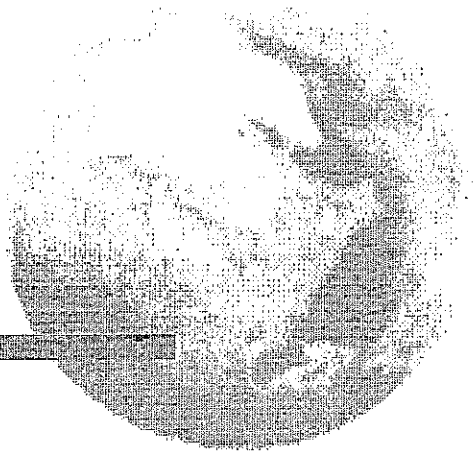
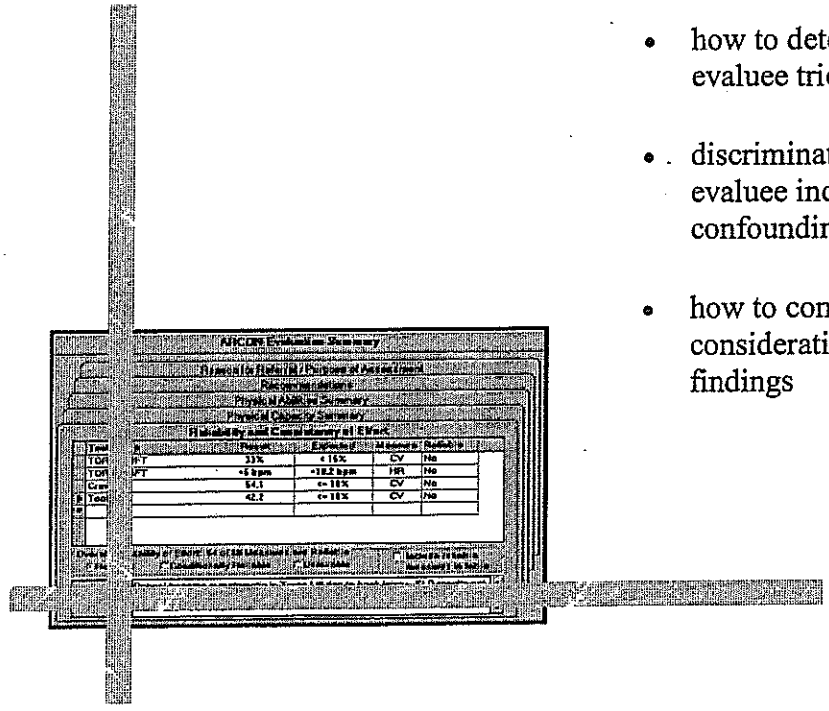
evaluee reliability



I. Introduction to Evaluee Reliability
II. Evaluation Protocols
III. Behavioral Indicators

WHAT YOU NEED TO KNOW

- how to determine the evaluee tried his/her best
- discriminate between evaluee inconsistency and confounding variables
- how to complete an FCE in consideration of positive findings



I. INTRODUCTION TO EVALUEE RELIABILITY

HOW DO YOU KNOW THE EVALUEE TRIED HIS/HER BEST?

This section will bring together the data from the preceding sections to develop an interpretation and opinion as to whether the evaluee tried his/her best. There is no algorithm or comprehensive guide that can fully assist in the final interpretation and conclusions. The medical-legal issues have to be considered very carefully because empirical evidence does not support making conclusions solely on one segment of data.

There is equivocal research on whether the existence of disability compensation has a confounding effect on disability duration and outcomes. It is estimated only 5% of disability claimants consciously alter their behavior to manipulate outcomes in their favor. Ultimately, the evaluator has to rely on clinical judgment as well as the objective evidence to make an opinion as to whether the evaluee tried his/her best, and what that may mean to the results of the testing.

DEFINITIONS:

Evaluee consistency is evidence of expected results in unconfounded testing, compared to criteria developed from research studies.

Non organic signs are evidence of biomechanically unexpected results in dissimulated physical examination.

Malingering is deliberate behavior representing greater impairment to influence compensation decisions.

Repeated Measures is successive performance on the same task.

Alternate Forms is a change in the content of a test while maintaining the criterion construct.

