PERRY N. BRUBAKER, PT, MS[1] • **FRANK J. FEARON,** PT, DHSc, OCS, FAAOMPT[2]
STEPHEN M. SMITH, PhD[3] • **RICHARD J. MCKIBBEN,** PT, MS, ECS[4] • **JAMES ALDAY,** MD[5]
STACIE S. ANDREWS, PT, MS[6] • **EVERALD CLARKE,** PT, MS[7] • **GEORGE L. SHAW JR,** PT, MS[8]

Sensitivity and Specificity of the Blankenship FCE System's Indicators of Submaximal Effort

STUDY DESIGN: Single-blinded, randomized, posttest only design.

OBJECTIVE: To help contribute to the body of evidence in defining the validity of functional capacity evaluations.

BACKGROUND: Functional capacity evaluations (FCEs) are tests used to help determine an individual's readiness to return to work. Most FCEs incorporate indicators of effort within the evaluation. Published evidence validating the use of these indicators is limited.

METHODS AND MEASURES: Forty-nine injured and noninjured individuals 18 to 65 years of age participated in this study. The participants were randomly assigned to 1 of 2 groups: 100% effort or 50% effort. Raters were blinded to participant group. The Blankenship Version 6.0 software was used to analyze the data and a Blankenship FCE validity profile was scored. A score of 70% or greater was deemed a valid FCE as adopted by the Blankenship protocol.

RESULTS: The sensitivity of the FCE components tested was demonstrated to be 80% and specificity was 84.2%. The positive likelihood ratio was 5 and the negative likelihood ratio was 0.2. A receiver operating characteristic (ROC) curve demonstrated the 70% cut-off value for scoring the FCE was optimal.

CONCLUSION: Four components of the Blankenship FCE system demonstrated good sensitivity and specificity for detecting submaximal effort. However, clinicians should note that false positives (maximum effort identified as submaximal effort) may occur and scores of "equivocal" are not scored in the "criteria passed" category. The rater should be aware that this method of scoring could potentially influence a client's overall FCE score.
J Orthop Sports Phys Ther 2007;37(4):161-168. doi:10.2519/jospt.2007.2261

KEY WORDS: ergonomics, false positives, functional capacity evaluation, sincerity of effort, work-related injuries

In 2004 the Bureau of Labor Statistics reported over 4.8 million nonfatal work-related injuries and illnesses, of which a substantial number were musculoskeletal in origin.[3] Such work-related musculoskeletal disorders result in marked expenses and decreased work productivity within American industries. The National Institute for Occupational Safety and Health (NIOSH) previously reported a $13-billion total annual cost for combined workplace injuries and illnesses.[13] As a result of these increasing costs, industries strive to find timely and objective means of assessing a worker's ability to perform job tasks before and after an injury.[12] Functional capacity evaluations (FCE) may be implemented to help reduce these costs and safely return injured workers to productivity.[9]

After an injury, a person may not want to return to work for both financial and emotional reasons. Consequently, the FCE typically incorporates means of determining a worker's sincerity of effort while performing functional tasks. There are many reasons why a person may not give full effort, including pain, fear of pain, lack of understanding of instructions, lack of understanding of test importance, secondary financial gain, and secondary emotional gains.[10]

There are 10 well-known marketed FCE systems. Each of these systems varies in terms of the type of equipment used and differs in standardization of instructions. However, they all have the primary goal of assessing a person's work-related abilities. King et al[9] reviewed the research related to the 10 major FCE systems. According to the review, the Blankenship, WorkHab, AssessAbility, and Key sys-

