

Sincerity of Effort Differences in Functional Capacity Evaluations

Tanya Rutherford Owen

Private Practice- Fayetteville, Arkansas

Melissa Jones Wilkins

Doctoral Student, University of Arkansas

American workers who sustain injuries in the course of employment are often referred for a functional capacity evaluation (FCE) to determine their work-related physical impairments. As part of the testing battery, the worker's sincerity of effort is often measured. The reliability and validity of the FCE is determined by this portion of the evaluation. This research examines valid versus invalid FCE outcomes of White and non-White workers and considers the primary spoken language, English or Spanish. Results of the study show significant differences in sincerity of effort determinations between White and non-White (majority Latino/a) groups and between English and Spanish speaking workers. In contrast, examination of sincerity of outcome differences between gender and age groups were non-significant. This research suggests that consideration should be given to a worker's ethnicity and primary language spoken when "unreliable" or "invalid" outcomes are obtained in a FCE.

A functional capacity evaluation (FCE) is a multi-hour or occasionally multi-day assessment of an individual's physical capabilities, most often performed by a physical or occupational therapist (Genovese & Galper, 2009; Gouttebarger, Wind, Kuijer, Sluiter, & Frings-Dresen, 2010; James & MacKenzie, 2009; L. Matheson, 2003). Functional capacity evaluations are also referred to as functional capacity assessments (FCA), physical capacity evaluations (PCE), work capacity evaluations (WCE) or functional abilities evaluations (FAE; Genovese & Galper, 2009). The American Physical Therapy Association (APTA) defines a functional capacity evaluation as a measure of "the ability of an individual to perform functional or work-related tasks and predicts the potential to sustain these tasks over a defined time frame" (APTA, 2011, p.2).

The FCE's purpose is to objectively determine the individual's functional limitations and physical capacities to work (Dresen, 2004; Gouttebarger, Wind, Paul, Kuier, & Frings, 2004; Gross, Battie, & Cassidy, 2004; R. Matheson, 2003; Reneman, Fokkens, Dijkstra, Geertzen & Froothoff, 2005).

The FCE report is utilized to compare one's demonstrated capacities following injury to the demands of one's job to determine the ability of the worker to safely return to work (Kaplan, Wurtele, & Gillis, 1996; R. Matheson, Iserhagen, & Hart, 2002). Other uses of the FCE include identifying on-the job accommodations, developing work conditioning programs, determining entitlement to disability-related benefits, and providing a framework for vocational rehabilitation services (Gouttebarger et al., 2004; Gross et al., 2004; Reneman et al., 2005).

After a worker has been diagnosed with a medical condition, the question of whether or not the worker can return to work must be answered. Historically, determining a worker's physical capacities was the task of the individual's physician (Genovese & Galper, 2009). However, as the requirement for detailed functional capacity information increased, the functional capacity evaluation process emerged (Genovese & Galper, 2009; Warren, Cupon, & Steinbaugh, 2004).

According to Genovese and Galper (2009), the first work capacity evaluation was developed in 1975 by Leonard Matheson at the Work Preparation Center at Rancho Los Amigos Hospital in California. It was developed in response to a change in the California Workers' Compensation Law which required physicians to complete a form addressing the work capacities for patients involved in workers' compensation. In

Tanya Rutherford Owen, Ph.D., CRC, CLCP, CDMS, Private Practice- Fayetteville, Arkansas, .

Email: owenvoc@gmail.com

response to these requests both in California and throughout the United States, physicians began to rely upon physical and occupational therapists to provide the requested information regarding work function capabilities. In 1983, the Polinsky Functional Capacity Assessment was the first widely available commercial FCE and, in the late 1980s, Blankenship FCEs became available. After 1990, many other commercial FCEs were in use. These functional capacity evaluations integrated the medical diagnosis provided by a physician with the measured functional abilities of the worker to perform the demands of work as outlined in the *Dictionary of Occupational Titles*, the *Selected Characteristics of Occupations as Defined in the Revised Dictionary of Occupational Titles* and *The Revised Handbook for Analyzing Jobs* (APTA, 2011).

The FCE process typically begins with the therapist obtaining the worker's informed consent, as recommended by the APTA and the American Occupational Therapy Association (AOTA). Informed consent includes discussion of what the FCE involves, risks associated with the test, and what is expected of the worker being evaluated (Genovese & Galper, 2009). Although consent can be obtained verbally, it is recommended that consent is obtained in writing and that there are procedures in place to communicate this information to illiterate and non-English speaking workers (Genovese & Galper, 2009).

Following informed consent, the evaluator interviews the worker to collect demographic data and information about activities of daily living. The evaluator also collects information about the worker's current symptoms, medications, and treatment history (Matheson, 2003). Medical records are reviewed if provided by the referral source(s) - typically physicians, insurance adjusters or attorneys. A brief physical exam is conducted to assess heart rate and blood pressure and physical testing commences (Matheson, 2003). Evaluations range in length from several hours to two-days (Genovese & Galper, 2009; James & MacKenzie, 2009). Once testing is completed, the evaluator issues a written report to the referral source(s) that contains FCE information including: 1) the physical demand level achieved by the worker, 2) answers to referral questions (e.g.: Can the worker perform past work?), and 3) the worker's sincerity of effort, which in FCE reports is referred to as reliability or validity (Chen, 2007b; Genovese & Galper, 2009; Matheson, 2003). These terms are used synonymously and do not reflect the classic definitions of reliability (i.e.: consistency) and validity (i.e.: the degree to which a test actually measures what it purports to measure) used in standardized testing (Anastasi, 1988). When inconsistencies between maximum effort and less than maximum effort are detected, the evaluator may deem the FCE "unreliable" or "invalid", with different evaluators using different terms (Innes & Straker, 1999; Matheson, 2003; Saunders, 1999). In functional capacity evaluations, an unreliable or invalid sincerity of effort assessment leads the evaluator to conclude that the FCE is not an accurate estimation of the worker's true functional capacities (Matheson, 2003; Saunders, 1999).