Dissimulation is deliberate misleading or leaving expectations ambiguous so as to prompt an evaluee to portray inappropriate behavior if they are so inclined.

Naturalistic testing is observation and measurement in an environment absent of the artificiality of the testing situation.

Standard Error of Measurement is an estimate of the absolute reliability of a test; the amount of error to expect in an evaluee's score represented by the Coefficient of Variance.

CONSISTENCY

Various protocols have been developed to test for client consistency. The most researched FCE protocols involve comparing the consistency of the evaluee's scores on *Repeated Measures* to the *Standard Error of Measurement* (SEM) as reflected in the *Coefficient of Variance* (CoVar). It is hypothesized that an evaluee's scores will be within the established CoVar if they are giving maximal effort, and that they will have higher than expected CoVar if they are giving sub-maximal effort. This hypothesis has shown equivocal results in empirical research

evaluee reliability

COEFFICIENT OF VARIANCE

Higher than expected coefficient of variance (CoVar) is not necessarily evidence of submaximal effort. Various confounding variables can contribute error measurement. Fatigue and impairment are common reasons why inconsistent effort may be evident, even with the evaluee performing at a maximal level on each repeated measure. Other factors include measurement error, task distraction, learning curve, evaluator error, etc. These factors should be carefully documented to assist the evaluator in their review of test results when considering if a high CoVar is representative of inconsistent effort.

REPEATED MEASURES

Repeated measures test protocols have been developed for the hand grip test and the MTM testing. Standard Error of Measurement is 8% for MTM and 14% for the Jamar hand grip. Since there is some learning curve on MTM tests and they are not performed at a maximal effort then it is suggested that evaluee consistency should not be considered as a concern until CoVar is greater than 10%, and even then the cautions mentioned must be adhered to.

It has been found that experimental subjects instructed to give submaximal effort have been able to do so consistently. Therefore it is reasonable to conclude that CoVar within the acceptable range for the test is evidence of consistent effort but not necessarily evidence that effort was maximal.

ALTERNATE FORMS

Alternate forms testing is the most established and validated method of testing evaluee consistency. It has been used in psychological paper and pencil testing for decades. The methodology involves repeat testing of the same function on a variance of the initial test. There need to be established relational outcomes against which to compare the evaluee's performance. Threats to validity of alternate form testing exist primarily due to change of evaluee function on that factor between testing events, and lack of control over learning curve that might exist on that function. Examples of alternates form testing in the FCE are the five hand grip positions on the Jamar, with the established bell shaped curve, and the Progressive Isoinertial Lifting Evaluation, with variation of lift distances resulting in predictable change in strength data. Alternate forms testing exists for the MTM tests, which have ratio criterion that maintains predictability via alternate forms of testing.

DISSIMULATION

Dissimulation has a long history of being used in psychological testing. Subjects are led to believe they are being tested for a particular trait or characteristic when in fact they are being monitored on another variable. There are ethical considerations when this methodology is used in experimental designs, usually monitored by a Human Subjects Research Committee. Dissimulation is the methodology inherent in the Waddel signs. Waddel developed physical examination techniques that lead the evaluee to believe he/she is being examined for some characteristics while other results are being monitored. Waddel tests also involve expectations of symptoms being simulated in evaluees, contrary to biomechanical function. Dissimulation is used in the clinical setting in tasks such as asking the evaluee to pick up a item dropped 'accidentally' by the evaluator, or writing paper and pencil tests to monitor sitting function.

"It is self evident to find disability behavior in a disability evaluation. Positive findings should not be allowed to diminish the validity of the report, but serve to increase the range of opinions and conclusions the evaluator has available to them."

NATURALISTIC TESTING

Naturalistic testing involves monitoring and measuring the evaluee while they are behaving outside the expected parameters of the evaluation environment. Threats to validity of this methodology involve conditioning that exists so strongly that disability behavior pervades all environments, not just the reinforcing environment. This would suggest false positive findings. Naturalistic testing captured via surveillance video misses some of the facilitative factors that enhance behavior. This can lead to false negative findings. Controlled naturalistic testing may be the optimal condition for this methodology. Observation and measurement of pre-determined tasks while the evaluee is on a 'break' from testing can be revealing. An example is arranging the break area so the evaluee has to walk to a break area, stoop and reach to get refreshment, arise, walk and sit. The results from this test can be compared to the 'test behavior' results from formal testing and measurement.