[1]Master's student (at time of study), North Georgia College and State University, Dahlonega, GA; Physical Therapist, Meridian, MS. [2]Professor, Department of Physical Therapy, North Georgia College and State University, Dahlonega, GA. [3]Professor, Department of Psychology, North Georgia College and State University, Dahlonega, GA. [4]Physical Therapist, EMG of Georgia, Inc, Pine Mountain, GA. [5]Medical Director (at time of study), Northeast Georgia Medical Center's Industrial Rehabilitation Program, Gainesville, GA. [6]Physical Therapist, Stevens County Hospital, Cornelia, GA. [7]Physical Therapist, Walton Rehabilitation Hospital, Augusta, GA. [8]Physical Therapist, Piedmont Hospital, Atlanta, GA. This study was completed in partial fulfillment of requirements for the degree of Master of Science in Physical Therapy at North Georgia College and State University, Dahlonega, GA. The protocol for this study was approved by the Institutional Review Board of North Georgia College and State University. Address correspondence to Perry N. Brubaker, 2505 56th Street, Meridian, MS 39305. E-mail: perrynarro@yahoo.com

tems all lack peer-reviewed published research.

Little is known about FCEs and their ability to distinguish maximum effort from submaximal effort among workers. Furthermore, only 1 study to date has investigated validity indexes of indicators of effort within an FCE system.[8] Jay et al[8] reported excellent positive and negative predictive values (94.44% and 80.0%, respectively) of the Employment Potential Improvement Corp (EPIC) system of validating the effort given by tested subjects.

Documentation of limited effort during a FCE may result in negative legal and financial consequences for the individual tested. Therefore, it is imperative that statements regarding the validity of patient effort demonstrated during objective functional tests such as FCEs be based on validated procedures. Specifically, the screening tool should give clinicians the ability to recognize a target population. Because of the medico-legal nature of FCEs and the possible consequences facing the worker if documentation of limited effort is found, validity indexes of such diagnostic tools must be established. Clinicians should have available for use a tool that most accurately defines the presence of a condition such as submaximal effort. Therefore, the purpose of this study was to investigate the sensitivity and specificity of the indicators of validity within 4 components of the Blankenship FCE system. The following research question was addressed: what is the sensitivity and specificity of 4 components of the Blankenship FCE system in determining submaximal effort? A similar design and method to that utilized by Jay et al[8] in their investigation of the EPIC system's indicators of effort was adopted.

The Blankenship FCE contains 3 evaluative components and 7 functional-testing components. The evaluative components are performed prior to functional testing and include a symptom exaggeration profile, nonorganic profile (evaluation for illness behavior or symptom magnification syndrome), and a pre-FCE musculoskeletal evaluation. The 7 functional-testing components include repetitive-movement tests, static-strength tests, occasional-material-handling tests, hand tests, frequent-material-handling tests, nonmaterial-handling tests, and constant tests (evaluates nonmaterial and material handling over time). Each functional testing component has its respective indicators of validity. These indicators of validity, such as coefficient of variation, analysis of the bell curve with grip testing, and extrapolations of static strength to dynamic strength, to name a few, are established by software internal to the equipment. The term *indicators of validity* is specific to the Blankenship FCE. The authors have chosen to use this term to provide consistency with the FCE language and protocol. The term in no way indicates validity of the FCE and is simply adopted by the Blankenship Group to define the data analysis within the components to determine effort.

This study focused on 4 of the functional components with their respective 19 indicators of validity. The functional components investigated were repetitive-movement tests, static-strength tests, occasional-material-handling tests, and hand tests. Repetitive-movement tests consisted of 3 movement patterns to determine the participant's willingness to move. These movements consisted of bending, reaching, and squatting. Static strength tests required the participant to stand on a platform and perform static lifts in 6 different postures: arm lift, torso lift, leg lift, high-far lift, floor lift, and high-near lift. Occasional-material-handling tests assessed how much the participant could lift at an occasional frequency (0% to 33% of the workday). The postures tested were torso lift, leg lift, 12-inch (30-cm) leg lift, shoulder lift, overhead lift, and carrying 30 feet (9.1 m). Hand tests required participants to perform a series of grip and pinch tests, including maximal static grip in the number 2 position, maximal static grip in all 5 positions, rapid-exchange grip, and a maximum key, tip, and palmar pinch.

