

Simultaneous bilateral hand strength testing in a client population, Part II: Relationship to a distraction-based lifting evaluation

James D. St. James^a, Darrell W. Schapmire^{b,*}, Larry Feeler^c and Joe Kleinkort^d

^aMillikin University, Decatur, IL, USA

^bX-RTS Software Products, Inc., Hopedale, IL, USA

^cWorkSTEPS, Inc., Austin, TX, USA

^dJoseph Kleinkort, PC, Trophy Club, TX, USA

Received 16 April 2009

Accepted 18 June 2009

Abstract. *Objective:* To determine if passing or failing statistically-based validity criteria during a distraction-based hand strength assessment is related to test behavior during a lifting assessment.

Participants: 200 consecutive clients presenting for an FCE.

Methods: The two testing protocols, one involving a hand strength assessment, the other involving an assessment of lifting capacities, were administered to assess the variability between repeated measures.

Results: Clients failing two or more statistically-based hand strength validity criteria had significantly more variability between repeated measures in the lifting assessment, $p = 0.001$ and 0.014 for right and left unilateral lifts, respectively, and $p < 0.0005$ for three different bilateral lifts.

Conclusions: A pattern of performance related to the degree of variability in repeated measures protocols for these two distraction-based protocols is revealed. Passing or failing the hand strength assessment are each equally predictive of test outcome during the distraction-based lifting assessment. The failure of the validity criteria in these two distraction-based tests cannot be attributed to a history of surgery but, rather, is the result of abnormal test behavior.

Keywords: Pattern of performance, lifting assessment, validity of effort, functional capacity evaluation (FCE), maximum effort

1. Purpose

This study is concerned with sincerity of effort in strength testing, which is an ongoing concern in Functional Capacity Evaluation (FCE) testing. This study compares results from the X-RTS Hand Strength Assessment [1] to results from a test of lifting capacity that compares dynamic lifts of standard crates to physically identical dynamic lifts using a lever arm. A large difference between the claimed maximum lifts using the

crate and using the lever arm is suggestive of noncompliance. We predicted that persons whose performance on the hand strength assessment is strongly indicative of feigned weakness will be more likely to also have large discrepancies between the cross-referenced lifts on the crates and lever arm.

The X-RTS Hand Strength Assessment is described in further detail with regard to its use in a client population of persons taking part in a functional assessment in Schapmire [2].

2. Distraction-based methodology for lifting

In addition to demonstrating a link between client responses to benign physical maneuvers and observa-

*Address for correspondence: Darrell Schapmire, MS, X-RTS Software Products, Inc., P.O. Box 171, Hopedale, IL 61747, USA. Tel.: +1 309 449 5483; E-mail: ds@xrts.com.

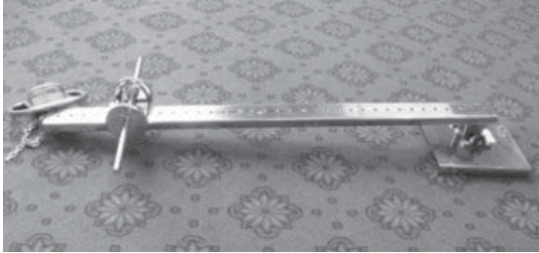


Fig. 1. X-RTS lever arm.

tions and client scores on psychometric measures, Waddell, McCulloch, Kummel and Venner [4] introduced the concept of distraction-based testing. The distraction test specifically mentioned by Waddell et al. was the “Flip Test,” a comparison between the seated and supine straight leg raises. Waddell et al. proposed that such distraction-based tests might be useful in identifying clients who present with exaggerated complaints of pain, stipulating that distraction-based tests must be “non-emotional, non-surprising and non-hurtful.” To the authors’ knowledge, only one previous study assessed the reproducibility of physical effort during a distraction-based test for the hands [1]. No such investigations prior to the current study involved a distraction-based protocol for assessing validity of effort during a lifting assessment.

The second class lever arm testing device in this study is a patented, non-computerized testing device (Fig. 1) developed by the second author. Its configuration replicates the biomechanics required to lift a box containing a workload. The handle plate which is held by the client is configured so as to position the hands 12” apart and place them the same distance from the body as would be required to lift an empty 12” × 12” container. An adjustable clip is used to regulate the length of the chain which connects the handle plate and lever arm, thereby controlling the height from which lifts are initiated. A handle on top of the handle plate is used for unilateral lifting. Thus, the biomechanical factors are controlled. Along the length of the lever are equally-spaced measurement points at which a movable carriage can be mechanically locked into position. Unmarked barbell weights can be affixed to a steel bar on the carriage. By changing the position of the weight and/or changing the amount of weight placed upon the bar, the actual workload can be regulated. Moving any given weight from one location to another results in an actual workload that is predictable because all such movements result in linear changes in the actual workload. Likewise, changing the amount of weight applied

to the bar at any given location also results in linear changes in the actual workload.