LIE SCALE

A methodology used in psychology, but not implemented in Functional Capacity Evaluation yet, is use of a validity or lie scale. The MMPI has a series of questions that are extremely rare to answer in the affirmative. An example of a validity question is "I never lie". Other tests use repetition of the same items twice within a lengthy test. An example of a repeated question is:

I get headaches:

a) one or more per day;
b) three or more per week;
c) three or more per month;
d) less than three per month.

This methodology holds some promise for development of a series of items that could be interspersed into a paper and pencil functional capacity self report. An examples of a validity item would be "My symptoms remain the same regardless of what I do". The affirmative would be very rare.

It does great disservice to the science of Functional Capacity Evaluation for a report to conclude "The evaluee demonstrated signs of inconsistent performance and symptom magnification throughout the evaluation, therefore the results of this evaluation are invalid".

CONCURRENT VALIDITY SELF REPORT SCALES

Scales have been developed that self report pain, exertion, and functional tolerance. Published concurrent validity data can assist in interpretation as to the minimization, normalcy or exaggeration of the self report scales. The Borg perceived exertion scale (RPE) has been used widely in concurrent validity studies, both as criterion and dependent variable (Carton and Rhodes, 1985). The Borg Scale has shown .85 correlation to heart rate. Consequently concurrent heart rate monitoring and RPE measurement can lead to interpretation of the evaluee's self perception as being within normal limits, minimized or exaggerated.. Heart rate monitoring has some validity problems however, as heart rate is subject to effects of anxiety, tobacco, caffeine and chronic pain attenuation. Using resting heart rate as benchmark diminishes some of these concerns.

Self report functional and pain scales can be compared to objective measures from the FCE. Caution must be applied when differences in self report and observed performance are noted because many evaluees are quite unaware of their limits and abilities and are unreliable sources of estimation.

When indicators of poor evaluee reliability and consistency are present the most realistic interpretation might be that the evaluee is unaware of their true abilities and would benefit from an activity program, work hardening, reconditioning, counseling, etc.

VIGILANT OBSERVATION

Skilled observation requires focus, sensitivity, objectivity and much practice. Most evaluators will have had the opportunity to hone these skills over years of practice.

Focus and perceptual skill requires suspension of judgment in favor of highly vigilant perception. Avoid judgment that filters sensory input into selective perception.

DOCUMENTING BEHAVIOR

Documentation of behavior supercedes judgment about the meaning of that behavior. It is more appropriate to document that "the evaluee stopped the lifting task, groaned and reached for their lumbar area", than it is to 'judge' that the evaluee "reached their maximal lifting capacity". The evaluator must avoid 'bias' based on personal values, expectations or other information in the evaluee file. Some evaluators practice a principle of not reading clinical information before the evaluation, while others have the clinic administration purposefully keep the evaluator naïve to the referral source to remove bias.

Behavior analysis serves to validate or invalidate test results. Poor dexterity scores would be further validated by observation of penmanship during paper and pencil intake questionnaires. Coping skills can be observed during the 'stress' of testing. Anxiety indicators, such as evaluee questions and voiced concerns, crying, request for feedback about performance can be helpful indicators for case management.

The evaluator should avoid influencing evaluee behavior outside the parameters of the test protocols. Behavior reinforcement is a powerful factor and should be carefully controlled.

Research has demonstrated that overt pain documentation increases its frequency (Eastwood et al., 1998).