These components were chosen because they are included in most FCEs and their indicators of validity have been the focus of much research.[1,10,14] The 19 indicators of validity incorporated into these components are based on the research behind grip strength force curves, coefficient of variation, static strength extrapolations to dynamic strength, frequency of lift with regard to amount of weight lifted, and the rater's subjective report of the client's behavior throughout the test in determining submaximal effort.

METHODS

The study used a single-blinded, randomized, posttest-only design. Sensitivity was defined as the system's ability to detect submaximal effort when submaximal effort truly existed. Specificity was defined as the system's ability to detect the absence of submaximal effort when submaximal effort was truly absent.[15]

The participants in the study were recruited from healthcare and academic institutions from 3 Georgia cities. Sixty participants with or without musculoskeletal pain or injury volunteered for the study. The participants were randomly assigned to 1 of 2 groups: 100% effort or 50% effort. The 100%-effort group was instructed to give 100% of their effort throughout testing while the 50%-effort group was instructed to give only 50% of their maximal effort during testing. Informed consent was obtained prior to participation in the study. The protocol for this study was approved by the Institutional Review Board of North Georgia College and State University.

Inclusion criteria included both males and females, aged 18 to 65 years, with or without previously reported musculoskeletal injury or pain. The participants' height was between 1.5 and 1.95 m to allow each participant enough vertical distance when lifting overhead. All participants were fluent in English.

Participants were excluded from the study if they were currently receiving medical treatment including a previous FCE, had pending workers compensation claims or litigation, had a resting heart rate below 50 beats per minute or above 120 beats per minute, had a systolic blood pressure greater than 200 mmHg and/or a diastolic blood pressure greater than 110 mmHg[4], were taking sedatives or illegal substances, had visual, auditory or balance impairments, had a history of cerebrovascular accident or heart disease, or were unable to grasp the handles of the lifting box. Exclusion criteria were adopted from the Blankenship procedural manual and further defined when necessary.[2,4,8,11,17]

The raters (a physical therapist and an occupational therapist) were certified in Blankenship FCE testing by the Blankenship Group and blinded to participant group assignment. Data for interrater reliability analysis were collected by having the raters simultaneously observe subject performance on the selected FCE components. Each rater applied an ordinal scale with 3 possible ratings (valid, invalid, or equivocal) to each of the 4 observed components. The ratings were recorded by each rater in response to 3 questions applied to each of the 4 components: Do movement patterns match the pain? Do movement patterns improve with distraction? Is overreaction behavior present? Although each FCE component consisted of multiple, individual, functional tests, the ratings were applied to observed subject performance during each component as a whole, considering performance across all tests comprising the respective components. The paired observations and ratings were performed with 4 subjects.

A trial run to enhance the internal validity of the study (as adopted by Jay et al[8]) was conducted prior to data collection to quantify the actual effort demonstrated among participants when they were verbally instructed to give 50% of their maximum. Ten volunteers participated in the trial run, where trials of a static-grip test and the leg lift of the occasional material-handling test were performed. Participants were first instructed to give 50% effort so they would be unaware of maximum performance. They were then allowed to repeat the testing, giving 100% effort. The actual percentage of maximal effort ([50% effort score/100% effort score] × 100) demonstrated when asked to give submaximal effort was as follows: right grip, 58.8%; left grip, 55.3%; leg lift, 57.8%.

The order of testing for the study followed Blankenship FCE protocol: repetitive-movement tests, static-strength tests, hand tests, and occasional material-handling test, including the leg lift, 12-inch (30-cm) leg lift, shoulder lift, overhead lift, and carrying 30 feet (9.1 m). The Blankenship interface system and Blankenship Version 6.0 software (HOGGAN Health Industries, Inc, West Jordan, UT) were used to analyze the indicators of validity.

