related validity), that the content of the test is representative of important aspects of performance on the job (content validity), or that the protocol measures the degree to which candidates have identifiable characteristics which have been determined important for successful job performance (construct validity) (14). The regulations state that when such a test cannot be conducted, that, selection procedures should be “as job related as possible.” Thus, aside from the scientific importance of properly assessing the job demands to worker capacity ratio discussed earlier, the same concept is critical for establishing a legal test.

The Uniform Guidelines on Employee Selection contain technical standards for validity studies. The validity study should include a review of the job, which may seem obvious when the goal is to assess the degree of match between job demands and worker capacity; however, not all FCE providers perform a job analysis. The job analysis is required to insure that measures of work behavior or performance have at least some relevance to the job. The first step should always be an accurate job description, from the perspective of both legal requirements and technical appropriateness. In fact, the job description is required when considering the ADA.

Title 1 of the ADA specifically addresses the timing, nature, scope, and use of results of FCEs, and significantly limits the ability of employers to require evaluations of those who are already employed. The Act protects those who have a significant disability, as well as those who are not actually disabled, but where an employer wrongly regards them as disabled (59). Agility tests, including measures of physical and functional capacity that might be part of an FCE, are allowed if they are consistently applied and job-related—in other words, have a valid relationship to ability to perform essential job functions. This is why the concept of a “generic FCE” should always raise concerns.

Thus employers who use FCEs of questionable validity to select workers or limit those who RTW after an injury may be subject to litigation based upon these antidiscrimination laws. Other legal considerations may include liability for injury occurring to patients as a result of FCE testing, and liability to employers for inaccurate results.

Practicality implies that the ease of FCE administration, acceptability, interpretation, and reporting are all reasonable, and that the benefits outweigh all associated costs. Extensive job simulation may be ideal from a legal and validity perspective, but prohibitively expensive in most cases. Safety is also a concern; for example, exacerbation of LBP is often noted with isometric tests (60). There is no infallible method to determine when FCE

maneuvers are unsafe, despite suggestions that body mechanics during lifting should be used (61).

It is not uncommon for organizations that have not previously used FCEs to be overly optimistic about the value of these tests, particularly when vendors provide a strong sales pitch. However, FCEs should only be considered as one component of a broader program that comprehensively addresses injury prevention and RTW. First and foremost, workplaces need to be designed to minimize mechanical stress on to workers, as well as to minimize pain and discomfort while performing their jobs after an injury occurs. Ergonomics efforts should be undertaken first to determine if the problematic job demands can be minimized or eliminated, as engineering control of risk is the ideal solution.

Criteria for acceptability of FCE use varies greatly by application. Accurate prediction of safe RTW outcomes, including reinjury, may be most important when RTW is the goal of an FCE. When the primary purpose is adjudication, the primary criterion for acceptability of an FCE approach may be quite different—in this instance, the function is administrative rather than rehabilitative, and thus consistency may be most important. In the third major instance of FCE use, the well-person preplacement evaluation, a different set of criteria might be most important—such as avoidance of discrimination, excessive cost relative to benefit, and predictive ability for future serious injury.

CONCLUSION

Several scientific and practical limitations are associated with FCEs. In a few instances, these have been overcome through thorough job analysis and careful work simulation, with protocols that closely parallel work activities, directed by expert evaluators, resulting in findings of reasonable certainty (62). Most FCEs do not achieve these standards of performance. Generally acceptable, accurate measures of voluntary self-limitation are not available. Until further research develops valid, reliable, and efficient measures that correlate well with safe and sustained return to work, FCEs will not be very helpful for practicing clinicians involved in return to work decisions. Perhaps it is appropriate to regard the majority of FCEs that are not conducted in relation to a specific job as primarily administrative exercises, allowing for a demonstration of a range of performance that is acceptable to the worker. In this context, the FCE may be of significant therapeutic value. Although the validity of FCE scores alone to accurately predict job performance is questionable, this form of observer-based functional evaluation may be helpful to chart improvements, specify functional disparities with respect to job demands, and identify nonmedical factors influencing the ability to work. Thus, when combined with other sources of information, FCE results may ultimately contribute to resolving issues of compensability, disability, and employability. Some recently published research is beginning to address these concerns, and much more research is needed to broaden the scope of applicability of these tests (63).