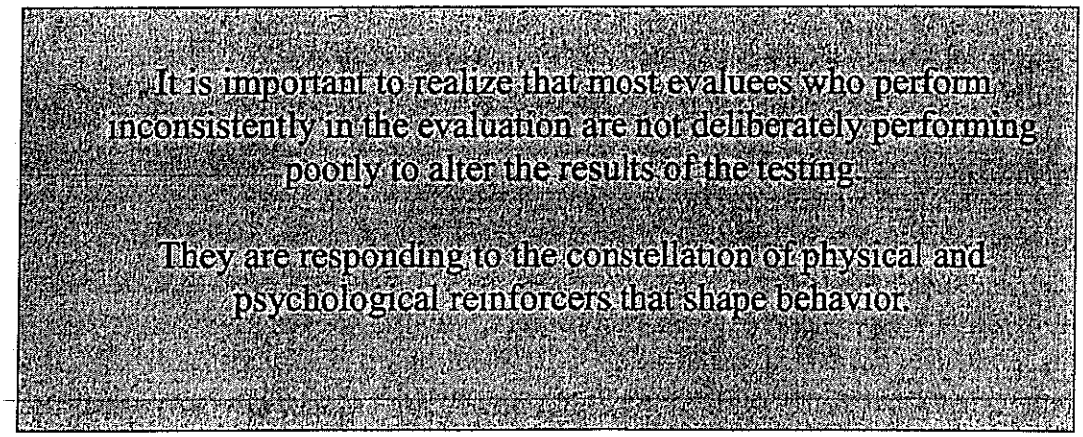
METHODS OF OBSERVATION

There are three methods of observation: narrative, event sampling and time sampling. The three methods are not mutually exclusive and any or all three can be used.

Narrative observation occurs at no set time or event. Observations can be documented in real time or at a convenient time to write the narrative. Observations usually document remarkable occurrences. The bias of narrative observation is that the preponderance of documentation is remarkable and the trend of 'normal' behavior is not revealed.

Event sampling looks for a particular behavior to occur and records it each time it happens. Event sampling is most appropriate when concerned about a particular behavior such as pain behavior. A checklist or other appropriate recording instrument should accompany this observation method. Caution must be exercised to try and limit the effect of behavior sampling on its reinforcement by performing this task as covertly as possible.

Time sampling is the most systematic method. This method assumes that overt behaviors occur on a fairly regular basis. It entails selecting a period of time averaging 10 minutes per hour and systematically recording functional and work related behaviors. Behavior criterion should be predetermined, appropriate to the particular evaluee, with rating scales. Criterion examples (and ratings) are down-time (minutes), verbal and non-verbal symptom indicators (1-3 indicators, 3-5 indicators, 5 or more indicators).



COMMON BEHAVIORAL OBSERVATION ERRORS

A common error in behavioral observation for health care professionals is the *tendency to be lenient.* The evaluator must try to use the most objective measures possible and leave judgment about the behavior to the interpretation stage of the report.

Another error is based *on central tendency.* Evaluators have to consider that there will be some high and low ratings and avoid always selecting ratings towards the average.

The *halo effect* concerns the evaluee with some attractive qualities not being objectively documented for their undesirable behavior.

Logical error in rating exists when a behavior is expected due to its relationship to behavior already documented. An evaluee who displayed stoic pain behavior during a carry from floor task at 8:00 a.m. might be expected to have the same behavior on repeated testing at 10:00 a.m. To read this subtle behavior into the second evaluation is a mistake unless it is clearly observed.

Contrast error exists when an evaluator rates others in the opposite direction from their perception of themselves on a trait. High regard for a personal trait leads to the tendency to rate all others as low on that trait. Athletic evaluators who have coped stoically with injury rehabilitation might be biased in their ratings of evaluees who overtly demonstrate their discomfort.

Proximity error exists when tasks are not separated by much time. In the FCE a noticeable gait behavior during the first walk task might be read into the second walk task, but in actuality was a spasm that resolved itself based on the first activity.

When remarkable findings are discovered in psychosocial and/or psychological evaluation then interpretation of FCE results should be considered in the direction the traits might suggest. A depressed evaluee who has other indicators of inconsistent or poor effort would 'probably demonstrate greater work capacity if diminished effort could be resolved'. Similarly, conclusions need to be adjusted for an evaluee with remarkable findings on substance abuse screening, or who has evident secondary gain for compensation <e.g., has no child care support>, and indicators of inconsistent or poor effort.

INTERPRETATION

Interpretation of behavior should follow the guideline of 'keep it simple'. A simple, commonsense explanation of behavior does not require a license in psychology. If the evaluee is unwilling to perform a task following a painful spasm in the previous task it can be concluded that they were symptomatic and not uncooperative. Restrict interpretation to the present situation. The behavior may not arise in a more natural setting.

Look for trends in behavior. If a male evaluee had a male evaluator, and was resistant and hostile, but smiled, joked and flirted with female clinic staff there may be advantage to scheduling the afternoon follow up with a female evaluator.

