Functional Capacity Evaluation

Part 2: Exposing the Most Common Myths in Validity of Effort Testing

James D. St. James*
Darrell Schapmire**

Abstract

Part 1 of this article described multiple variables in the Gordian Knot known as workers' compensation. These variables included and were aggravated by functional capacity evaluation (FCE) validity of effort testing. The weaknesses in using intake interviews, pain questionnaires and various clinical tests and observations as the basis for conclusions related to validity of effort were pointed out. The remaining element of the FCE validity of effort testing, various common assessments of physical strength, were referred to as myths on the same order as the legendary Gordian Knot. The conclusion of

---

* *Partner, X-RTS Software Products, Inc, Professor of Psychology, Millikin University, Decatur, Illinois. Email: jstjames@millikin.edu*
** *Partner, X-RTS Software Products, Inc., Owner, Industrial Rehabilitation Consultants, Hopedale, Illinois. Email: ds@xrts.com*

this article discusses these myths and proposes two previously peer-reviewed distraction-based testing methods as viable alternatives to discredited testing practices.

Introduction

Functional capacity evaluation (FCE) is one of the most controversial areas in the field of disability management and rehabilitation because it purports to objectify how an individual should perform in highly complex situations in the face of physical and psychological limitations created by injury. There are many methodological flaws with clinical *observations* when used to assess validity of effort, as was pointed out in Part 1 of this article. The purpose of Part 2 of this article is to address commonly used methods which are intended to objectively quantify functional abilities during an FCE. This process involves actual measurement of physical output. However, the quantification of function is utterly impossible unless the physical performance data is objectively shown to reflect a claimant's true function. Assessing functional abilities requires a robust assessment of validity of effort. Lacking a sound analysis of validity of effort, the FCE report is nothing more than a description of what the claimant did during the test—and conclusions drawn from the report are legally indefensible.

There are three popular myths in the field of functional assessment. Despite their lack of objective support, the three testing practices related to these myths constitute the majority of the opportunities to collection physical performance data, i.e., actually measurements, which can actually be assessed for validity of effort during an FCE. Historically, however, FCE protocols have failed to take advantage of this opportunity. These myths are:

1. Validity of effort during *hand strength assessments* can be objectively assessed by using the coefficient of variation (COV), Rapid Exchange Grip (REG) testing, and the so-called "Bell Curve Analysis."

2. Dynamic (real life) *lifting capacities* can be estimated adequately by isometric and isokinetic testing, and such testing is capable of objectively classifying validity of effort.
3. "Visual estimation of effort" during lifting assessment is an accurate method of classifying relative levels of exertion, ("light," "moderate," or "heavy") and is capable of objectively classifying validity of effort.

Like all myths, for these ideas to be accepted as "true," they must have at least a plausible element of truth. But we intend to demonstrate here that they are myths, legends and tall tales that have passed for "science" since the first days of the field known as industrial rehabilitation. Following is our defense of this rather startling and profound claim about the "standard" testing practices associated with functional capacity evaluations.

Hand Strength Assessment

The Jamar Dynamometer is by far the most widely used device for measuring grip strength. It is a hydraulic device that is adjustable to five widths. A pinch gauge is a dynamometer which measures pressure with a compression spring and is used to measure finger strength. Standard positions for holding the gauges are recommended by the American Society of Hand Therapists. For all grip and pinch strength testing, the arm of the subject is held next to the body with the elbow in 90 degrees of flexion. During grip testing, all fingers are in contact with the device. The *two-point pinch* is measured with the device held between the extended thumb and index finger. A *three-point pitch* is measured with the device held between the thumb and Digits 2 and 3. A lateral pinch trial is conducted with Digits 2 – 5 fully flexed (making a fist) and the thumb is placed on the uppermost surface of the gauge. During all pinch trials, the face of the gauge should be oriented away from the subject. Grip and pinch trials are typically five seconds in duration.

The three "standard" methods used to classify validity of effort during hand strength testing have an error rate of at least 30%, based on numerous studies that will be cited below. Most of the error occurs in the failure to detect feigned weakness. Many published studies and reviews that have appeared in print in the past 21 years that have called the Unholy Trinity—the COV, the REG, and the Bell Curve—into question. These studies will be cited in this article as they apply to each of these three testing methods. Nevertheless, these methods remain the most widely used in the classification of validity of effort during hand strength tests. Curiously, a serious debate regarding these methods never took place. Instead, many practitioners in the field still routinely use "standard" hand strength assessments, even in the face of substantial published evidence that such testing practices are highly inaccurate.

Coefficient of Variation

The coefficient of variation (also COV or CV) has traditionally been used as a measure of consistency of performance. The COV is a measure of variability. The greater the differences from trial to trial, the higher the COV. (Mathematically, the COV is the standard deviation divided by the mean and expressed as a percentage.) The traditional cutoff point for separating good effort from poor effort has been 15%. The assumption is that a COV above a given cutoff such as 15% is indicative of an invalid effort and variability less than the cutoff indicates a valid effort.

Table 1 illustrates the COVs for three successive grips (in pounds) of different degrees of variability:

There are three primary weaknesses in the use of the COV to classify effort during hand strength assessments. The first is an error in logic. The second arises from the basic anatomy and physiology of the hand. The third involves the choice of a cutoff for declaring that a COV indicates an invalid effort.

TABLE 1

Grip 1	Grip 2	Grip 3	COV
65	70	75	6%
57	70	83	15%
57	70	83	15%
50	70	90	23%

The error in logic is this: If a person makes a series of squeezes that are true attempts at a maximum, they should apply about the same force each time, and the COV should be low. If the COV is large, this is presumptive evidence that the effort was not valid. (We ignore for the moment what constitutes a "high" COV.) So far, the logic is fine. But most users of the standard FCE methodology make the further assumption that *a low COV indicates a valid effort*. It may, of course. But it could also indicate a proficient cheater. It is actually fairly easy for most people to repeatedly reproduce a submaximal effort. Incidentally, this likely accounts for the fact that the errors in classifying effort are more likely to miss a fake than they are to declare a valid effort to be invalid.