Sixty participants volunteered for this study and 11 were excluded. Eight were excluded based on the exclusion criteria, 2 due to disclosure of group assignment, and 1 chose to withdraw from the study. Therefore, the results of this study are based on 49 participants. The average age of the participants was 36 years old (age range, 18-65 years). There were 17 (34.7%) males and 32 (65.3%) females. The participants included 17 (34.7%) African-Americans, 31 (63.3%) Caucasians, and 1 multiracial participant (**TABLE 1**). Thirty-one participants reported having

TABLE 1 Demographic Data of Participants

Subject Characteristics	100% Effort	50% Effort	Significance Test
Age (y)			$t = 1.3$ ($P = .16$)
Mean	31.7	37.5	
Range	20-60	18-65	
Gender			$\chi^2 = 0.1$ ($P = .8$)
Male	7	10	
Female	12	20	
Race			$\chi^2 = 0.7$ ($P = .3$)
Caucasian	12	19	
African American	7	10	
Other	0	1	
Hand dominance			$\chi^2 = 0.7$ ($P = .3$)
Right	18	28	
Left	1	1	
Both	...	1	
Height (m)			$t = 1.4$ ($P = .2$)
Mean	1.6	1.7	
Range	1.5-2.0	1.5-1.8	
Body mass (kg)			$t = 0.7$ ($P = .5$)
Mean	76.0	72.7	
Range	54.0-127.8	54.0-103.5	
Injury/pain			$\chi^2 = 1.3$ ($P = .2$)
Yes	11	18	
No	8	12	
Employment			$\chi^2 = 0.1$ ($P = .8$)
Employed	11	18	
Not employed	7	11	
Retired	1	1	

TABLE 2 Sensitivity and Specificity for Various Functional Capacity Evaluation Cutoff Scores

Cutoff Score	Sensitivity	Specificity
55%	36.7	100.0
60%	53.0	89.5
65%	60.0	89.5
70%	80.0	84.2
75%	86.7	68.4
80%	100.0	42.0

presence or a history of musculoskeletal pain or injury and 18 reported no musculoskeletal pain or injury.

SPSS statistical software (SPSS, Inc, Chicago, IL) was used to perform the data analysis. Interrater reliability was assessed by computation of percent agreement: dividing the number of exact agreements by the number of possible exact agreements and multiplying by 100.[15] Independent *t* tests and chi-square analyses were used to analyze differences between groups. Sensitivity, specificity, and likelihood ratios were computed from contingency tables. Receiver operating characteristic (ROC) curve analysis was used to help determine the optimal FCE cutoff score.

RESULTS

Nineteen participants were randomly assigned to the 100%-effort group and 30 were assigned to the 50%-effort group. There were no significant differences between groups for age, height, body mass, gender, race, hand dominance, injury/pain, or employment (**TABLE 1**).

Interrater reliability calculated by using percent agreement between the 2 raters was demonstrated to be 81.6%.

The sensitivity and specificity of 4 components of the Blankenship FCE, utilizing the Blankenship standard method of rating validity according to its benchmark of meeting at least 70% of the validity criteria, were demonstrated to be 80.0% and 84.2%, respectively. The average (±SD) Blankenship FCE validity score was 77.3% (±10.4%) for the 100%-effort group and 58.3% (±12.7%) for the 50%-effort group.

The positive likelihood ratio for the Blankenship FCE was demonstrated to be 5 and the negative likelihood ratio was 0.2. A receiver operating characteristic (ROC) curve demonstrated that the optimal FCE cutoff score was 70% (**FIGURE, TABLE 2**).

DISCUSSION

A tool incorporating assessments of effort, such as an FCE, should strive to be both highly sensitive and specific in order to correctly identify all clients' performance. Poor case outcome could result if the tool were not able to distinguish between maximal and submaximal effort and were to incorrectly identify a maximal effort as a submaximal effort (a false positive). The importance of validating indicators of effort within FCEs cannot be overstated. In many cases FCEs are being utilized to make decisions regarding safe return to work, based on demonstrated physical abilities, and their results may be used in legal settlements. When an individual is stated to have performed in a submaximal manner the results may have serious consequences. Therefore, it is critical that such statements regarding effort be based on sound research that establishes the validity of the testing. It is also incumbent upon the tester to recognize the potential negative consequences that could result from improperly designating effort as submaximal when it is in fact maximal effort. Only tests with demonstrated appropriate validity indexes, such as high sensitivity and specificity, should be utilized in making such evaluative statements.