FCE Sincerity of Effort

As part of the FCE process, an assessment is made of the sincerity of effort put forth by the worker during testing. Sincerity of effort is defined as "a patient's conscious motivation to perform optimally during an evaluation" (Lechner, Bradbury & Bradley, 1998, p. 868). When FCEs were first developed, they did not incorporate a sincerity of effort measure (Genovese & Galper, 2009). However, due to the suspicion of disability exaggeration, also called "symptom magnification" and "malingering", sincerity of effort measures have been incorporated into most, if not all, FCE systems (Lechner et al., 1998). Malingering is defined in the *DSM-IV-TR* as the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs" (American Psychiatric Association, 2000, p.683). It has been found to occur in 25-30 percent of workers involved in personal injury, worker's compensation and disability benefits systems (Genovese & Galper, 2009). Some authors caution that the entire concept of symptom magnification is theoretically unsound, as there is no objective measure of a symptom (Lechner et al., 1998; Saunders, 1999). Lechner, Bradbury, and Bradley (1998) suggest that terms such as "symptom magnification" and "exaggerated pain behavior" should be avoided as they provide little information that leads to improved treatment and recovery.

As most clinicians and physicians have no special expertise in determining sincerity of effort, several methods have been adopted that purport to measure a worker's sincerity of effort during FCE testing. These include heart rate intensity, repeated measures assessed by coefficients of variation, and documentation of pain behavior through visual observation (Genovese & Galper, 2009; Lechner et al., 1998; Reneman et al., 2005). To better understand how these measures are utilized within a FCE to measure sincerity of effort, a brief review of these measures is included.

Based upon the heart rate theory, the more effort being exerted by the worker during the FCE, the higher his or her heart rate (Schapmire, St. James, Townsend, & Feeler, 2011). Maximum heart rate is typically calculated using the equation $\text{Max HR} = 220 - \text{Age}$ (Schapmire et al., 2011). However, heart rate as a measure of exertion has been debated in FCE literature (Morgan, Allison, & Duhon, 2012; Schapmire et al., 2011). In a review of literature on measuring maximum heart rate, the standard error of estimate within a 95% confidence interval was 40 beats per minute (Schapmire et al., 2011). These data demonstrate significant variability in this measurement but to obtain a more precise measure of heart rate, a hospital-based exercise test is required. It is also noted that heart rate is affected by medication, test anxiety and other physical conditions (Genovese & Galper, 2009; Gross & Battié, 2005c; Gross, 2006; Kaplan, Wurtele, & Gillis, 1996; St. James & Schapmire, 2011). Therefore, using maximum heart rate as a measure of effort is questionable (Morgan et al., 2012). Schapmire et al., 2011).

Coefficients of variation (CV) are another measure utilized by FCE evaluators to determine sincerity of effort in FCE participants. Statistically, the coefficient of variation is an expression of variability within a sample, some of which reflects measurement error and some of which is variability within subjects (Lechner et al., 1998). It is derived mathematically by dividing the standard deviation by the mean with the result expressed as a percentage (Kaplan et al., 1996; Lechner et al., 1998; Townsend, Schapmire, St. James, & Feeler, 2010). Historically, 15% variability is allowed in FCEs with more than 15% score variability reflecting unreliable or invalid effort (Schapmire et al., 2011). It is noted, however, that the 15% cutoff score has never been validated through a controlled study (Lechner et al., 1998; Schapmire & St. James, 2011).

The larger objection to the use of CVs in measuring sincerity of effort is that CVs are measures of reliability and not validity. For example, a worker can consistently demonstrate less than his or her maximum effort, yet their coefficient of variation score would be less than 15% and their effort would be deemed valid. This occurred in a measure of isometric strength assessments where more than half of the subjects produced CVs of 15% or less for leg and arm lifts during feigned weakness sessions (Townsend et al., 2010). Townsend et al. (2010) indicate that a "high" CV is likely to indicate feigned weakness but "low" CV may not reflect sincere effort. An additional concern in using CVs to measure effort is that CVs are influenced by the presence of pain, the instruments used, and the tasks performed (Lechner et al., 1998). In the first controlled study of static leg lift and static arm lifts (isometric assessments), Townsend, Schapmire, St. James and Feeler (2010), found that 40-80% of the subjects tested produced coefficients of variation less than 15% in tests where workers put forth less than maximum effort. They concluded that neither of these assessments is appropriate for classifying effort and should be discontinued for this purpose.