Regarding basic anatomy, the hands have a very powerful biofeedback loop which facilitates the production of consistent, submaximal forces. This loop consists of two parts—motor and sensory. The muscles in the hand and forearm that control gripping and pinching have a very high level of enervation. Indeed, the area of the brain devoted to control of the hand is about as large as the area of the brain devoted to movement of the entire trunk of the body. This is why we have such fine motor control of the hands. Furthermore, each bundle of muscle fibers in the hands has relatively fewer fibers, as compared to other parts of the body, such as the biceps. This anatomic feature also accounts for the increased (more precise) control of activities involving the hands. On the sensory side, the hands are especially sensitive to touch, with many nerve endings that indicate touch and stretch of the skin. The muscles that control the hands are also especially sensitive in indicating stretch of the muscle fibers. As with movement, the brain area devoted to interpreting sensory input from the skin and muscles of the hand and forearm is about the same as the area devoted to sensory input from the entire trunk of the body. It should come as no surprise, then, that many people can easily reproduce a submaximal effort—they can fake weakness in hand strength as judged by the COV.

Cutoff for "High" or "Low" COV

Although "15%" is widely used cutoff to classify validity of effort, no study has ever validated that cutoff in a controlled study. Furthermore, no study has ever validated a specific *frequency* of COVs in excess of 15% (or any other cutoff) as being accurate in distinguishing between good effort and feigned weakness in "standard" protocols using only unilateral force testing. Finally, it is noted that there are significant differences between the "standard" commercially available FCE protocols in terms of the number of positions on the Jamar Dynamomter that are considered in the classification of validity of effort and in the percentage of data sets which must have a COV less than 15% to be considered as "valid" or maximum efforts. In other words, there is a general lack of standardization within the field in how to interpret intra-data set variability. It should also be pointed out that some completely cooperative persons may have occasional COVs which exceed the mythical 15% cutoff. Lastly, it is noted that while the industry standard seems to be 15% the authors have not identified any published article which describes a controlled study validating the use of the COV as an index of effort. In fact, multiple studies have rejected the COV—at least as it is most commonly employed during grip testing—as a proper method for use in a clinical setting (Ashford, Nagelburg, & Adkins, 1996; Dvir, 1999; De Smet & Londers, 2003; Fairfax, Balnave, & Adams, 1995, Fishbain, Cutler, Rosomoff, & Rosomoff, 1999; Hamilton, Balnave, & Adams, 1994;

Hoffmaster & Niebuhr, 1993; Lechner, Bradbury, & Bradley, 1998; Shechtman, 1999, 2000, 2001a, 2001b; Shechtman, Anton, Kanasky, & Robinson, 2006).

Rapid Exchange Grip (REG) Testing

REG testing is also based on a valid physiological principle. The REG method was first described by Hildreth, who noted that maximum voluntary grip strength requires 1.0 – 1.5 seconds to generate (Hildreth, Breidenbach, Lister, & Hodges, 1989). Hildreth, however, did not provide specific instructions with regard to standardizing the testing process or standardizing the interpretation of the data.

In REG testing, the claimant is required to make a series of very quick grips, alternating between the hands. If, under these conditions, a rapid, explosive grip is substantially greater than a standard grip of longer duration, this indicates that the longer duration grip is likely to be a submaximal effort. A careful read of the preceding sentence should alert the reader to the challenges in standardizing such a testing approach. For example, one issue is the rate at which the dynamometer is exchanged from one hand to the other. Another difficulty is controlling the time during which force is exerted. Still another basis for challenging this methodology is the number of exchanges between the hands that might theoretically be required to standardize the testing process (Shechtman & Taylor, 2000, 2002; Taylor & Shechtman, 2000). The continuous movement of the needle on the gauge makes it very difficult to accurately track the amount of force that is being generated by either hand. Attempting to reset the gauge to zero after each squeeze defeats the purpose of the technique, which is to rapidly transfer the gauge from one hand to the other several times. In fact, the needle on the Jamar Dynamometer or any similar device displays an apparent "force" measurement when the device is simply bumped with the palm of the hand—and when no grip whatsoever has been performed.

Most disturbing from the standpoint of interpreting REG data is that no controlled study has ever validated the following: a method which prescribes the rate of exchange of the testing instrument, the duration of the contractions or the number of trials in each hand, the number of positions REG testing should be conducted, or the specific position on a dynamometer that is useful for this alleged purpose. As a result, there is no standardization *between the various testing systems* which use REG as an index of effort. Lastly, no controlled study has ever identified the maximum degree of variability between an REG and a grip of longer duration that is either acceptable or unacceptable in classifying validity of effort. Multiple published studies, in fact, have challenged the appropriateness of the REG for the purposes of classifying validity of effort (Shechtman & Taylor, 2000, 2002; Taylor & Shechtman, 2000; Tredgett & Davis, 2000; Westbrook, Tredgett, Davis, & Oni, 2002).

In addition to the methodological problems with REG testing, there is a logical flaw parallel to that of the COV. If the rapid grips are stronger than standard grips of greater duration, that is presumptive evidence of an invalid effort. But if they are not, that does not necessarily indicate that the effort was valid. Again, it could also indicate a proficient cheater.

The Bell-Shaped Curve

The "Bell Curve Analysis," like the above statistical concepts, is based on a valid physiological principle, but misuses it. The tenet in this case is the fact that maximum strength is produced in the mid-range of motion. This is called the length-tension relationship, the result of maximum overlapping of the actin fibrils and myosin heads in striated muscle being found in the mid-range of motion. Therefore, Stokes hypothesized that maximum grip on the Jamar Dynamometer would be registered in Position 2 or Position 3, the mid-range of motion, and less force would be produced in Positions 1, 4 and 5 (Stokes, 1983). Thus, when a graph is produced by drawing a line indicating the amount of force in each of the five positions a Jamar Dynamometer, the "Bell Curve" supposedly would be identifiable by persons trained in the mysterious art of graph interpretation.