BEHAVIOR CHECKLIST

- ☑ Hygiene and Dress
- ☑ Initiating Behaviors
- ☑ Odd Behaviors
- ☑ Communication Skills
- ☑ Vitality
- ☑ Stamina
- ☑ Steadiness
- ☑ Quality
- ☑ Production Rate
- ☑ Attendance
- ☑ Punctuality
- ☑ Frustration Tolerance
- ☑ Personal Complaints
- ☑ Distractibility
- ☑ Safety
- ☑ Social Skills

It is reasonable for the appropriately licensed evaluator to conclude "The evaluee demonstrated signs of inconsistent performance and symptom magnification throughout the evaluation, therefore the results of this evaluation are adjusted to project a range of abilities that would be expected with resolution of the causes of the evaluee inconsistency".

SUMMARY

The learning objective of this section was to:

- ✓ Introduce evaluee reliability indicators
- ✓ Acquaint the evaluator with interpretation guidelines
- ✓ Outline the major issues in evaluee reliability

---

evaluee reliability

LEARNING EXERCISE:

The participants will develop an effort rating scale. Each group will develop three items for their assigned parameter and design a concurrent validity test.

(see next page)

EFFORT RATING SCALE

Parameter	Item
Vocal Indicators: Verbal	1)
Concurrent validity criterion:	2)
	3)
Vocal Indicators: Non-Verbal	1)
Concurrent validity criterion:	2)
	3)
Down-Time (time spent inactive):	1)
Concurrent validity criterion:	2)
	3)
Body Language:	1)
Concurrent validity criterion:	2)
	3)
Physiological:	1)
Concurrent validity criterion:	2)
	3)
Biomechanical:	1)
Concurrent validity criterion:	2)
	3)
Psychophysical:	1)
Concurrent validity criterion:	2)
	3)
Motivational:	1)
Concurrent validity criterion:	2)
	3)
Emotional State	1)
Concurrent validity criterion:	2)
	3)
Locus of Control	1)
Concurrent validity criterion:	2)
	3)
Other parameters	1)
Concurrent validity criterion:	2)
	3)

evaluee reliability

REFERENCES

1. "Disability exaggeration as a predictor of functional restoration outcomes for patients with chronic LBP." SPINE 16.9 (1991).
2. "Functional Disability Scales for Back Pain." SPINE 20 (1995): 1943.
3. "Hannover Functional Questionnaire in Ambulatory diagnosis in Functional Disability caused by backache." Rehabilitation 35.1 (1996): I-VIII.
4. "Measuring the Functional Status of Patients with Lower Back Pain (4 Questionnaires)." SPINE 20.9 1017-1028.
5. "Validity of patient reporting and predictive values of industrial physical work demands." SPINE 19.8 (1994): 888-893.
6. Applewhite, P. B., Paulhe, G. P., and Thompson, D.A. (1965) "Frequency Distribution Shape and Work Output". Perceptual and Motor Skills, 20: 407-8
7. Berryhill, Bill H., "Horizontal Strength Changes: An Ergometric Measure For Determining Validity of Effort in Impairment Evaluations a Preliminary Report". Journal of Disability Volume 3, Numbers 1-4
8. Birmingham TB; Kramer JF "Identifying Submaximal Muscular Effort: Reliability of Difference Scores Calculated From Isometric and Isokinetic Measurements". Percept Mot Skills, 87(3 Pt 2): 1183-91 1998 Dec
9. Birmingham TB; Kramer JF; Speechley M; Chesworth BM; MacDermid J " Measurement Variability and Sincerity of Effort: Clinical Utility of Isokinetic Strength Coefficient of Variation Scores". Ergonomics, 41(6):853-63 1998 Jun
10. Blakley,-Barry-R.;Quinones,-Michael-A.,Crawford,-Marnie-Swedlin; Jago,-I.-Ann Jeanneret and Assoc, Houston, TX, USA "The Validity of Isometric Strength Tests". Personnel-Psychology. 1994 Sum; Vol 47(2): 247-274
11. Burgess, PR; Jones LF "Perceptions of Effort and Heaviness During the Size-Weight Illusion". Somatosens Mot Red, 14(3):189-202 1997
12. Carlson, Jodi L "Evaluating Patient Motivation in Physical Disabilites Practice Settings". American Journal of Occupational Therapy. 1997May; Vol 51(5):347-351.
13. Chan CW, et al. "The pain drawing and Waddell's nonorganic physical signs in chronic low-back pain." SPINE 18.13 (1993): 1717-1722.
14. Chengalar SN, et al. "Assessing sincerity of effort in maximal grip strength tests." Am J Phys Med Rehab 69.3 (1990): 148-153.
15. Department of Physical Therapy, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Israel. "Coefficient of Variation in Maximal and Feigned Static and Dynamic Grip Efforts". Am J Phy Med Rehabil, 78(3):216-21 1999 May-Jun
16. Dvir Z "Reproducibility of Performance and Certainty of Judgment in Maximal vs Feigned Muscular Effort". Percept Mot Skills, 88(3 Pt 2):1078-80 1999 Jun
17. Eastwood, J. D., Gaskovski, P., Bowers, K.S. "The folly of effort: Ironic effects in the mental control of pain". Int J of Clinical and Experimental Hypnosis. 1998 Jan; Vol 46 (1):77-91.
18. Etcoff, Lewis M.; Kampfer, Karen M. "Practice Guidelines in The Use of Symptom Validity and Other Psychological Tests To Measure Malingering and Symptom Exaggeration in Traumatic Brain Injury Cases". Neuropsychology Review. 1996 Dec; Vol 6(4); 171-201.