3. Study one

We report here a preliminary experiment designed to test the accuracy with which untrained observers can estimate the force required to lift the lever arm. This is directly pertinent to the issue of whether persons can feign weakness during strength testing in the protocol used for the main study. The lifting protocol compares workloads reported by the subjects to be maximum safe lifts, obtained when lifting unmarked weights in a lifting crate and from those obtained on the lever arm. Since the two lifts are, physically, nearly identical, there is no basis for a major disparity in performance, unless a person is attempting perform to estimate lifting the force required to lift the lever instead of simply giving a maximum voluntary effort.

The literature on intuitive physics finds that people generally have a poor knowledge of the physics of simple mechanics, such as motion and force, Sherrin [3], though we know of none that have examined the kind of second class lever used herein.

4. Method

4.1. Subjects

A convenience sample of eight males and 17 females had a mean age of 33.4 years (SD = 15.8). Seven of the subjects (age range 18–22 years) were students at Millikin University in Decatur, Illinois, and were paid \$5 for their participation. The other subjects (age range 19 to 62 years) were employees at two physical therapy clinics. The experiment was approved by the Institutional Review Board of Millikin University.

4.2. Procedure

Subjects were tested in groups of up to five at a time. The subjects stood at the “user’s end” of the device – the end of the device that is lifted – which was lying on the floor. They watched as various configurations of barbell weights were placed on the lever arm.

Subjects were asked to estimate the force, in pounds, that would be required to lift the lever. They were shown a line drawing consisting of a representation of the device, suspended from a scale and were told that

their task was to estimate the reading that would be registered if workloads were to be placed at various locations along the length of the device. They were explicitly told that their task was not to estimate how much weight had been placed on the device, but rather to estimate the actual workload that would result for each of the configurations they would be shown during the experiment. Each subject recorded his or her answer on a data sheet on a clipboard. The subjects were cautioned not to look at each others' answers, or give their answers out loud.

A total of 25 workloads of various configurations of 2.5-pound (1.13 kg), 5.0-pound (2.27 kg) and 10.0-pound (4.54 kg) barbell weights (markings obscured). The same sequence was presented to all subjects. The sequence was not random, but was intended to avoid repetitions of the same position or of the same number of weights, and to cover close to the maximum range of positions and weights. After each estimate, the subjects turned their backs on the lever arm while the experimenter changed the number of weights and their position. When told to by the experimenter, the subjects turned back around to look at the lever arm with the weights attached in a new position and make another estimate. Subjects did not make any actual lifts of the lever arm.

Across the 25 estimations, the amount of weight placed on the lever arm varied from 2.27 kg to 58.97 kg (5–130 lbs). The positions varied from 0 inches from the center of the fulcrum to 64 inches. The actual force required to lift the lever ranged from 5.85 kg to 50.41 kg.

4.3. Analysis

Data analysis was conducted using SPSS and the Excel statistical functions.

4.4. Results

Each trial was scored for each subject as the difference between the actual workload and their estimate. The mean average variation across subjects between the actual workloads and the estimates was 43.0% (SD = 81.7) for signed change and 84.1% (SD = 55.9) for absolute unsigned change. Of the 625 individual estimates, 473 (75.7) had an unsigned error of 25% or more, and 141 (22.6) had an unsigned error of 100% or more.

For individual subjects, the range of average errors was from -106.8% to 301.7% for signed errors and from 38.6% to 301.9 for unsigned errors.

4.5. Discussion

Because the weights used were standard size barbell weights, many of the subjects doubtless knew the amount of weight positioned on the lever, though any advantage gained from this appears to have been more than offset by an inability to also consider the position of the weight on the lever arm. The findings of this experiment are in keeping with the literature on intuitive physics, in replicating the general finding that most people have, at best, a very poor understanding of simple mechanics.

In the case of a client attempting to control the outcome of an FCE to avoid return to work, they may avoid making a lift above an amount needed for return to work. When tested using the lever arm, they would face the difficulty of estimating the force needed to lift the lever. It would be difficult for a client to control the outcome of a test in which workloads were placed upon the device, using a visual estimation of the workloads. As such, the device is useful in a distraction-based, repeated measures lifting protocol, particularly in situations for which secondary gain issues might affect test behavior. Furthermore, the use of such a device meets the aforementioned criteria for distraction-based tests (i.e. "non-emotional, non-surprising and non-hurtful") [4].