Consider the graphs below. Which one(s) have "significantly more" grip force depicted in Positions 2 and 3? In other words, which one(s) indicate a good effort during grip testing?

FIGURE 1

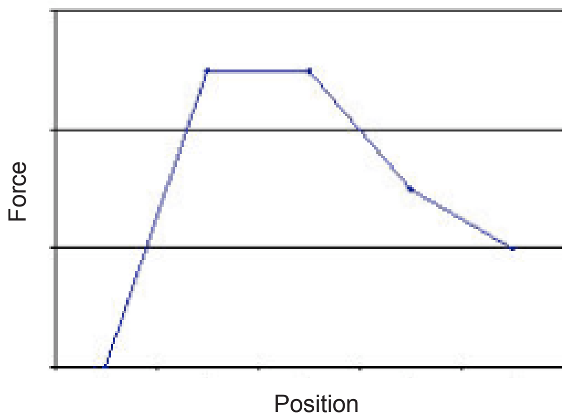


FIGURE 2

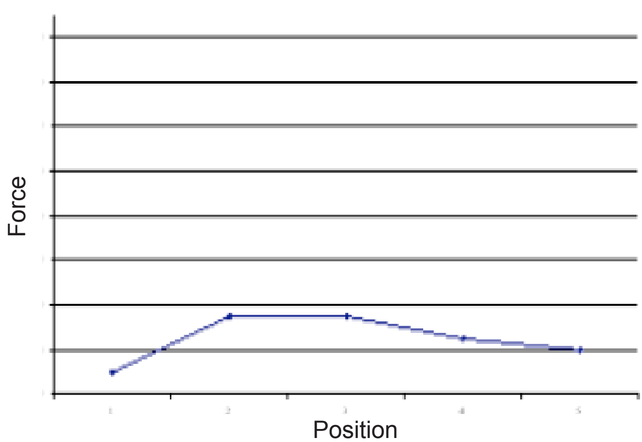


FIGURE 3

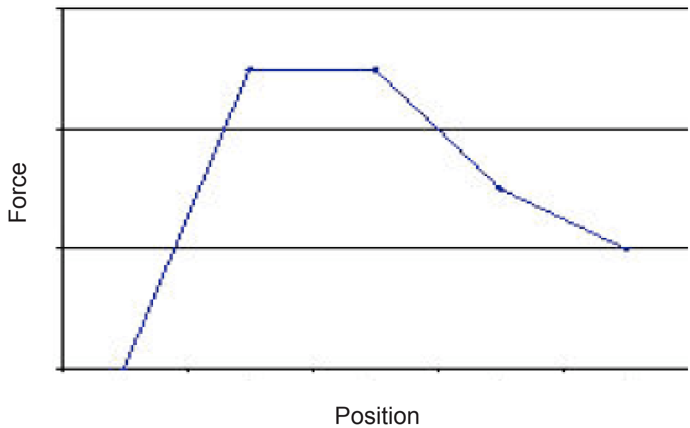
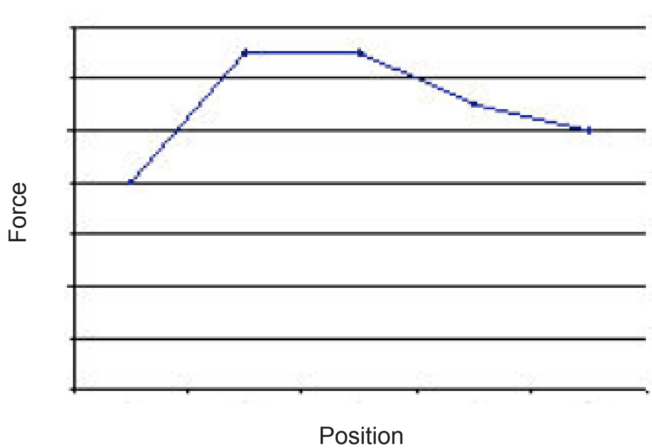


FIGURE 4



These are the correct answers: A) None of them. B) All of them. C) Any graph you chose. The graphs all depict identical data. Only the sizes, scales of the axes and the amount of space above and below the line have been altered. As shown, any interpretation of a line on a graph is not objective because:

1. The size of graphs is not standardized—and size can affect the impression of whether or not the curve is sufficiently high or whether it is in the mid range of motion.
2. The relative scaling of the x and y axes on such graphs are not standardized. In other words, the display can be distorted by "stretching" the graph horizontally or vertically. Doing so distorts the appearance of the degree of curvature.
3. There is no standardized approach with regard to how much space should occur above and below a line in a line graph. This space also affects impression.
4. There is no way to standardize interpretations.

Although the original Stokes study is one of the most widely cited "supportive" references in FCE bibliographies, few people realize that the study had only two subjects. In this case, again, an "evidence-based" and "reasonable" assumption has not been subjected to rigorous scrutiny. Rather, impressions such as the "interpretation" of the "Bell Curve," which are impossible to standardize between observers, have become accepted and promoted as being accurate measures of validity of effort—and the mantle of "analysis" has been conferred upon the process.

In 1995, Stokes revisited the concept of the Bell Curve. Using the Greenleaf System, a fully computerized testing apparatus, the combined data pertaining to "the curve" and REG, he reported accuracy which approached 100% (Stokes, Landrieu, Domangue, & Kunen, 1995). Very few clinicians use the Greenleaf testing device, possibly because it costs more than $15,000. In the same article, Stokes described a variation of this high-tech approach which used the standard mechanical Jamar Dynamometer. There were 38 subjects in this part of the study, all predicted to give "low effort" during hand dynamometry, based on the use of operational definitions proposed by Stokes. Sensitivity to "low effort" in this group of subjects was 84.2%. However, the study did not investigate the specificity of Stokes' operational definitions, i.e., the ability to correctly identify persons who are *not* predicted to give low effort. As a result, the false positive rate for low effort is unknown. By extension, the accuracy of the method is also unknown.