ACKNOWLEDGMENTS

Special thanks to Lyn Dempsey for editorial support, and to Bill Shaw and Ray McGorry for scientific review.

REFERENCES

1. National Council on Compensation Insurance. *Countrywide workers' compensation experience including certain competitive state funds, first report basis*. Boca Raton, FL: NCCI, 1982.
2. Simpson SJ, Richlin D. The role of functional capacity evaluations in occupational health settings. *AAOHN J* 2003; 51(5): 202–203.
3. Hoffman S, Pransky G. Pre-placement evaluations and the Americans with Disabilities Act: An analysis. *J Occup Rehabil* 1998; 8: 255–263.
4. Bigos SJ, Battie MC, Fisher LD, Hansson TH, Spengler DM, Nachemson AL. A prospective evaluation of preemployment screening methods for acute industrial back pain. *Spine* 1992; 17(8): 922–926.
5. Chaffin DB, Herrin GD, Keyserling WM. Preemployment strength testing: An updated position. *J Occup Med* 1978; 20(6): 403–408.
6. Dempsey P, Ciriello V, Clancy E, McGorry R, Pransky G, Webster B. *Quantitative assessment of upper extremity capacity and exposure*. Proceedings of the IEA 2000/HFES 2000 Congress, 2000, Santa Monica, P5/724–727.
7. Sinclair M, Drury C. On mathematical modelling in ergonomics. *Appl Ergon* 1979; 10(4): 225–234.
8. Armstrong TJ, Franzblau A, Haig A, Keyserling WM, Levine S, Streilein K, Ulin S, Werner R. Developing ergonomic solutions for prevention of musculoskeletal disorder disability. *Assist Technol* 2001; 13(2): 78–87.
9. Innes E, Straker L. Reliability of work-related assessments. *Work* 1999; 13(2): 107–124.
10. Dusik LA, Menard MR, Cooke C, Fairburn SM, Beach GN. Concurrent validity of the ERGOS work simulator versus conventional functional capacity evaluation techniques in a workers' compensation population. *J Occup Med* 1993; 35(8): 759–767.
11. Christian B. Return to work outcomes following accident compensation corporation work capacity assessment. *N Z Med J* 2002; 115(1153): 209–211.
12. Ferguson SA, Marras WS, Gupta P. Longitudinal quantitative measures of the natural course of low back pain recovery. *Spine* 2000; 25(15): 1950–1956.
13. Dempsey PG, Ayoub MM, Westfall PH. Evaluation of the ability of power to predict low frequency lifting capacity. *Ergonomics* 1998; 41(8): 1222–1241.
14. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13(2): 125–152.
15. Abdel-Moty E, Fishbain DA, Khalil TM, Sadek S, Cutler R, Rosomoff RS, Rosomoff HL. Functional capacity and residual functional capacity and their utility in measuring work capacity. *Clin J Pain* 1993; 9(3): 168–173.
16. Peterson N, Mumford M, Borman W, Jeanneret P, Fleishman E, Levin K, Campion M, Mayfield M, Morgenson F, Pearlman K, Gowing M, Lancaster A, Silver M, Dye D. Understanding work using the occupational information network (O*NET). *Pers Psychol* 2001; 54(2): 451–492.
17. Innes E, Straker L. Workplace assessments and functional capacity evaluations: Current practices of therapists in Australia. *Work* 2002; 18(1): 51–66.
18. Chan G, Tan V, Koh D. Ageing and fitness to work. *Occup Med (Lond)* 2000; 50(7): 483–491.
19. Lindstrom I, Ohlund C, Nachemson A. Validity of patient reporting and predictive value of industrial physical work demands. *Spine* 1994; 19(8): 888–893.
20. Jones T, Kumar S. Functional capacity evaluation of manual materials handlers: A review. *Disabil Rehabil* 2003; 25(4–5): 179–191.
21. Bilzon JL, Allsopp AJ, Tipton MJ. Assessment of physical fitness for occupations encompassing load-carriage tasks. *Occup Med (Lond)* 2001; 51(5): 357–361.
22. Svensson T, Karlsson A, Alexanderson K, Nordqvist C. Shame-inducing encounters. Negative emotional aspects of sickness-absentees' interactions with rehabilitation professionals. *J Occup Rehabil* 2003; 13(3): 183–195.
23. Colella A, DeNisi AS, Varma A. The impact of ratee's disability on performance judgments and choice as partner: The role of disability-job fit stereotypes and interdependence of rewards. *J Appl Psychol* 1998; 83(1): 102–111.
24. Isernhagen S. Functional capacity evaluation: Rationale, procedure, utility of the kinesiophysical approach. *J Occup Rehabil* 1992; 2(3): 157–168.
25. Smith RL. Therapists' ability to identify safe maximum lifting in low back pain patients during functional capacity evaluation. *J Orthop Sports Phys Ther* 1994; 19(5): 277–281.
26. Ting W, Wessel J, Brintnell S, Maikala R, Bhambhani Y. Validity of the Baltimore therapeutic equipment work simulator in the measurement of lifting endurance in healthy men. *Am J Occup Ther* 2001; 55(2): 184–190.
27. Mazanec DJ. The injured worker: Assessing "return-to-work" status. *Cleve Clin J Med* 1996; 63(3): 166–171.
28. Schonstein E, Kenny DT. The value of functional and work place assessments in achieving a timely return to work for workers with back pain. *Work* 2001; 16(1): 31–38.
29. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther* 1998; 78(8): 852–866.