19. Fairbanks JCT, FRCS, et al. "The OSWESTRY Low Back Questionnaire." BMJ 300 (1992).
20. Fairbanks JCT, FRCS, et al. "The OSWESTRY Low Back Questionnaire." Physiotherapy 66.8 (1980): 271-273.
21. Fishbain DA; Abdel-Moty E; Cutler RB: Rosomoff HL; Steele-Rosomoff R " Detection of a "Faked" Strength Task Effort in Volunteers Using a Computerized Exercise Testing System". Am J Phys Med Rehabil, 78(3):222-7 1999 May-Jun
22. Goldman S; Cahalan TD; An KN "The Injured Upper Extremity and the JAMAR five-handle Position Grip Test". Am J Phys Med Rehabil, 70(6):306-8 1991 Dec
23. Harber P; SooHoo K. "Static ergonomic strength testing in evaluating occupational back pain". J Occup Med, 26(12):877-84 1984 Dec
24. Hattori Y; Ono Y; Shimaoka M; Hiruta S; Kamijima M; Takeuchi Y "Test-Retest Reliability of Isometric and Isoinertial Testing in Symmetric and Asymmetric Lifting". Ergonomics, 41(7):1050-9 1998 Jul
25. Hildreth, Davis H., Lister, Graham D. "Detection of submaximal effort by use of the rapid exchange grip". Journal of Hand Surgery. 14(4):742-5 1989 Jul
26. Kaplan GM; Wurtele SK; Gillis D "Maximal Effort During Functional Capacity Evaluations: An Examination of Psychological Factors". Arch Phys Med Rehabil, 77 (2):161-4 1996 Feb
27. Kumar S; Simmonds M "The Accuracy of Magnitude Production of Submaximal Precision and Power Grips and Gross Motor Efforts". Ergonomics, 37(8):1345-53 1994 Aug
28. Lawlis GF, et al. "The development of the Dallas Pain Questionnaire: an assessment of the impact of spinal pain on behavior." SPINE 14.5 (1989): 511-515.
29. Lim, Levan: Browder,-Diane-M.; Sigafoos, Jeff "The Role of Response Effort and Motion Study in Functionally Equivalent Task Designs and Alternatives". Journal-of-Behavioral-Education. 1998 Mar; Vol 8(1): 81-102
30. Luoto S; Hupli M; Alaranta H; Hurri H "Isokinetic Performance Capacity of Trunk Muscles. Part II: Coefficient of Variation in Isokinetic Measurement in Maximal Effort and in Submaximal Effort". Scand J Rehabil Med, 28(4):207-10 1996 Dec
31. MacKinnon,-Scott-Netson "Relating Heart Rate and Rate of Perceived Exertion in Two Simulated Occupational Tasks". Ergonomics.1999 May; Vol 42 (5): 761-766
32. Matheson, Leonard N., Carlton, Richard, Niemeyer, Linda O. "Grip Strength in a Disabled Sample". Industrial Rehabilitation Quarterly 1 (3); 1-7 Fall 1998
33. 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, 19(12);1359-66 1994 Jun 15
34. Neibuhr BR; Marion R "Voluntary Control of Submaximal Grip Strength". Am J Phys Med Rehabil, 69(2):96-101 1990 Apr
35. Niebuhr BR "Detecting Submaximal Grip Exertions of Variable Effort by Electromyography". Electromyogr Clin Neurophysiol, 36(2):113-20 1996 Mar
36. Niebuhr BR; Marion R; Hasson SM "Electromyographic Analysis of Effort in Grip Strength Assessment". Electromyogr Clin Neurophysiol, 33(3):149-56 1993 Apr-May
37. Norregaard J; B" ulow PM; Lykkegaard JJ; Mehlsen J; Danneskiold-Samsooe B " Muscle Strength, Working Capacity and Efforts in Patients With Fibromyalagia". Scand J Rehabil Med, 29(2):97-102 1997 Jun