No study has ever confirmed the accuracy of Stoke's "analysis" of the so-called "Curve." There are, however, several studies that call the concept into question, both in the attempt to classify the curve simply by visual inspection or by using various mathematical descriptors (Goldman, Cahalan, & An, 1991; Guitierrez & Shechtman, 2003; Niebuhr & Marion, 1987, 1990; Shechtman, Guitierrez, & Kokendofer, 2005; Sindhu & Shechtman, 2011; Tredgett, Pimble, & Davis, 1999).

Integrated methods

Some testing protocols attempt to use a "holistic" or "whole picture" approach by "integrating" all three standard concepts pertaining to hand strength assessment. A defensible approach would require establishing a system that would assign a relative mathematical weighting to each of these variables, unless it is established that all of them are equally important. However, no research on the relative importance of these three variables has been brought forth. In some FCE protocols, though, "majority rules," and, thus, the interpretation with regard to validity of effort is in accordance with the preponderance of the evidence. In other protocols, a decision is rendered by the evaluator, and offered as a "professional opinion." The absence of standardization in the classification process when attempting to integrate the three "standard" hand strength methods perhaps illustrates the field's confusion on this issue.

Testing: Isometric and Isokinetic Strength Measurement

Isometric Tests of Lifting Capacity

Isometric strength is measured when a subject exerts force against an immovable object. Because there is no movement of the body isometric testing is also called "static" testing. During isometric testing, the subject assumes the test position, grasps the handles of a strain gauge and exerts force. Typically, isometric trials last five seconds. Subjects are asked to gradually increase force for the first one or two seconds of the activity, then progress rapidly to a maximum level of exertion. The most commonly administered tests are the Static Leg Lift, Static Arm Lift, High-Near Lift, High-Far Lift, and the Torso Lift. Static pushing and pulling strengths are also measured in some protocols.

Most jobs that require lifting require dynamic lifting, in which an object is moved through space. Lifting against an immovable object is seldom required for most jobs. Thus, isometric resistance is typically encountered only The chief motivation for using isometric testing is ease of use. To test *dynamic* lifting, a person must lift an increasing workload until they perceive that they are at their maximum safe lift. On the other hand, a maximum isometric strength is measured in a single lift or, at the most, by averaging the results of three trials. It would be very convenient if one simple measurement in a five second period of time could predict how much physical work an individual is be capable of performing. This *unproven assumption* has been used successfully market isometric testing equipment.

There are several problems with the use of isometric testing. These concern the accuracy of predicting dynamic lifting capacity from static measures of strength, the problem of assessing validity of effort, and the basic safety of isometric testing. These problems are discussed below.

Isometric Testing: Predicting dynamic capacity from static testing

It is now known that measures of isometric strength do not accurately predict dynamic lifting capacity. (Feeler, St. James & Schapmire, 2010) In this study of data from tests of 134,000 job applicants to predict floor-to-waist dynamic lifting from static leg lifts. That prediction will be within 31 pounds of the correct answer 68% of the time. Put another way, it will be in error by *more* than 31 pounds 32% of the time. Such predictions may result in these potential outcomes:

1. If used to make hiring decisions, there will be a disparate impact on female job applicants because males have higher isometric strengths than females.
2. If used for hiring purposes, or for the purpose of predicting the dynamic lifting capacity of an injured worker who is returning to work, it is possible to place persons in jobs that are too physically demanding. Therefore, the prospect of a negligent assignment arises as a potential legal liability.
3. In many cases, indemnity for returning workers will be grossly miscalculated—with errors in both directions, i.e., over- and under-compensation.

Isometric Testing: Assessing validity of effort

Isometric strength testing has also been promoted as a method to classify validity of effort. The "reasonable premise" for using isometric assessment to classify effort is similar to the thinking behind standard hand strength assessment: "Repeated measures will be consistent when an individual gives a maximum voluntary effort." The weakness of standard hand strength testing, however, is also present during isometric assessment. Specifically, the methodology ignores the same biofeedback loop that defeats the standard use of the COV in classifying effort during hand strength assessments. In fact, the biofeedback loop that was discussed previously may arguably be enhanced during isometric strength testing, simply because additional sensory biofeedback is being received from the multiple joints and muscles involved in the testing process.

The isometric lifts most often used for this purported purpose are the so-called National Institutes of Occupational Safety and Health (NIOSH) lifts. Of these, the Static Leg Lift and Static Arm Lift were two of the most commonly administered tests. Although these isometric tests have been used to classify validity of effort for more than 30 years, until recently there had not been a single controlled study to verify the accuracy of such testing. The results of the first controlled study on this topic have now been published (Townsend, Schapmire, St. James & Feeler, 2010). This study demonstrates that isometric testing *does not*, in fact, accurately classify validity of effort. In fact, it is now known with a 95% confidence interval that 40% - 80% of the persons who are tested with the Static Leg Lift and Arm Lift can produce COVs less than 15% during *submaximal* exertion!

Isometric Testing: The issue of safety

If practitioners were paying attention to published studies, the potential safety hazard associated with exerting maximum voluntary effort against an immovable object would have killed this technology more than 20 years ago. Grasping the handles of a testing device and exerting maximum force, even when the handles of the device remain in a fixed position, is analogous to an average person attempting to lift a 500-pound workload, i.e., the workload is not going to move, no matter how hard the person tries to lift the weight. Is this safe? In one study, it was demonstrated that even at submaximal levels of exertion, "structural failure at L3" could be anticipated during isometric testing, based on the sheering and compression forces generated when exerting force against and immovable strain gauge (the functional equivalent of attempting to lift 500 pounds; Hansson, Stanley, Bigos, Wortley, & Spengler, 1984). On the basis of safety alone, isometric testing may put a subject at risk and should not be used by the conscientious clinician.