5. Study two

The main study examines the relationship between physical performance data in two distraction-based tests in which comparisons are made between repeated measures to classify effort. A more complete description of those tests is presented in Part I of this article [2] and in the original study [1] which demonstrated the effectiveness of a distraction-based test for hand strength. In the protocol, which used simultaneous bilateral testing of the hands as the distraction-based technique, accuracy as 99.5% in classifying validity of effort (199/200 proper classifications) in a non-client population. The authors have found no previous studies identifying a pattern of performance with regard to the reproducibility of physical performance data during multiple distraction-based tests.

6. Research hypothesis

The research hypothesis is that subjects who fail two or more validity criteria during a distraction-based pro-

protocol for assessing consistency of effort during a hand strength assessment will have more variability between repeated measures of a distraction-based lifting protocol than subjects who pass all of the validity criteria for the distraction-based hand strength assessment. In essence, the hypothesis is that compliance during a hand strength assessment is related to consistency of effort during a lifting evaluation.

7. Methods

Test results of 200 consecutive clients who had undergone a functional capacity evaluation (FCE) were compiled. All subjects in this study had applied for benefits in connection to reported work-related injuries or for long term disability status. The Institutional Review Board of Millikin University exempted review of this retrospective analysis of anonymous archival data.

8. Hand strength validity criteria

The hand strength protocol used in this study consists of a randomized order of 66 trials, 48 of which involve unilateral Jamar Dynamometry or pinch strength assessment and 18 of which involve simultaneous testing of both hands. A statistical analysis as described by Schapmire et al. [1] consisting of seven validity criteria is used to classify sincerity of effort as follows:

1. All validity criteria are passed = valid effort.
2. One failed validity criterion = equivocal, or 'gray zone' results.
3. Two or more validity criteria are failed = invalid effort.

9. Lifting activities

If lifting was a critical component of job duties of the claimants, an attempt was made to administer a repeated measures lifting protocol. In its entirety, the lifting protocol was a two-step process consisting of baseline testing with lifted crates that was followed by cross-reference testing on the lever arm. During baseline testing, the workloads were comprised of unmarked rectangular steel bars placed symmetrically in a heavy duty plastic container weighing 1.29 kg (2.85 lbs) and having top side dimensions of 0.30 m × 0.30 m (12" × 12"). For both modes of lifting, the height from which the lifts were initiated was referenced to the distance

of the client's knuckles from the floor. Instructions and a demonstration of safe lifting mechanics were given to each client. It was explained to each client that the goal was to identify a "one-time, safe maximum lifting capacity" for each of the various lifts performed during the test. Each client was also instructed to immediately terminate any lifting activity if he/she believed the workload would be unsafe to lift. Limited only by the client's demonstrated functional ranges of motion, safety considerations and/or willingness to participate, the following lifts were assessed:

1. Bilateral 0.51 m (20") to Waist Lift.
2. Bilateral 0.38 m (15") to Waist Lift.
3. Bilateral 0.25 m (10") to Waist Lift.
4. If right side-involved, Right Unilateral Lift from either 0.25 m or 0.51 m.
5. If left-side-involved, Left Unilateral Lift from either 0.25 m or 0.51 m.

Lifting activities were terminated when any of the following conditions were met:

1. If the client indicated that a "maximum safe level of lifting" had been attained.
2. If the evaluator believed that performing a heavier lift would be unsafe because of radiating pain in an extremity.
3. If the evaluator believed the client's presentation was grossly unsafe secondary to behavioral factors such as refusal to fully grasp the handles of the object being lifted, or gross unsteadiness suggestive of imminent risk of fall.
4. If the client dropped any workload.

Lifting activities were not performed if any of the following conditions were present:

1. The client indicated the need to use a cane or walker on a continuous basis.
2. The client demonstrated the inability to squat to assume the position to initiate a bilateral lift from 0.51 m above the floor.
3. The client refused to participate.
4. The client was not required to perform lifting tasks on the job or if the referral was solely for hand strength assessment.

The results of the baseline testing were cross-referenced by having the client perform corresponding lifts on the class one lever unless the subject lifted the maximum amount of weight required on the job during the baseline testing.

10. Lifting validity criteria

Results were classified as having 'acceptable consistency' between repeated measures during a lifting evaluation if *all* of the following criteria were met:

1. No single set of comparative lifts had variability $\geq 30\%$.
2. At least half of all comparisons had variability $< 25\%$.
3. The average variation between all comparative lifts was $< 20\%$.

Results were classified as 'equivocal consistency' of effort between repeated measures if *all* of the following criteria were met:

1. No single set of comparative lifts had variability $\geq 30\%$.
2. At least half of all comparisons had variability $< 25\%$.
3. The average variation between all comparative lifts were $\geq 20\%$ and $< 25\%$.