30. Curtis L, Mayer TG, Gatchel RJ. Physical progress and residual impairment quantification after functional restoration. Part III: Isokinetic and isoinertial lifting capacity. *Spine* 1994; 19(4): 401–405.
31. Hazard RG, Fenwick JW, Kalisch SM, Redmond J, Reeves V, Reid S, Frymoyer JW. Functional restoration with behavioral support. A one-year prospective study of patients with chronic low-back pain. *Spine* 1989; 14(2): 157–161.
32. Rainville J, Ahern DK, Phalen L, Childs LA, Sutherland R. The association of pain with physical activities in chronic low back pain. *Spine* 1992; 17(9): 1060–1064.
33. Boadella JM, Sluiter JK, Frings-Dresen MH. Reliability of upper extremity tests measured by the Ergos work simulator: A pilot study. *J Occup Rehabil* 2003; 13(4): 219–232.
34. Innes E, Straker L. A clinicians' guide to work-related assessments: 3—Administration and interpretation problems. *Work* 1998; 11(2): 207–219.
35. Newton M, Waddell G. Trunk strength testing with iso-machines. Part 1: Review of a decade of scientific evidence. *Spine* 1993; 18(7): 801–811.
36. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgments of level of lift effort in a kinesiophysical functional capacity evaluation. *Work* 1999; 12(2): 145–150.
37. Rainville J, Sobel J, Hartigan C, Monlux G, Bean J. Decreasing disability in chronic back pain through aggressive spine rehabilitation. *J Rehabil Res Dev* 1997; 34(4): 383–393.
38. Hazard RG, Reid S, Fenwick J, Reeves V. Isokinetic trunk and lifting strength measurements: Variability as an indicator of effort. *Spine* 1988; 13(1): 54–57.
39. Jay MA, Lamb JM, Watson RL, Young IA, Fearon FJ, Alday JM, Tindall AG. Sensitivity and specificity of the indicators of sincere effort of the EPIC lift capacity test on a previously injured population. *Spine* 2000; 25(11): 1405–1412.
40. Robinson M, Geisser M, Hanson C, O'Conner P. Detecting submaximal efforts in grip strength testing with the coefficient of variation. *J Occup Rehabil* 1993; 3: 45–50.
41. Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: A summary of methods and approaches. *Phys Ther* 1998; 78(8): 867–888.
42. Innes E, Tuckwell N, Straker L, Barrett T. Test–retest reliability on nine tasks of the Physical Work Performance Evaluation. *Work* 2002; 19(3): 243–253.
43. Hirsch G, Beach G, Cooke C, Menard M, Locke S. Relationship between performance on lumbar dynamometry and Waddell score in a population with low-back pain. *Spine* 1991; 16(9): 1039–1043.
44. Grabiner M, Jeziorowski J, Divekar A. Isokinetic measurements of trunk extension and flexion performance collected with the Biodex Clinical Data Station. *J Orthop Sports Phys Ther* 1990; 11: 590–598.
45. Croft PR, MacFarlane GJ, Papageorgiou AC, Thomas E, Silman AJ. Outcome of low back pain in general practice: A prospective study. *BMJ* 1998; 316(7141): 1356–1359.