Visual observation by the FCE evaluator is another method by which sincerity of effort is assessed in FCEs. However, because consistency among ratings does not reflect accuracy of ratings, many have cautioned against the use of visual observation as a way to measure effort (Genovese & Galper, 2009; Schapmire et al., 2011; Schapmire & St. James, 2011). Schapmire et al. (2011) reported no difference in the classification of effort observed by untrained observers and trained/experienced medical professionals. Reneman et al. (2005) caution that, despite the wide use of visual observations as a sincerity of effort measure, no evidence has been published that addressed its reliability and validity for use with FCEs.

The FCE evaluator's measurement of sincerity of effort, within FCE reports, is often termed "validity" or "reliability." However, this is an inappropriate use of these terms as validity and reliability are specific scientific terms (Lechner et al., 1998). Genovese and Galper (2009) caution that tests of sincerity of effort need to be further examined and that evaluators should avoid terms such as "valid" or "invalid" in the context of describing performance effort because the meaning can be

misconstrued. In 1998, Lechner et al. cautioned that reporting that a patient has intentionally given less than full effort is a violation of the American Physical Therapy Association measurement standards. In addition to evaluation system limitations, factors such as fatigue, anxiety, pain, fear of re-injury, depression, medications, work satisfaction, lack of understanding of procedures and anxiety can impact performance consistency (Chen, 2007a; Genovese & Galper, 2009; Gross & Battie, 2005b; Gross, 2006; Kaplan et al., 1996; Lechner et al., 1998; Robinson & Dannecker, 2004). Despite this, many times these factors are not assessed during the FCE and at most are noted as anecdotes in the FCE report (Kaplan et al., 1996). Kaplan et al. (1996) examined the role of psychological factors on FCE performance and found that workers who exerted less than "maximal effort" during a FCE demonstrated more depression, anxiety, higher perceived disability ratings and were less ready to return to work than those who demonstrated "maximal effort." With this finding, the authors recommended that psychological testing be conducted with workers prior to scheduling the FCE in order to identify those who will benefit from psychological counseling to address depression, anxiety, perceived disability and/or self-efficacy barriers that may interfere with FCE performance. Their proposed process would address both psychological and physical readiness in returning to work following injury.

FCE Validity

Functional capacity evaluations have been criticized due to the lack of standardization in terminology, test length, evaluator qualifications, report format, and determination of material handling and problems with predictive validity of FCE for outcomes (Streibelt, Blume, Thren, Reneman, & Mueller-Fahrnow, 2009). While the FCE is a widely available product utilized by rehabilitation counselors in assisting workers with disabilities in returning to work, research has been mixed about the predictive validity for functional capacity evaluation results, particularly for sustaining work (Reneman, & Dijkstra, 2009).

Test validity is a test measuring what it says it is measuring and its ability to have the results used to make inferences (Gouttebarger et al., 2004; Innes & Straker, 1999; Lechner et al., 1998; Mitchell, 2008). The various types of validity (e.g., face validity, content validity, criterion-related validity, and construct validity) are all applicable to functional capacity evaluations. For a full discussion of how various types of validity can be applied to FCEs, the reader is referred to Innes & Straker (1999). Any type of validity, however, is established by research and is concerned with the results of an assessment (Innes & Straker, 1999). Innes and Straker (1999) caution that there is "no peer-reviewed scientific justification for the use of the term validity profile as that term relates to functional testing" (p. 126). Lechner et al. (1998, p.868) agree, noting "There is no evidence reported in the peer-reviewed literature that any of the tests designed to provide a 'validity' profile of the patient are valid in the scientific sense." Utilizing the term "validity" to describe a worker's sincerity of effort during testing is a misnomer as a test's validity should not and does not change based upon the test-taker's level of

effort (Innes & Straker, 1999; Lechner et al., 1998).

The type of validity that is often of interest to vocational rehabilitation counselors pertains to criterion related validity of FCEs. That is, can the FCE predict the worker's ability to successfully perform work duties? While some have reported that completion of the FCE may result in closure of a medico-legal claim (and suspension of disability benefits), a worker's performance on the FCE may not predict the worker's successful return to work (Chen, 2007a; Gross, 2006; Gross & Battie, 2005a; Kaplan et al., 1996; Matheson et al., 2002). Matheson et al. (2002) examined three dynamic lift and isometric grip force tests (typically found in FCE protocols) to determine the validity of these tests in predicting return to work in a sample of 650 adults. They determined that higher weight lifted from floor to waist by a worker was associated with a greater likelihood of return to work. Grip force was found not to relate to return to work, causing the authors to caution that the use of grip force to predict return to work should be reconsidered. In the Matheson et al. (2002) study, factors that more strongly predicted return to work (than did FCE factors) included time off work and male gender.

Gross and Battie (2005a) found similar results in a study of 130 individuals with chronic back problems. In this group, the median number of days between FCE completion and benefit suspension was 45 days. One year following the FCE, 57% of workers reported that they were working, with employed individuals reporting less pain and disability than those not working. Higher weight lifted and lower numbers of failed tasks during the FCE were weakly associated with faster benefit suspension and claim closure. They found that physical factors, perceptions of disability and pain intensity all influenced results of the FCE.

Gross et al. (2004) examined the validity of the Iserhagen Work System's evaluation in predicting return to work in a sample of adults with low back injury. They found that while only 4% of the sample met all job demands, 95% had their temporary total disability benefits suspended during the year following the FCE with the median number of days to receive temporary total disability (TTD) benefits post-FCE reported at 32. The median time to claim closure following FCE was 97 days with higher amount of weight lifted on the floor to waist lift associated with time to claim closure. Additionally, they concluded that while FCE systems assess functional limitations of an employee, they do not assess psychosocial factors which have been found to influence sincerity of effort during FCEs.