However, sensitivity and specificity have limitations in their interpretation. Sensitivity alone cannot describe how often patients testing positive actually have the condition of interest.[16] Likewise, specificity cannot describe how often patients testing negative are absent of the condition of interest. Oftentimes, predictive values are used in addition to sensitivity and specificity to better illustrate the proportion of patients testing positive or negative that actually have or do not have the condition of interest. The closer the predictive value is to 100% the more likely the condition is present or not, depending on whether the test is positive or negative.[16] In the EPIC study, positive and negative predictive values (94.44% and 80.0%, respectively) were discussed in addition to sensitivity and specificity to provide more meaning to the clinician administering the test. Predictive values are most useful when the diagnostic test has only 2 outcomes.[16] In the case of the EPIC, these were "sincere" or "insincere" effort. In the current study, the Blankenship Group had developed a cutoff value of 70% or greater to define a valid FCE. Because the cutoff value had not been

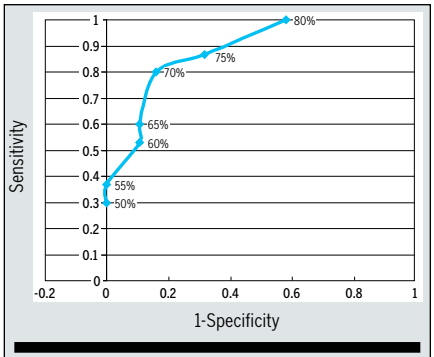

FIGURE. Receiver operating characteristics. Plot of sensitivity (true positives) and 1-specificity (false positives) relationship using various functional capacity evaluation cut-off values. The 70% cutoff value provides the greatest diagnostic accuracy by balancing the objectives of maximizing true positives while minimizing false positives.

previously studied, likelihood ratios instead of predictive values were chosen to give more meaning to sensitivity and specificity.

The likelihood ratio helps the clinician determine the probability of the condition of interest in the population. When used with pretest probability, it allows the clinician to determine how likely the condition of interest is present after the test results have been obtained (posttest probability).[5,16] Jaeschke et al[7] have given meaning to interpreting likelihood ratios and their values; they have shown that positive likelihood ratios between 5 and 10 or negative likelihood ratios between 0.2 and 0.1 generate moderate shifts from pretest to posttest probability.[7] In the current study, 4 components of the Blankenship FCE, using a FCE cutoff score of 70%, demonstrated an aggregate sensitivity of 80% and a specificity of 84.2%, producing a positive likelihood ratio of 5 and a negative likelihood ratio of 0.2. According to Jaeschke et al,[7] these values may indicate a moderate shift from pretest to posttest probability.

The 70% cutoff value was originally determined to be the cutoff point for passing the validity criteria by the Blankenship Group. This cutoff was based on an intuitive analysis of a clinical database obtained by the Blankenship Group during its development. Therefore, our study analyzed the sensitivity and specificity using the 70% cutoff criteria, as has been the practice integrated into the Blankenship FCE system. Additionally, we looked at the sensitivity and specificity at multiple cutoff points. As demonstrated by the ROC curve (**FIGURE**), the 70% cutoff value does appear to provide the maximum true positives while minimizing the false positives and, therefore, provides the criteria for the greatest diagnostic accuracy. It is critical in FCE testing to maximize true positives while minimizing false positives to prevent incorrect identification. Because documentation of limited effort could result in negative legal and financial consequences for the worker, the clinician should strive to avoid incorrect identification. Most critically, the clinician should avoid documenting a person as giving submaximal effort when it is in fact a true maximum effort (false positive).