46. van den Hoogen HJ, Koes BW, van Eijk JT, Bouter LM, Deville W. On the course of low back pain in general practice: A one year follow up study. *Ann Rheum Dis* 1998; 57(1): 13–19.
47. Waddell G, McCulloch JA, Kummel E, Venner RM. Nonorganic physical signs in low-back pain. *Spine* 1980; 5(2): 117–125.
48. Reneman MF, Jorritsma W, Dijkstra SJ, Dijkstra PU. Relationship between kinesiophobia and performance in a functional capacity evaluation. *J Occup Rehabil* 2003; 13(4): 277–285.
49. Beimborn DS, Morrissey MC. A review of the literature related to trunk muscle performance. *Spine* 1988; 13(6): 655–660.
50. Cooke C, Menard MR, Beach GN, Locke SR, Hirsch GH. Serial lumbar dynamometry in low back pain. *Spine* 1992; 17(6): 653–662.
51. Matheson LN, Isernhagen SJ, Hart DL. Relationships among lifting ability, grip force, and return to work. *Phys Ther* 2002; 82(3): 249–256.
52. Ruan CM, Haig AJ, Geisser ME, Yamakawa K, Buchholz RL. Functional capacity evaluations in persons with spinal disorders: Predicting poor outcomes on the Functional Assessment Screening Test (FAST). *J Occup Rehabil* 2001; 11(2): 119–132.
53. Spieler EA, Barth PS, Burton JF, Jr, Himmelstein J, Rudolph L. Recommendations to guide revision of the Guides to the Evaluation of Permanent Impairment. American Medical Association. *JAMA* 2000; 283(4): 519–523.
54. Menard MR, Cooke C, Locke SR, Beach GN, Butler TB. Pattern of performance in workers with low back pain during a comprehensive motor performance evaluation. *Spine* 1994; 19(12): 1359–1366.
55. Reisine S, McQuillan J, Fifield J. Predictors of work disability in rheumatoid arthritis patients. A five-year followup. *Arthritis Rheum* 1995; 38(11): 1630–1637.
56. Loisel P, Poitras S, Lemaire J, Durand P, Southiere A, Abenhaim L. Is work status of low back pain patients best described by an automated device or by a questionnaire? *Spine* 1998; 23(14): 1588–1594.
57. Cohen JE, Goel V, Frank JW, Gibson ES. Predicting risk of back injuries, work absenteeism, and chronic disability. The shortcomings of preplacement screening. *J Occup Med* 1994; 36(10): 1093–1099.
58. Dempsey P. How to choose a strength-testing program. *Ergonomics in Design*, April 1999, pp. 18–23.
59. 42 United States Code § 19102(2)(C).

60. Hansson TH, Bigos SJ, Wortley MK, Spengler DM. The load on the lumbar spine during isometric strength testing. *Spine* 1984; 9(7): 720–724.
61. Stregre DW, Cooney WP, Wood MB, Johnson SJ, Metcalf BJ. Chronic peripheral nerve pain treated with direct electrical nerve stimulation. *J Hand Surg [Am]* 1994; 19(6): 931–939.
62. Frings-Dresen MH, Sluiter JK. Development of a job-specific FCE protocol: The work demands of hospital nurses as an example. *J Occup Rehabil* 2003; 13(4): 233–248.
63. Reneman MF, Dijkstra PU. Introduction to the special issue on functional capacity evaluations: From expert based to evidence based. *J Occup Rehabil* 2003; 13(4): 203–206.