Gross and Battie (2006) reported that in a group of 336 individuals with upper extremity disorders, 95% experienced benefit suspension within one year following FCE. Here, higher lifting performance during FCE was associated with faster benefit suspension and claim closure. The median time between FCE and benefit suspension was 47 days. The injury recurrence rate in this sample was 39% and 24% of the sample resumed benefits following suspension. They concluded that a better performance during a FCE weakly predicted faster

benefit suspension but was unrelated to sustained recovery from injury.

Relevant FCE Research

While the field of rehabilitation has emphasized the need for culturally sensitive service provision, not all tools utilized by rehabilitation counselors have undergone scrutiny for cultural sensitivity or test bias. According to Reynolds and Suzuki (2012), a biased test systematically underestimates or overestimates the value of the variable it is designed to assess. A specific type of bias, called cultural bias, may be found in any type of assessment instrument. Cultural test bias has occurred if the bias is due to a specific cultural variable such as ethnicity (Reynolds & Suzuki, 2012). Unfortunately, unequal test results produced by a culturally biased test may produce inequitable social consequences (Brown, Reynolds, & Whitaker, 1999; Reynolds, 1982a, 1982b).

A review of FCE literature pertaining to cultural test bias was undertaken. Previous FCE research investigating the reliability of FCEs was identified (Gross et al., 2004; Gross & Battie, 2005a; Reneman, Schiphorts Preuper, Kleen, Geertze, & Dijkstra, 2007; Reneman et al., 2004; Smith, Cunningham, & Weinbery, 1986). Validity studies of the FCE were also identified (Gross et al., 2004; Smith et al., 1986; Reneman, Joling, Soer, & Goeken, 2001). While research supporting the influence of ethnicity on pain behavior was identified (Lechner et al., 1998), a search for previous studies of FCE outcomes based upon ethnicity or language spoken by the worker revealed no such studies.

A review of FCE literature relating to the influence of age or gender on FCE outcomes revealed a limited number of studies. Gross and Battie (2005c) investigated factors influencing results of functional capacity evaluations in worker's compensation claimants with low back pain. The researchers found that the Pain Disability Index (PDI), pain intensity, age, and sex independently contributed to floor-to-waist lift performance. However, only the PDI, pain intensity, and duration of injury contributed to the number of failed tasks. Baldasseroni et al. (2013) evaluated the correlation between depressive symptoms and 6-minute walking test (6WT) in patients with coronary artery disease (CAD) and the role of age on this relationship. They found that depressive symptoms negatively affected 6WT performance among older CAD subjects.

In the Gross and Battie (2005a) study that investigated factors influencing FCE results, language was identified as a potential determinant of FCE performance but was not investigated. Therefore, it does not appear that there has previously been an examination of FCE outcomes by ethnicity or primary language spoken. This research investigated sincerity of effort measures (i.e., "validity") in multiple FCE systems between White and non-White and English and Spanish speaking workers involved in the workers compensation system.

The following research questions were addressed:

1. Are non-English speaking and non-White workers who undergo FCE assessments more likely to have

invalid outcomes than English speaking and White workers?

The following hypotheses were developed:

Hypothesis 1: Non-English speaking workers will have higher rates of invalid sincerity of effort measures in FCE reports.

Hypothesis 2: There will be statistically significant sincerity of effort group differences pertaining to primary language spoken and ethnicity.

Hypothesis 3: There will be no significant sincerity of measure group differences pertaining to age and gender.

Methods

Participants

Demographic variables were collected including ethnicity, gender, age, highest level of education, primary language, and the sincerity of effort measure (i.e., valid or invalid effort). A review of 69 FCEs ($N=69$) was made. All participants were adults who participated in functional capacity evalua-

tions following a worker's compensation injury. Participants ranged in age from 29-66 ($M=49.70$, $SD = 10.53$) with 51% ($N=35$) between the age of 29 and 50 and 49% ($N=34$) over the age of 50. The ethnicity of the participants was 73.13% White/ Caucasian, 17.91% Latino and 8.96% African American. The highest level of education obtained by participants was less than high school (25.40%), high school (41.27%) and some college (33.33%). The most common disability type was back injury (57.14%) with other injuries reported to be knee, shoulder, pelvis and upper extremity orthopedic injuries. Some participants were monolingual English speakers, some were monolingual Spanish and some were bilingual English/ Spanish.

Procedures

An ex-post facto design was utilized in this study. A review was made of 69 functional capacity evaluations conducted with adults residing in Arkansas. Multiple FCE vendors were included in the study and the types of FCEs performed were unknown to the researchers. The cases were selected randomly from the caseload of a rehabilitation counselor and two attorneys, all of whom are regularly involved in the workers compensation system.

Statistical Analysis

In this research, differences between sincerity of effort outcomes (termed validity) were examined. These scores were examined for differences between Spanish and English speaking workers and between workers who were White and non-White. Dependent variables included valid or invalid sincerity of effort measures. Independent variables included race (White and non-White), language (English and Spanish), age (50 and under and over 50) and gender (male and female). To measure differences, a chi-square test was used as there were two nominal dependent variables, each with two measured levels. A Fisher's exact test correction was utilized where any cell contained fewer than five subjects.