The primary purpose of this study was to determine the sensitivity and specificity of the validity criteria of 4 components of Blankenship FCE. In clinical practice it is also useful to know which, if any, indicator of effort is the most predictive of submaximal effort. The authors attempted to identify which of the Blankenship indicators of validity were most predictive in determining submaximal effort using cross-tabulations.[19]

Within the 4 components tested in this study, 19 different indicators of validity were used to determine the participant's effort. Each component within the Blankenship system has its respective indicators of validity, and these indicators can be tested multiple times throughout a FCE. Cross-tabulations were only performed on those variables tested with 38 or more participants so that cross-tabulations could have sufficient statistical power. Only 5 of the indicators of validity tested scored greater than 70% sensitivity (**TABLE 3**). Likewise, 12 indicators had 100% specificity (**TABLE 4**). However, these variables had low sensitivity (less than 70%). Only 1 indicator had both sensitivity and specificity greater than 70%. This indicator of validity was "OMH is greater than the high extrapolation from the leg static-strength test." The sensitivity was 78.6% and the specificity was 72.2% (**TABLE 3**).

For this study, pain was defined as any current or past musculoskeletal injury or symptom experienced by the participant. Some studies have suggested that pain responses and submaximal effort can be differentiated, while others have suggested that they cannot be as easily differentiated.[6,18,20] Because identification of effort is such a controversial issue, participants with and without musculoskeletal pain or injury were included in this study to challenge both the raters' and the system's ability to correctly identify a

TABLE 3 Variables With 70% Sensitivity or Greater and Participants' Scores on These Variables

Variable/Score	100% Effort (n Participants)	50% Effort (n Participants)	Sensitivity	Specificity
Fatigue Percent Index of the high-far SST			70.0	57.9
Invalid	5	13		
Equivocal	3	7		
Valid	11	9		
OMH greater than high extrapolation of the leg test			78.6	72.2
Invalid	5	22		
Equivocal	0	0		
Valid	13	6		
REG on right			83.3	68.4
Invalid	6	24		
Equivocal	0	1		
Valid	13	5		
REG on left			83.3	57.9
Invalid	8	24		
Equivocal	0	1		
Valid	11	5		
30-cm lift greater than the leg lift			72.4	47.4
Invalid	10	21		
Equivocal	0	0		
Valid	9	8		

Abbreviations: OMH, occasional-material-handling test; REG, rapid-exchange grip test; SST, static-strength test.

participant's effort. There were 31 participants who reported having 1 or multiple sites of musculoskeletal injury or pain. Of these, 20 were in the 50%-effort group and 11 in the 100%-effort group. A chi-square analysis revealed no significant association of injury diagnosis and group assignment (**TABLE 1**).

During the FCE, the indicators of validity are scored as valid, equivocal, or invalid. A valid score according to Blankenship protocol means the participant passed the criteria of the indicator of validity. An equivocal score represents uncertainty regarding the level of effort the participant gave, while an invalid score means that the participant failed the criteria. The authors found that a change in the method of scoring equivocal ratings may further reduce the false positive rate. The overall Blankenship FCE score is calculated by dividing the variables passed (numerator), known as the *criteria passed*, by the total variables scored (denominator), known as the *criteria scored*. The Blankenship 6.0 software gives an individual a score of equivocal when the participant scores between the coefficient of variation ranges or during any part of testing where the rater or the system cannot definitively determine effort. According to the Blankenship protocol, the number of equivocal scores is added to the criteria-scored category (denominator) but not to the criteria-passed category (numerator).