Isokinetic Strength Testing

Isokinetic resistance is also called "accommodating resistance." Isokinetic workloads are encountered only in the clinical setting on specialized pieces of machinery which are typically used to test the shoulders, knees and (rarely) backs. During the test, movement is maintained at a constant velocity, i.e., 90, 120, 150 or 180 degrees per second. The choice of velocity to conduct the test is completely arbitrary and has no particular clinical or practical significance, although some individuals and some protocols consistently use a specific velocity. Isokinetic equipment adjusts immediately to changes in force produced by the person being tested. If the subject produces more force, the machine responds by increasing resistance. If less force is produced by the subject, the machine responds by decreasing the workload. The first study on isokinetic strength we know of was published in 1974. Typically, the parameters investigated in isokinetic studies are:

1. Peak torque.
2. Time to achieve peak torque.
3. Range of motion.
4. Work (distance x torque).
5. Power (work/time).

The hope was that isokinetic strength could be used to predict dynamic function, to determine functional status (possibly identify orthopedic pathology) and to classify validity of effort. The basic error made in the isokinetic studies is the attempt to use group data to predict function for individuals. Major literature reviews have repeatedly concluded isokinetic strength testing does not predict function, does not identify orthopedic pathology or classify validity of effort have been published (Rothstein, Lamb, & Mayhew, 1987; Dvir, 1991; Newton & Waddell, 1993). A recent dissertation found that even using a sophisticated statistical analysis based on signal detection theory, sensitivity to feigned weakness during isokinetic testing of the knee is only 31% (Sivan, Stevenson, Day, Bardana, Diaconescu & Dvir, 2011).

Visual Estimation of Effort During a Lifting Task

The last, and perhaps the most damaging, myth is related to the "visual estimation of effort" (VEE) during a lifting task. The underlying assumption of this testing approach is simply this: It is possible to visually estimate the relative level of exertion and to assess cooperation simply by making a series of visual observations. "Operational definitions" are provided to define what "light," "medium," and "heavy" lifts presumably look like. They are instructions that tell the observer "what to look for" during the lifting event. By far, the most widely used method of assessing validity of effort during a lifting assessment involves the "visual estimation of effort" approach.

The literature for the visual estimation of effort has focused on:

1. *Inter-tester reliability*. These studies describe the degree to which one tester or rater agrees with another when viewing persons performing lifting tasks.
2. *Intra-rater reliability and the related topic of test-retest reliability*. These studies describe the degree to which a tester is in agreement with himself/herself when watching the same videos of lifting tasks on more than one occasion.
3. *Inter-protocol reliability*. These studies are conducted to determine if two different commercially-available protocols agree with one another.

There is a considerable literature demonstrating generally high levels of agreement between and within observers, protocols and test sessions (Brouwer, Reneman, Groothoff, Schellenkens, & Goeken, 2003; Durand, Loisel, Poitras, Mercier, & Lemaire, 2004; Gouttebarge, Wind, Kuifer, Sluiter, & Frings-Dresen, 2006; Gouttebarge, Wind, Kuijer, & Frings-Dresen, 2004, 2005; Gross & Battie, 2002; Hodelsmans, Dijkstra, van der Shans, & Geertzen, 2007; IJmker, Gerrits, & Reneman, 2003; Isernhagen, Hart, & Matheson, 1999; Lechner, Jackson, Roth, & Straaton, 1994; Lygren, Dragesund, Joensen, Ask, & Moe-Nilssen, 2005; Matheson et al., 1995; Reneman, 2003; Reneman et al., 2004; Reneman, Dijkstra, Westmaas, & Goeken, 2002; Reneman, Fokkens, Dijkstra, Geertzen, & Groothoff, 2005; Reneman, Jaegers, Westmaas, & Goeken, 2002; Rustenburg, Kuijer, & Frings-Dresen, 2004; Smeets, Hijdra, Kester, Hitters, & Knottnerus, 2006; Smith, 1994; Soer, Gerrits, & Reneman, 2006; Soer, Poels, Geertzen, & Reneman, 2006; Tuckwell, Straker, & Barrett, 2002). However, there are few that look at the other equally important question of the *accuracy* of the observations. Note that "reliability" can be perfect (1.0) if the observers all agree—*even if they are all wrong!* In 2005, though, and one of these studies did, in fact, report the following with regard to identifying maximum lifting capacities: "'Maximal' performances were correctly rated in 46% to 53% (healthy subjects) and in 5% to 7% (patients with chronic nonspecific low back pain) of the cases" (Reneman et al., 2005, p. E40).

There are many physical/perceptual explanations, logical reasons with regard to the application of operational definitions, and questions with regard to standardization of the VEE that are discussed fully in a study that will be published in the last quarter of 2011(Schapmire, St. James, Townsend, & Feeler, in press). In that study, it is demonstrated that the VEE is accurate—at a level that is only marginally higher than chance—and, just as important, trained and experienced therapists are no more accurate in their estimations than untrained lay subjects. A pre-publication copy of this study can be obtained from the authors.

Distraction-Based Tests - Cutting the Gordian Knot

Distraction-based testing was first described by Waddell (Waddell, McCulloch, Kummel, & Venner, 1980). It is a concept that employs non-obvious methods in re-testing the same activity. In his study, Waddell compared the straight leg raise to the seated straight leg raise. Since they are essentially both measures of hip flexion, Waddell hypothesized that they were explained by non-physical, i.e., behavioral, factors. He emphasized that such tests must be non-emotional, non-surprising and non-hurtful. None of the protocols based on the most popular myths associated with validity of effort testing have incorporated distraction-based testing into the methodology.