To assess differences in sincerity of effort scores among demographic variables, a series of chi-square tests were conducted. A chi-square test is a measure of association between two or more variables, testing whether a single predictor variable is related to a single criterion variable (Hatcher, 2003). No relationship between

Table 1

Demographic characteristics and valid and invalid sincerity of effort scores of participants.

Variable	Level	# and (%) of Participants	Valid	Invalid	χ^2	p value	ϕ
Age (total=69)	Ages 29- 50	35 (51%)	21 (60.00%)	14 (40.00%)	$\chi^2 (1, 68) = .4365$	$p = .51$	$\phi = -.08$, a small effect size
	Over the age of 50	34 (49%)	23 (67.65%)	11 (32.35%)			
Gender (total=69)	Female	23 (33.33%)	12 (52.17%)	11 (47.83%)	$\chi^2 (1, 68) = 2.0073$	$p = .16$	$\phi = -.17$, a small effect size
	Male	46 (66.67%)	32 (69.57%)	14 (30.43%)			
Ethnicity (total=68)	White/ Caucasian	50 (73.53%)	36 (72.0%)	14 (28.0%)	$\chi^2 (1, 67) = 6.2$	$p = .01$	$\phi = -.30$, a medium effect size
	Non-White	18 (26.47%)	7 (38.9%)	11 (61.1%)			
Language (total=68)	English	57 (83.82%)	40 (70.2%)	17 (29.8%)	$\chi^2 (1, 67) = 7.3$	$p = .01$	$\phi = -.33$, a medium effect size
	Spanish	11 (16.18%)	3 (27.3%)	8 (72.7%)			

the variables results in a chi-square of zero with a stronger relationship reflected in a larger chi-square statistic (Hatcher, 2003). For chi-square tests, the usual assumptions of normal distribution and homogeneity of variance need not be met; however, independence in scores must hold (Aron, Aron, & Coups, 2005).

Phi coefficients with scores ranging from -1.00 to +1.00 are the resultant measures of chi-square tests with a 2 x 2 table (Aron, Aron, & Coups, 2005). If the obtained *p* value is less than .05, the null hypothesis (that there is no relationship between the two variables) will be rejected. A statistically significant result suggests that the two variables are probably related in the population (Hatcher, 2003).

Effect size was measured in this study by Cohen's *d* (Cohen, 1988). Effect size is defined as the magnitude (or size) of an effect (Kirk, 1995). Cohen (1988) indicates that a small effect size is indicated by a .10, a medium effect size by .25 and a larger effect size by .40. Larger effect sizes indicate stronger relationship of the measured effect. SAS 9.2 Software was utilized to compare the relationships among variables.

Results

The first analysis examined rate of FCE sincerity of effort scores by ethnic category. Ethnicity was analyzed by White and non-White categories using a chi-square test of independence. This analysis revealed a significant relationship between ethnicity and sincerity of effort scores, $\chi^2(1, 67) = 6.2$, $p = .01$. For this analysis, the phi coefficient (ϕ) was used as the index of effect size. For this analysis, ϕ equaled .30, a medium effect size. Table 1 illustrates the number of workers in White and non-White workers who obtained invalid and valid sincerity of effort scores.

The second analysis examined FCE sincerity of effort scores by language. Language was analyzed by Spanish and English categories, with bilingual workers included in the Spanish-speaking category. Data were analyzed using a chi-square test of independence with Fisher's exact test correction as one cell had fewer than five subjects. This analysis revealed a significant relationship between primary language spoken and sincerity of effort scores, $\chi^2(1, 67) = 7.3$, $p = .01$. For this analysis, ϕ equaled .33, a medium effect size. Table 1 illustrates the number of workers in Spanish and English workers who obtained invalid and valid sincerity of effort scores.

The third analysis examined FCE sincerity of effort by age. Age categories were analyzed by using those 50 and over and those below age 50. This analysis revealed a non-significant relationship between age and sincerity of effort scores, $\chi^2(1, 68) = .4365$, $p = .51$. For this analysis, ϕ equaled -.08, a small effect size. Table 1 illustrates the number of workers in 50 and above and below 50 age categories who obtained invalid and valid sincerity of effort scores.

The fourth analysis examined FCE sincerity of effort by

gender. This analysis revealed a non-significant relationship between gender and sincerity of effort scores, $\chi^2(1, 68) = 2.0073$, $p = .16$. For this analysis, a phi coefficient (ϕ) was used as the index of effect size. In this analysis, ϕ equaled -.17, a small effect size. Table 1 above illustrates the number of male and female workers who obtained invalid and valid sincerity of effort scores.

Discussion

Many rehabilitation counselors now work with clients who are Latino, as many Latino workers engage in dangerous occupations resulting in injury (Breeding, Harley, Rogers & Crystal, 2005; Moreno, 2004; United States Department of Labor, 2013). If the Latino population in the United States more than doubles as expected by the year 2060, this service provision trend will continue (U.S. Census Bureau, 2013). As Latino workers become increasingly involved in workers compensation disability systems, rehabilitation counselors working in these systems must be able to provide the required culturally sensitive rehabilitation services (Lewis & Arango-Lasprilla, 2010; Rubin, & Roessler, 2008; Smart & Smart, 1994; Wong-Hernandez & Wong, 2002). If FCEs remain a regular part of rehabilitation service planning, it critical that FCEs demonstrate acceptable measurement properties (Gross & Battie, 2005a).