This method of scoring results in the equivocal scores counting against the participant because the denominator increases while the numerator only includes the number in the criteria-passed category. This role of equivocals in scoring could potentially negatively influences a score. For example, a participant in the 100%-effort group who had a FCE score of 57% passed 27 of the 47 variables scored. Of the 47 variables scored, 7 were equivocals. If the score were calculated by adding the 7 equivocals to the 27 criteria passed, the FCE score would increase to 72%. Adding the equivocals to the criteria-passed category in the data obtained

TABLE 4 Variables With 100% Specifity and Participants' Scores on These Variables

Variable/Score	100% Effort (n Participants)	50% Effort (n Participants)	Sensitivity	Specificity
Overreaction for static			4.0	100
Invalid	0	1		
Equivocal	0	0		
Valid	14	24		
Do movement patterns match pain for static?			28.0	100
Invalid	0	2		
Equivocal	0	5		
Valid	14	18		
Do movement patterns improve with distraction for static?			25.0	100
Invalid	0	3		
Equivocal	0	3		
Valid	14	18		
OMH greater than high extrapolation for shoulder			46.7	100
Invalid	0	14		
Equivocal	0	0		
Valid	19	16		
OMH greater than high extrapolation for overhead			65.5	100
Invalid	0	19		
Equivocal	0	0		
Valid	19	10		
REG consistent right			40.0	100
Invalid	0	10		
Equivocal	0	2		
Valid	19	18		
REG consistent left			50.0	100
Invalid	0	11		
Equivocal	0	4		
Valid	19	15		
Right key pinch			30.0	100
Invalid	0	7		
Equivocal	0	2		
Valid	19	21		
Movement pattern matches pain for HT			18.5	100
Invalid	0	2		
Equivocal	0	3		
Valid	16	22		
Movement patterns improve with distraction for HT			22.2	100
Invalid	0	3		
Equivocal	0	3		
Valid	16	21		
Overreaction (OMH)			10.3	100
Invalid	0	1		
Equivocal	0	2		
Valid	19	26		
Distraction (OMH)			37.9	100
Invalid	0	7		
Equivocal	0	4		
Valid	19	18		

Abbreviations: HT, hand test; OMH, occasional-material-handling test; REG, rapid-exchange grip test; SST, static-strength test.

from the study would increase the specificity to 94%, preventing a false positive by raising the FCE score above the 70% cutoff value, thereby reducing the rate of false positives to 4%. However, when the data were reanalyzed with all the equivocals added to their respective criteria passed categories, there was an overall decrease in the system's sensitivity. The recalculated sensitivity was 56%. Therefore, this change in calculating scores is not recommended. Additionally, when the equivocal scores were taken out of the calculations altogether (that is, only criteria passed was divided by true criteria scored), the overall sensitivity and specificity were 63.3% and 89.5%, respectively. This calculation increased the false negatives to 11 and decreased the false positives to 2. Based on the above examples, the strongest sensitivity and specificity for the 4 components tested was shown by allowing the system to score equivocals as designed. However, further exploration of the scoring of equivocal ratings is recommended.

The primary limitation of this study is that the subjects participating may not be identical to a typical population of individuals who would participate in a FCE. Therefore, the external validity may be at risk. However, it would not be possible to conduct such studies with subjects who may indeed have emotional or financial incentives to perform less than optimally. The inclusion criterion to select subjects with histories of musculoskeletal pathologies was designed with this in mind and to obtain a sample as close as possible to the typical individual tested. Interrater reliability was assessed using percent agreement and could be a limitation of the study. A stronger statistic, such as kappa, was not feasible in calculating interrater reliability due to small sample size. Lastly, the amount of weight to be lifted could have been known to the participant during the occasional material-handling test, as the Blankenship materials included known weights.

Recommendations for future research of the Blankenship FCE system include conducting a study where discriminant analysis can be used to establish the variables that are most correlated with effort. Also, a study to examine the role of the equivocals and their influence on sensitivity and specificity is warranted, based on the role we found the equivocals to play in scoring. Both this study and the related study by Jay et al[8] demonstrated good sensitivity and specificity of different systems' methods of assessing effort. The EPIC system primarily utilizes a simple dichotomous method of rater observation of maximum effort by criteria, whereas the Blankenship system uses a complex computerized criteria checklist that confirms effort consistency in different tasks as well as incorporates the trained rater's opinion based upon observation. The systems basically represent high-tech versus low-tech methodologies, with the common feature being trained rater observation. The current study demonstrated good sensitivity and specificity of 4 of 7 functional testing components of the Blankenship FCE system. During FCE testing, clinicians typically choose the appropriate testing components based on the worker's diagnosis, job demands, and current level of function. The authors feel this study may support future research in determining if the system could be condensed in its ability to determine effort. Lastly, further research validating other FCE systems may help move the medical professions using FCEs toward consensus regarding the optimal means of determining effort.