In 2002, a distraction-based test that incorporated simultaneous testing of both hands into a hand strength assessment was shown to be 99.5% accuracy, 100% specificity, 99% sensitivity (199/200 correct classifications in a controlled study; Schapmire et al., 2002). The statistically based validity criteria for the protocol described in Schapmire et al. are failed because of abnormal test behavior (inconsistent performance) and is appropriate for use in a patient population (Schapmire, St. James, Feeler, & Kleinkort, 2010). The pattern of test behavior (consistent or inconsistent performance) on this test predicts test behavior during a lifting assessment in which validity is classified according to variability between repeated measures in a distraction-based protocol (St. James, Schapmire, Feeler, & Kleinkort, 2010). The distraction in the lifting protocol is the visual appearance of a workload when place on a patented lever arm testing apparatus, Figure 5. In this protocol, baseline measurements are compared with the corresponding activities on the testing device. The design of the testing device allows for duplication of the biomechanical variables in the baseline testing (point of initiating lift, distance between hands, distance between feet, etc.). The odds of "successful" cheating during the simultaneous bilateral hand strength test are 1%. The odds of estimating three workloads on the lever arm to the extent that deception would be undetected are also approximately 1%. Therefore, odds of one feigning weakness passing both protocols are .001. These methods appear to be more legally defensible than the methods associated with the common myths of the field.

FIGURE 5

Conclusion

The challenges to all parties involved in managing or trying workers' compensation cases are many. In addition to the medical condition of the claimant, there are challenges which arise from various physical, psychological, financial and economic factors and considerations. The degree to which any of these factors are present or affect a case is unpredictable and cannot be measured with precision. The failure of common tests for validity of effort is now well-understood. These tests have become part of the Gordian knot known as workers' compensation. They feed, not discourage, an expert witness culture, adding yet another challenge to those administering compensation cases. With the new and emerging research cited in this article, it is likely the most commonly administered validity of effort tests may be difficult to defend during legal testimony.

By using validity of effort methods that do, in fact, classify validity of effort objectively and with a high degree of accuracy, the fruitless task of trying to solve an unsolvable problem, essentially trying to untie Gordian Knot, can be discontinued. A more direct approach which assesses test behavior (arguably, the collective result of physical, psychological, financial factors) has been demonstrated to be highly accurate in classifying validity of effort. This approach cuts the Knot.

Subsequently, questions about function and validity of effort which are now "settled in court" or negotiated by attorneys could instead be answered in the clinical setting in most circumstances if the medical condition is properly diagnosed. Combined with accurate medical data, an accurate assessment of validity of effort makes possible the next step in the evolution of compensation systems—evidence-based case management decisions.

References

Sivan A., Stevenson, J.M., Day, A.G., Bardana, D.D., Diaconescu E.D. & Dvir, Z. (2011). *Differentiating between types and levels of isokinetic knee musculature efforts*. Unpublished doctoral dissertation. Queen's University. Kingston, Ontario, Canada.

Ashford, R. F., Nagelburg, S., & Adkins, R. (1996, May). Sensitivity of the jamar dynamometer in detecting submaximal grip effort. *Journal of Hand Surgery [Am]*, 21(3), 402-405.

Brouwer, S., Reneman, M. F., Groothoff, J. W., Schellenkens, J. M., & Goeken, L. N. (2003). Test-retest reliability of the of the Isernhagen Work Systems functional capacity evaluation in patients with chronic low back pain. *Journal of Occupational Rehabilitation*, 14, 119-129.

De Smet, L., & Londers, J. (2003, April). Repeated grip strength at one month interval and detection of voluntary submaximal effort. *Acta Orthopaedica Belgica*, 59(2), 142-144.

Durand, M. J., Loisel, L., Poitras, R., Mercier, S. R., & Lemaire, J. (2004). The interrater reliability of a functional capacity evaluation: the physical work performance evaluation. *Journal of Occupational Rehabilitation*, 14(2), 119-129.

Dvir, Z. (1991). Clinical applicability of isokinetics: A review. *Clinical Biomechanics*, 6, 133-144.

Dvir, Z. (1999, May-June). Coefficient of variation in maximal and feigned static and dynamic grip efforts. *American Journal of Physical Medicine and Rehabilitation*, 78(3), 216-221.

Fairfax, A. H., Balnave, R., & Adams, R. D. (1995, September). Variability of grip strength during isometric contraction. *Ergonomics*, 38, 1819-1830.

Feeler, L., St. James, J. D., & Schapmire, D. W. (2010). Isometric strength assessment, Part I: Static testing does not accurately predict dynamic lifting capacity. *Work*. 37(3), 301-308.

Fishbain, D. A., Cutler, R., Rosomoff, H. L., & Rosomoff, R. S. (1999, December). Chronic pain disability exaggeration/malingering and submaximal effort research. *Clinical Journal of Pain*, 15(4), 244-274.

Goldman, S., Cahalan, T. D., & An, K. N. (1991, December). The injured upper extremity and the JAMAR five-handle position grip test. *American Journal of Physical Medicine and Rehabilitation*, 70(6), 306-308.

Gouttebarge, V., Wind, H., Kuifer, P. P., Sluiter, J. K., & Frings-Dresen, M. H. (2006). Reliability and agreement of 5 Ergo-Kit functional capacity evaluation lifting tests in subjects with low back pain. *Archives of Physical Medicine and Rehabilitation*, 87, 1365-1370.

Gouttebarge, V., Wind, H., Kuijer, P. P., & Frings-Dresen, J. H. (2004). Reliability and validity of functional capacity evaluation methods: a systematic review with reference to Blankeship system, Ergo work simulator, Ergo-Kit and Iserhange work system. *International Archives of Occupational and Environmental Health* , *77*, 527-537.