Beginning with the 1978 Amendments to the Rehabilitation Act which required public rehabilitation counselors to communicate with non-English speaking clients in their native language, the field of rehabilitation has incorporated multiculturally informed practices. This practice includes requiring that all CORE approved rehabilitation programs provide multicultural training that stresses the need for using reliable and valid instruments which are appropriately normed for the population served (CORE, 2014; Rubin, & Roessler, 2008). Understanding the measurement properties of evaluation and assessment instruments is foundational to rehabilitation.

Findings of this research highlight the need for rehabilitation counselors to consider functional capacity evaluation sincerity of effort outcomes, when the FCE is conducted with Latino clients. If FCE sincerity of effort determinations are influenced by immutable characteristics such as a worker's ethnicity or primary language spoken, classifying the FCE as a purely objective assessment may be inaccurate.

When it is all said and done, a worker's maximum effort demonstrated in any setting can be nothing more than the effort that the worker is willing to produce (Reneman et al., 2005). With the multiple documented problems with assessing sincerity of effort in FCEs, persons involved in medico-legal proceedings should be cautioned against utilizing the "validity" or "reliability" scale as a reflection of anything more than a report of the worker's behavior during the test. Schapmire and St. James (2011) caution that without a sound analysis of the effort demonstrated by the worker during the FCE, the FCE report is "nothing more than a description of what the claimant did during the test" (p. 66).

While FCEs can be a useful tool for assisting workers with injuries to return to work, it is important that those who utilize FCEs do not rely solely upon the FCE in the disability determination process. As early as 1988, it was suggested that a multidisciplinary team should determine a worker's functional capacity (Chen, 2007a). With all of the instruments and methods developed since then to assess functional capacity, it appears that a multidisciplinary approach to disability determination continues to be a superior approach to relying solely on the FCE.

Limitations

Additional research is needed to examine the impact of ethnicity and language on FCE outcomes, particularly outcomes that address sincerity of effort. Examination of actual administrations across ethnic and language groups was not undertaken in this study and may provide information regarding causes of invalid assessments for non-English speaking workers. In this sample, it is unknown which assessments were conducted with an English/Spanish translator. Comparison of outcome differences in FCEs conducted with translators and with native born Spanish speaking FCE evaluators should also be undertaken. Interviewing non-English speaking clients post FCE may assist in furthering our understanding of between group differences.

Conclusions

The results of this study are important for rehabilitation counselors and those involved in the medico-legal environment where FCEs are commonly used for decision making purposes. This research demonstrates that an "invalid" sincerity of effort FCE report should be interpreted with caution, particularly for non-English speaking and non-White clients. Gross and Battie (2005a), caution that if psychosocial factors influence performance during FCEs, the data derived from FCEs would be questionable. In this research, psychosocial factors including ethnicity and language spoken were found to influence performance outcomes.

If those involved in medico-legal systems, including rehabilitation counselors, are unaware of the potential bias contained in the FCE report, certain workers with disabilities may receive unfair treatment following disability onset. By increasing the knowledge base of rehabilitation counselors about potential bias, the role of the rehabilitation counselor as advocate for the individual with a disability becomes potentially restored. This research demonstrated systematic differences in outcomes for the same test administered to White and non-White and English and non-English speaking workers. These differences may lead to unfairly disastrous consequences for the workers who lose access to retraining or to disability related payments, as those reading the FCE reports are lead to believe that the worker did not fully engage in testing and is engaging in symptom magnification or disability exaggeration.

When a worker loses credibility within a system of disability adjudication, his ability to continue to receive benefits becomes significantly compromised. For workers engaged in disability benefits systems, an invalid functional capacity outcome may result in benefit denial, premature claim closure, lost access to rehabilitation services, diminished medico-legal settlement, and lost access to medical treatment and / or limited offers of employment (Genovese & Galper, 2009; Gross, 2006; Lechner et al, 1998). With such importance given to FCEs, it is concerning that some FCE evaluators may be disproportionately reporting invalid sincerity of effort scores for non-White/ non-English speaking workers. A review of literature indicates that sincerity of effort scores were not originally included in FCE development and many have cautioned against their use in FCEs. If FCEs do not successfully predict safe, sustained return to work without recurrence of injury for all workers, regardless of language spoken and ethnicity, relying solely upon FCE outcomes to determine entitlement to disability related benefits appears to be problematic.

Armed with this research, rehabilitation counselors will be better able to recognize the potential problems with determining that non-White and non-English speaking workers provided less than full effort during a FCE. As rehabilitation counselors often educate parties within the medico-legal system (e.g. judges, claims adjusters, physicians) about FCEs, we will now be better prepared to discuss the potential bias implied when a FCE report determines that a non-White or non-English speaking worker gave less than full effort during testing. In addition to rehabilitation counselors, FCE evaluators may be better able to determine the cause of the differences found in this research with hopes of correcting the bias that has been heretofore undiscovered.