CONCLUSION

Four components of the Blankenship FCE demonstrated a sensitivity of 80% and a specificity of 84.2% in determining submaximal effort. The 70% cutoff score developed by the Blankenship Group was shown to provide the greatest diagnostic accuracy for determining effort. Five indicators of validity were shown to have 70% sensitivity or greater and 12 indicators had 100% specificity. The clinical relevance for this study is that the validity indicators of 4 components of the Blankenship FCE had good sensitivity and specificity; however, raters should recognize that a small percentage of false positives (maximum effort identified as submaximal effort) might occur. Also, the clinician should note that scores of equivocal are not scored in the criteria-passed category and could potentially negatively affect a worker's overall FCE validity score. ◉

REFERENCES

1. Baker JC. Burden of proof in detection of submaximal effort. *Work.* 1998;10:68-70.
2. Blankenship K. *The Blankenship System Functional Capacity Evaluation: The Procedural Manual.* Macon, GA: The Blankenship Corporation; 1994.
3. Bureau of Labor Statistics. Incidence Rates of Nonfatal Occupational Injuries and Illnesses by Industry and Case Types, 2004. Available at: http://data.bls.gov. Accessed August 16, 2006.
4. Fletcher BJ, Dunbar S, Coleman J, Jann B, Fletcher GF. Cardiac precautions for non-acute inpatient settings. *Am J Phys Med Rehabil.* 1993;72:140-143.
5. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther.* 2001;81:1546-1564.
6. Hildreth DH, Breidenbach WC, Lister GD, Hodges AD. Detection of submaximal effort by use of the rapid exchange grip. *J Hand Surg [Am].* 1989;14:742-745.
7. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA.* 1994;271:703-707.
8. Jay MA, Lamb JM, Watson RL, et al. Sensitivity and specificity of the indicators of sincere effort of the EPIC lift capacity test on a previously injured population. *Spine.* 2000;25:1405-1412.
9. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther.* 1998;78:852-866.
10. Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: a summary of methods and approaches. *Phys Ther.* 1998;78:867-888.
11. Lechner DE, Roth D, Straaton K. Functional capacity evaluation in work disability. *Work.* 1991;1:37-47.
12. Menard MR, Hoens AM. Objective evaluation of functional capacity: medical, occupational, and legal settings. *J Orthop Sports Phys Ther.* 1994;19:249-260.
13. National Institute for Occupational Safety and Health. Musculoskeletal disorders (MSDs) and

workplace factors: a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back. DHHS (NIOSH) Publication No 97-141. Cincinnati, OH: US Department of Health and Human Services; 1997.

14. Niebuhr BR, Marion R. Detecting sincerity of effort when measuring grip strength. *Am J Phys Med.* 1987;66:16-24.

15. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* East Norwalk, CT: Appleton & Lange; 1993.

16. Riddle DL, Stratford PW. Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. *Phys Ther.* 1999;79:939-948.

17. Sanderson PL, Todd BD, Holt GR, Getty CJ. Compensation, work status, and disability in low back pain patients. *Spine.* 1995;20:554-556.

18. Shechtman O, Gutierrez Z, Kokendofer E. Analysis of the statistical methods used to detect submaximal effort with the five-rung grip strength test. *J Hand Ther.* 2005;18:10-18.

19. Stevens J. *Applied Multivariate Statistics for the Social Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.

20. Stokes HM. The seriously uninjured hand--weakness of grip. *J Occup Med.* 1983;25:683-684.