Gouttebarge, V., Wind, H., Kuijer, P. P., Sluiter, J. K., & Frings-Dresen, M. H. (2005). Intra- and interrater reliability of the Ergo-Kit functional capacity evaluation method in adults without musculoskeletal complaints. *Archives of Physical Medicine and Rehabilitation*, 86, 2354-2360.

Gross, D. P., & Battie, M. C. (2002). Reliability of safe maximum lifting determinations of a functional capacity evaluation. *Physical Therapy*, 82, 364-371.

Guitierrez, Z., & Shechtman, O. (2003). The effectiveness of the five-handle position grip strength test in detecting sincerity of effort in men and women. *American Journal of physical Medicine and Rehabilitation*, 82, 847-855.

Hamilton, A., Balnave, R., & Adams, R. (1994, July-September). Grip strength testing reliability. *Journal of Hand Therapy*, 7(3), 163-170.

Hansson, T. H., Stanley, J., Bigos, S. J., Wortley, M. K., & Spengler, D. M. (1984). The load on the lumbar spine during isometric testing. *Spine*, 9, 720-724.

Hildreth, D. H., Breidenbach, W. C., Lister, G. D., & Hodges, A. D. (1989, July). Detection of submaximal effort by use of rapid exchange grip. *Journal of Hand Surgery [Am]*, 14(4), 742-745.

Hodelsmans, A. P., Dijkstra, P. U., Van der Shans, C., & Geertzen, J. H. (2007). Test-retest reliability of psychophysical lift capacity in patients with non-specific chronic low back pain and healthy subjects.

Hoffmaster, E., & Niebuhr, B. R. (1993, August). Consistency of sincere and feigned grip exertions with repeated testing. *Journal of Occupational Medicine*, 35(8), 788-794.

IJmker, S., Gerrits, E. H., & Reneman, M. F. (2003). Upper lifting performance of healthy young adults in functional capacity evaluations: a comparison of two protocols. *Journal of Occupational Rehabilitation*, 297-305.

Isernhagen, S. L., Hart, D. L., & Matheson, L. M. (1999). Reliability of independent observer judgments of level of lift effort in a kinesiophysical functional capacity evaluation. *Work*, *12*, 145-150.

Lechner, D. E., Bradbury, S. F., & Bradley, L. A. (1998, August). Detecting sincerity of effort: A summary of methods and approaches. *Physical Therapy*, *78*(8), 867-888.

Lechner, D. E., Jackson, J. R., Roth, D. I., & Straaton, K. V. (1994). Reliability and validity of a newly developed test of physical work performance. *Journal of Occupational Medicine*, *36*, 997-1004.

Lygren, H., Dragesund, T., Joensen, J., Ask, T., & Moe-Nilssen, R. (2005). Test-retest reliability of the progressive isoinertial lifting evaluation (PILE). *Spine*, *30*, 1070-1074.

Matheson, L. N., Mooney, V., Grant, J. E., Affleck, M., Hall, H., Melles, T., et al. (1995). A test to measure lift capacity of physically impaired adults, Part I: DeMecham, J. (2008, June). Under closer scrutiny: Functional capacity evaluations must be evidence based and valid. *Directors in Rehabilitation*, 47-48.

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

Niebuhr, B. R., & Marion, R. (1987, February). Detecting sincerity of effort when measuring grip strength. *American Journal of Physical Medicine and Rehabilitation*, 66(1), 16-24.

Nieburh, B. R., & Marion, R. (1990, April). Voluntary control of submaximal grip strength. *American Journal of Physical Medicine and Rehabilitation*, 69(2), 96-101.

Reneman, M. (2003, December). Introduction to the special issue on functional capacity evaluations: from expert based to evidence based. *Journal of Occupational Rehabilitation*, 13(4), 203-206.

Reneman, M. F., Brouwer, S., Meinema, A., Dijkstra, P. U., Geertzen, J. H., & Groothoff, J. W. (2004). Test-retest reliability of the Isernhagen work systems functional capacity evaluation in healthy adults. *Journal of Occupational Rehabilitation*, 14, 295-305.

Reneman, M. F., Dijkstra, P. U., Westmaas, M., & Goeken, L. N. (2002). Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *Journal of Occupational Rehabilitation*, 12, 269-275.

Reneman, M. F., Fokkens, A. S., Dijkstra, P. U., Geertzen, J. H., & Groothoff, J. W. (2005). Testing lifting capacity: Validity of determining effort level by means of observation. *Spine (Epub)*, 30, E40-46. Retrieved from http://dissertations.ub.rug.nl/FILES/faculties/medicine/2004/m.f.reneman/c10.pdf

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, 22-37.

Rothstein, J. M., Lamb, R. L., & Mayhew, T. P. (1987). Clinical uses of isokinetic measurements. *Physical Therapy*, 67, 1840-1844.

Rustenburg, G., Kuijer, P.P., & Frings-Dresen, M. H. (2004). The concurrent validity of the ERGOS work simulator and the Ergo-Kit with respect to maximum lifting capacity. *Journal of Occupational Rehabilitation*, 14, 107-118.

Schapmire, D. W., St. James, J. D., Townsend, R., Stewart, T., Delheimer, S., & Focht, D. (2002, July-September). Simultaneous bilateral testing: validation of a new protocol to detect insincere effort during grip and pinch strength testing. *Journal of Hand Therapy*, 15(3), 242-250.

Schapmire D. W., St. James J. D., Feeler L., Kleinkort J. (2010). Simultaneous bilateral hand strength testing in a client population, Part I: Diagnostic, observational and subjective complaint correlates to consistency of effort. *Work. 37*(3):309-320

Schapmire, D., St. James, J. D., Townsend, R., & Feeler, L. (in press). Accuracy of visual estimation of effort during a lifting task. *Work*.

Shechtman, O. (1999). Is the coefficient of variation a valid measure for detecting sincerity of effort of grip strength? *Work, A Journal of Prevention, Assessment and Rehabilitation*, *13*(2), 163-169.