References

- American Physical Therapy Association. (2011). Occupational health physical therapy: Evaluating functional capacity guidelines. Retrieved from http://www.or-thopt.org/uploads/content_files/OHSIG_Guidelines/Occupational_Hlth_PT_Evaluating_Functional_Capacity_040610_2.pdf
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Anastasi, A. (1988). *Psychological testing* (6th edition). New York, NY: MacMillan Publishing
- Aron, A., Aron, E. N., & Coups, E. J. (2005). *Statistics for the behavioral and social sciences: A brief course* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Bartoli, N., Mossello, E., Bari, M., Marchionni, N. & Tarantini, F. (2013). Age-related impact of depressive symptoms on functional capacity measured with 6-minute walking test in coronary artery disease. *European Journal Of Preventive Cardiology*.
- Breeding, R., Harley, D. A., Rogers, J. B., & Crystal, R. M. (2005). The Kentucky migrant vocational rehabilitation program: A demonstration project for working with Hispanic farm workers. *Journal of Rehabilitation*

- tion, 71(1), 32-41.
- Brown, R.T., Reynolds, C.R., & Whitaker, J.S. (1999). Bias in mental testing since "Bias in Mental Testing." *School Psychology Quarterly*, 14, 208-238.
- Chen, J. J. (2007a). Functional capacity evaluation. *Iowa Orthopaedic Journal*, 27, 121-127.
- Chen, J. (2007b). Functional capacity evaluation & disability. *The Iowa Orthopaedic Journal*, 27, 121-127.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.
- Council on Rehabilitation Education. (2014). What is core? *Council on Rehabilitation Education*. Retrieved from <http://www.core-rehab.org/WhatIsCORE>
- Genovese, E. & Galper, J. (2009). Guide to the evaluation of functional ability. *Chicago, IL: AMA Press*
- Gouttebauge, V., Wind, H., Kuijer, P. M., Sluiter, J. K., & Frings-Dresen, M. W. (2010). How to assess physical work-ability with Functional Capacity Evaluation methods in a more specific and efficient way?. *Work*, 37(1), 111-115. doi:10.3233/WOR20101084
- Gouttebauge, V., Wind, H., Kuijer, P., & Frings-Dresen, M. (2004). Reliability and validity of functional capacity evaluation methods: a systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen work system. *International Archives Of Occupational And Environmental Health*, 77(8), 527-537.
- Gross, D., Battié, M., & Cassidy, J. (2004). The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 1: timely return to work. *Spine*, 29(8), 914-919.
- Gross, D., & Battié, M. (2005a). Functional capacity evaluation performance does not predict sustained return to work in claimants with chronic back pain. *Journal Of Occupational Rehabilitation*, 15(3), 285-294.
- Gross, D., & Battié, M. (2005b). Predicting timely recovery and recurrence following multidisciplinary rehabilitation in patients with compensated low back pain. *Spine*, 30(2), 235-240.
- Gross, D., & Battié, M. (2005c). Factors influencing results of functional capacity evaluations in worker's compensation claimants with low back pain. *Physical Therapy*, 85(4), 315-322.
- Gross, D., & Battié, M. (2006). Does functional capacity evaluation predict recovery in workers' compensation claimants with upper extremity disorders?. *Occupational & Environmental Medicine*, 63(6), 404-410.
- Gross, D. (2006). Are functional capacity evaluations affected by the patient's pain?. *Current Pain And Headache Reports*, 10(2), 107-113.
- Hatcher, L. (1994). A step-by-step approach to using the SAS® System for factor analysis and structural equation modeling. Cary, NC: SAS Institute Inc.
- Innes, E., & Straker, L. (1999). Validity of work-related assessments. *Work*, 13(2), 125.
- James, C., & MacKenzie, L. (2009). The clinical utility of functional capacity evaluations: The opinion of health professionals working within occupational rehabilitation. *Work*, 33(3), 231-239. doi:10.3233/WOR-2009-0871
- Jenson, A.R. (1980). Bias in mental testing. New York, NY: Free Press.
- Jenson, A.R. (1984). Test bias: Concepts and criticisms. In C.R. Reynolds & R.T. Brown (Eds.) *Perspective on bias in mental testing* (p507-586) New York, NY: Plenum Press.
- Kaplan, G., Wurtele, S., & Gillis, D. (1996). Maximal effort during functional capacity evaluations: an examination of psychological factors. *Archives Of Physical Medicine And Rehabilitation*, 77(2), 161-164.
- Kirk, R.E., (1995). Experimental design. Pacific Grove, CA: Brooks Cole.
- Lechner, D.E., Bradbury, S.F., & Bradley, L.A. (1998). Detecting sincerity of effort: a summary of methods and approaches. *Physical Therapy*, 78(8), 867-888.
- Lewis, A. N., & Arango-Lasprilla, J. (2010). Multicultural Challenges in Employment of People with Disabilities. *Journal Of Vocational Rehabilitation*, 33(1), 1-2. doi:10.3233/JVR-2010-0510
- Matheson, L. (2003). The functional capacity evaluation. In G. Andersson & S. Demeter & G. Smith (Eds.), *Disability Evaluation*. 2nd Edition. Chicago, IL: Mosby Yearbook.
- Matheson, L., Isernhagen, S., & Hart, D. (2002). Relationship between lifting ability and grip force and return to work. *Physical Therapy*, 82(3), 249-256.
- Matheson, R. (2003). Rehab products: industry viewpoint. High ground: FCE companies should focus on bettering the field of industrial rehab. *Advance For Directors In Rehabilitation*, 12(5), 78.
- Mitchell, T. (2008). Utilization of the functional capacity evaluation in vocational rehabilitation. *Journal Of Vocational Rehabilitation*, 28(1), 21-28.
- Morgan, M., Allison, S., & Duhon, D. (2012). Heart rate changes in functional capacity evaluations in a workers' compensation population. *Work*, 42(2), 253-257.
- Moreno, L. (2004). Understanding Worker Safety and Health Plenary Address. Department of Labor-OSHA Hispanic Safety & Health Summit; Orlando, FL
- Reneman, M.F., Joling, C.I., Soer, E.L., & Goeken, L.N. (2001). Functional capacity evaluation: Ecological validity of three static endurance tests. *Work*, 16(3), 227-234.
- Reneman, M.F., Jaegers, S.M., Westmaas M., Goeken, L.N. (2002). The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *Work*, 18(1), 23-27.
- Reneman, M., Brouwer, S., Meinema, A., Dijkstra, P., Geertzen, J., & Groothoff, J. (2004). Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in healthy adults. *Journal Of Occupational Rehabilitation*, 14(4), 295-305.
- Reneman, M., Fokkens, A., Dijkstra, P., Geertzen, J., & Groothoff, J. (2005). Testing lifting capacity: validity of determining effort level by means of observation. *Spine*, 30(2), E40-E46.
- Reneman, M.F., Schiphorts Preuper, H.R., Kleen, M., Geertzen, J.H., Dijkstra, P.U., (2007). Are pain intensity