evaluee reliability

38. Owens, Lester, Buchholz, Rodney L., "Assessing Reliability of Performance in the Functional Capacity Assessment". Journal of Disability Volume 3, Numbers 1-4
39. Rapport,-Lisa-J.; Farcchione,-Todd-J., Coleman,-Renee-D.; Axelrod,-Bradley-N "Effects of Coaching on Malingered Motor Function Profiles". Journal –of-Clinical-and –Experimental –Neuropsychology. 1998 Feb; Vol 20(1): 89-97
40. Robertson, - Loarn-D.; Brodowicz,-Gary-R.; Swafford,-Albert-R "Improved Detection of Submaximum Effort in Upper Extremity Strength and Strength Endurance Performance Testing". Journal-of-Occupational-Rehabilitation. 1997 Jun; Vol 7(2): 83-95
41. Robinson,-Michael-E.; Sadler,-Ian-J.; O'Connor,-Patrick-D.; Riley,-Joseph-L. "Detection of Submaximal Effort and Assessment of Stability of the Coefficient of Variation". Journal of Occupational Rehabilitation. 1997 Dec; Vol 7(4): 207-215
42. Rogers, Richard (Ed); et-al. (1997) In "Clinical Assessment of Malingering and Deception (2nd ed) (pp 301-327) New York, Guilford Press. "Structured Interviews and Dissimulation".
43. Rogers,-Richard; Cruse,-Keith-R. "Assessment of Malingering With Simulation Designs: Threats to External Validity". Law-and-Human-Behavior.1998 Jun; Vol 22 (3): 273-285
44. Roland M, Morris R. "A study of the Natural History of Lower Back Pain. Part I Development of reliable and sensitive measure of disability in lower back pain." SPINE 8 (1983): 141-144.
45. Simonsen JC "Coefficient of Variation as a Measure of Subject Effort". Arch Phys Med Rehabil, 76(6):516-20 1995 Jun
46. Smith et-al. In (1996). "Synopsis of Treatments of Psychiatric Disorders" (2nd ed.). (pp.723-768). Washington, DC, USA: American Psychiatric Press, Inc. . "Somatoform and Factitious Disorders".
47. Smith, Mike "A Theory of the Validity of Predictors in Selection". Journal of Occupational and Organizational Psychology. 1994 Mar; Vol 67(1): 13-31.
48. Stokes, HM. "The seriously uninjured hand--weakness of grip". J Occup Med, 25 (9):683-4 1983 Sep
49. Tredgett M; Pimble LJ; Davis TR "The Detection of Feigned Hand Weakness Using the Five Position Grip Strength Test". J Hand Surg (BR), 24(4):426-8 1999 Aug
50. Vernon H, Silvano M. "The Neck Disability Index. A study of reliability and validity." Journal of manipulation and psychological therapeutics 14.7 (1991).
51. Waddell G, et al. "A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability." PAIN 52 (1993):157-168.
52. Wright, Rex-A "Brehm's Theory of Motivation as a Model of Effort and Cardiovascular Response". New York, NY, USA: The Guilford Press. xv, 683 pp.
53. Yarnitsky D; Sprecher E; Tamir A; Zaslansky R; Hemli JA "Variance of Sensory Threshold Measurements: Discrimination of Feigners From Trustworthy Performers". J Neurol Sci 1994 Dec 20; 127(2):237)

evaluee reliability