Shechtman, O. (2000, January-March). Using the coefficient of variation to detect sincerity of effort of grip strength: A literature review. *Journal of Hand Therapy*, *13*(1), 25-32.

Shechtman, O. (2001a, July-September). The coefficient of variation as a measure of sincerity of effort of grip strength, Part II: Sensitivity and specificity. *Journal of Hand Therapy*, *14*(3), 188-194.

Shechtman, O. (2001b, July-September). The coefficient of variation as a measure of sincerity of grip strength, Part I: The statistical principle. *Journal of Hand Therapy*, *14*(3), 180-187.

Shechtman, O., & Taylor, C. (2000, July-September). The use of the rapid exchange grip test in detecting sincerity of effort, Part II: The validity of the rapid exchange grip test. *Journal of Hand Therapy*, *13*(3), 203-210.

Shechtman, O., & Taylor, C. (2002, January-March). How do therapists administer the rapid exchange trip test? A survey. *Journal of Hand Therapy*, *15*(1), 53-61.

Shechtman, O., Anton, S., Kanasky, W. F., & Robinson, M. E. (2006). The use of the coefficient of variation in detecting sincerity of effort: A meta-analysis. *Work*, *26*(4), 335-341.

Shechtman, O., Gutierrez, Z., & Kokendofer, E. (2005, January-March). Analysis of methods used to detect submaximal effort with the five-rung grip strength test. *Journal of Hand Therapy*, *18*(1), 10-18.

Sindhu, BS & Shechtman O. (2011, January-March). Using the force-time curve to determine sincerity of effort in people with upper extremity injuries. *Journal of Hand Therapy*. *24*(1):22-9.

Simons, G. (2006, Oct). Credibility crisis in FCEs. *Physical Therapy Products*. Retrieved March 23, 2011, from http://www.ptproductsonline.com/issues/articles/2006-10_03.asp.

Smeets, R. J., Hijdra, H. J., Kester, A. D., Hitters, M. H., & Knottnerus, J. A. (2006). The usability of six physical performance tasks in a rehabilitation population with chronic low back pain. *Clinical Rehabilitation*, 20, 989-997.

Smith, R. L. (1994). Therapists' ability to identify safe maximum lifting in low back pain patients during functional capacity evaluation. *Journal of Orthopaedic and Sports Physical Therapy*, 19(5), 277-281.

Soer, R., Gerrits, E. H., & Reneman, M. F. (2006). Test-retest reliability of the WRULD functional capacity evaluation in health adults. *Work*, *26*, 273-280.

Soer, R., Poels, B. J., Geertzen, J. H., & Reneman, M. F. (2006). A comparison of two lifting assessment approaches in patients with chronic low back pain. *Journal of Occupational Rehabilitation*, *16*, 639-646.

Schapmire D. W., St. James J. D., Feeler L., Kleinkort J. (2010). Simultaneous bilateral hand strength testing in a client population, Part I: Diagnostic, observational and subjective complaint correlates to consistency of effort. *Work*. *37*(3), 309-320.

St. James, J. D., Schapmire, D., Feeler, L., & Kleinkort, J. (2010). Simultaneous bilateral hand strength testing in a client population, Part II: Relationship to a distraction-based lifting evaluation. *Work. 37*(4), 395-403.

Stokes, H. M. (1983, September). Weakness of grip—the seriously uninjured hand. *Journal of Occupational Medicine*, *25*(9), 683-684.

Stokes, H. M., Landrieu, K. W., Domangue, B., & Kunen, S. (1995). Identification of low-effort patients through dynamometry. *Journal of Hand Surgery [Am], 20*, 1047-1056.

Taylor, C., & Shechtman, O. (2000, July-September). The use of the rapid exchange grip test in detecting sincerity of effort, Part I: Administration of the rapid exchange grip. *Journal of Hand Therapy*, *13*(3), 195-202.

Townsend R., Schapmire D. W., St. James J. D., Feeler L. (2010). Isometric strength assessment, Part II: Static testing does not accurately classify validity of effort. *Work. 37*(4), 387-394.

Tredgett, M. W., & Davis, T. R. (2000, August). Rapid repeat testing of grip strength for detection of faked hand weakness. *Journal of Hand Surgery [Br]*, *25*(4), 372-375.

Tredgett, M. W., Pimble, L. J., & Davis, T. R. (1999, August). The detection of feigned hand weakness using the five position grip strength test. *Journal of Hand Surgery [Br]*, *24*(4), 426-428.

Tuckwell, N. L., Straker, L., & Barrett, T. E. (2002). Test-retest reliability on nine tasks of the Physical Work Performance Physical Evaluation. *Work*, *19*, 997-1004.

Waddell, G., McCulloch, J., Kummel, E., & Venner, R. (1980). Non-organic physical signs in low-back pain. *Spine*, *5*(2), 117-125.

Westbrook, W. P., Tredgett, M. W., Davis, T. R., & Oni, J. A. (2002, March). The rapid exchange grip strength test and the detection of submaximal grip effort. *Journal of Hand Surgery [Am]*, *27*(2), 329-333.

James D. St. James, PhD (experimental psychology), is a full-time faculty member of Millikin University, Decatur, IL. In addition, he is a partner in X-RTS Software Products, Inc. He and Mr. Schapmire co-developed the concept of simultaneous bilateral testing of the hands for validity of effort testing. They have worked closely on several successful research projects that have brought new insights to the field for the purpose of objectively classifying test behavior of insurance claimants.

Darrell Schapmire, MS (exercise physiology), is a 1991 graduate of Benedictine University, Lisle, IL. He is a partner in X-RTS Software Products, Inc. and sole proprietor of Industrial Rehabilitation Consultants. In addition to performing functional capacity evaluations, he has patented two proprietary devices and worked closely with his business partner, James St. James, to conduct research focusing on validity of effort testing.