- and pain related fear related to functional capacity evaluation performances of patients with chronic low back pain? *Journal of Occupational Rehabilitation*, 17(2), 247-58.
- Reneman, M., & Dijkstra, P. (2009). Predictive validity of FCE? Lechner DE, Page JJ, Sheffield G. Predictive validity of a functional capacity evaluation: the physical work performance evaluation. *Work*, 2008;31:21-5. Functional Capacity Evaluations. *Work*, 32(1), 105-106. doi:10.3233/WOR-2009-0835.
- Reynolds, C.R. (1982a). Construct and predictive bias. In R.A. Berk (Ed.) *Handbook of methods for detecting test bias*. (p199-227). Baltimore, MD: John Hopkins University Press.
- Reynolds, C.R. (1982b). The problem of bias in psychological assessment. In C. R. Reynolds & T.B. Gutkin (Eds.) *The handbook of school psychology* (p178-208). New York, NY: Wiley.
- Reynolds, C.R., & Suzuki, L.A. (2012). Bias in psychological assessment an empirical review and recommendations. *Handbook of Psychology*, Volume 10, Assessment Psychology, 2nd Edition. Chapter 4, 82- 113.
- Robinson, M. E., & Dannecker, E. A. (2004). Critical issues in the use of muscle testing for the determination of sincerity of effort. *Clinical Journal of Pain*, 20, 392-398
- Rubin, S., & Roessler, R. (2008). *Foundations of the vocational rehabilitation process*. Austin, TX: PRO-ED.
- Saunders, R. (1999). Sincerity of effort. *Physical Therapy*, 79(1), 94-96.
- Schapmire, D.W., St. James, J.D., Townsend, R., & Feeler, L., (2011). Accuracy of visual estimation of effort during a lifting task. *Work*, 40(4), 445-457.
- Schapmire, D. W., & St. James, J. D. (2011). Letter to the editor. *Work: Journal Of Prevention, Assessment & Rehabilitation*, 38(2), 197-199.
- St. James, J.D, & Schapmire, D. (2011). Functional capacity evaluation. Part 2: Exposing the most common myths in validity of effort testing. *IAIABC Journal*, 48 (1), 65-83.
- Smart, J. F., & Smart, D. W. (1994). Rehabilitation of Hispanics: Implications for training and educating service providers. *Rehabilitation Education*, 8(4), 360-368.
- Smith, S.L., Cunningham, S., Weinberg, R. (1986) The predictive validity of the functional capacities evaluation. *American Journal of Occupational Therapy*, 40(8), 564-7.
- Streibelt, M., Blume, C., Thren, K., Reneman, M., & Mueller-Fahrnow, W. (2009). Value of functional capacity evaluation information in a clinical setting for predicting return to work. *Archives Of Physical Medicine And Rehabilitation*, 90(3), 429-434. doi:10.1016/j.apmr.2008.08.218
- Townsend, R., Schapmire, D.W., St.James, J., & Feeler, L. (2010). Isometric strength assessment, part II: Static testing does not accurately classify validity of effort. *Work*, 37, 387-394.
- United States Department of Labor. (2013). The Latino Labor Force in the Recovery. Retrieved from <http://www.dol.gov/sec/media/reports/hispaniclaborforce/>
- U.S. Census Bureau. (2012). U.S. Census Bureau projections show a slower growing, older, more diverse nation a half century from now. Retrieved from <https://www.census.gov/newsroom/releases/archives/population/cb12-243.html>
- Warren, T.J., Cupon, L.N., & Steinbaugh, J.H. (2004). Functional and work capacity evaluation issues. *Journal of Chiropractic Medicine*, 3, 1-5. Retrieved from http://www.forensic-sciences.org/member/forensicpearls/fce_issues.pdf
- Wong-Hernandez, L., & Wong, D.W. (2002) The effects of language and culture variables to the rehabilitation of bilingual and bicultural consumers: A review of literature study focusing on Hispanic Americans and Asian Americans. *Disability Studies Quarterly*, 22(2), 101-19

Copyright of Journal of Rehabilitation is the property of National Rehabilitation Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.