Results were classified as having 'unacceptably high variation' between repeated measures if *at least three* of the following criteria were met:

1. At least one set of comparative lifts had variation $\geq 40\%$.
2. Two or more sets of comparative lifts had variation $\geq 30\%$.
3. Mean variation between comparative lifts was $\geq 25\%$.
4. At least half of all comparative lifts have variation was $\geq 25\%$.

It is mathematically possible to obtain test results which do not fit into any of the aforementioned categories. Such results necessarily include data sets with high variability as well as data sets with low variability, an apparent contradiction in behavior that demonstrates neither an obvious pattern of consistency nor an obvious pattern of inconsistency. Lacking any objective evidence of other physical performance testing data which would call into question the test behavior of the subject, such results are classified as 'atypical' and re-testing would be recommended. If other objective indices of effort indicate noncompliance, the lifting assessment classification of effort is a judgment call, left to the discretion of the test administrator.

10.1. Analysis

Data analysis for this portion of the study was also performed with SPSS and the Excel statistical functions.

11. Results

The mean time between injury and the hand strength testing was 18.2 months (SD = 16.1) for persons who failed none of the validity criteria. The mean time between injury and testing for those who failed two or more criteria was 17.7 months (SD = 15.8). In 15 instances no precise date of injury could be identified secondary to conflicting medical records or significant differences between insurance company records and the client's subjective statements regarding the date of accident. For these cases, the date of injury was treated as 'missing data'. Seven of these cases occurred during the testing of subjects who passed all validity criteria, seven during the testing of clients who failed none of the criteria and one for a client producing equivocal results.

Clients whose baseline lifting met job requirements were not tested on the lever arm. Some of the subjects were excluded from all lifting activities for the reasons stated beneath Table 1. Details about diagnoses and behavioral presentations are provided. Chi-square differences between the two groups represented in the table are shown. For subjects passing all hand strength validity criteria, the frequency of lifting weight equal to the amount required on the job was statistically higher than for subjects failing two or more criteria. There were no statistically significant differences between the two groups of clients with regard to the number of subjects who had no lifting on the job, the number of clients who demonstrated the inability to assume the proper posture to perform a bilateral lift from 20", or in the frequency of clients who were unable to complete the lifting assessment because of pain. Five clients who failed two or more hand strength criteria demonstrated the inability to stand without a cane or walker and, therefore, did not take part in a lifting assessment. It is noted, however, that four of these subjects were back clients who failed validity criteria associated with hand strength assessment, i.e., failed validity criteria for the testing of uninvolved parts of the body. No such limitations occurred in the group of subjects passing all hand strength criteria. Similarly, in the group of clients failing two or more hand strength criteria, there were 12 clients who demonstrated the inability to perform at least three lifts of five pounds or more. No such result was obtained for clients passing all hand strength criteria. Lastly, in the group failing two or more criteria, there were 10 clients whose presentation contraindicated conducting a lifting evaluation, for the reasons listed beneath Table 1. By any reasonable standard,

Table 1
Number (percentage) of clients excluded from lifting assessment per test outcome of classification of effort during hand strength assessment

	Group 1: Met on the job lifting requirements (baseline testing only)	Group 2: No lifting on the job or referred for hand strength testing only	Group 3: Demonstrated inability to assume lifting posture	Group 4: Client unable to complete evaluation due to pain or unsafe	Group 5: Client unable to stand without support of cane or walker	Group 6: Completed fewer than three lifts of less than 5 lbs	Group 7: Refusal to participate or other unusual presentation or circumstances
Passed All Hand Strength Assessment Validity Criteria	40/83 (48.2%)	2/83 (2.4%)	3/83 ^[1] (3.6%)	1/83 ^[2] (1.2%)	0	0	0
Group differences per χ^2 , p values ^[3]	χ^2 (1) = 6.81 $p < 0.0005$	χ^2 (1) = 0.03 $p = 0.874$	χ^2 (1) = 0.12 $p = 0.728$	χ^2 (1) = 0.04 $p = 0.851$	NA	NA	NA
Failed Two or More Hand Strength Assessment Validity Criteria	10/108 (9.3%)	3/108 (2.8%)	5/108 ^[4] (4.6%)	1/108 ^[5] (0.9%)	5/108 ^[6] (4.6%)	12/108 ^[7] (11.1%)	10/108 ^[8] (9.3%)

^[1]One client with elbow pain, one cervical spine client with a below the knee amputation, one client lumbar diskectomy.

^[2]One rotator cuff/closed reduction client who lifted 30 pounds, complained of significant increase in symptoms, and was believed by evaluator to be unsafe for additional lifting after baseline testing.

^[3]Compares frequency of Passed All to Failed Two or More Criteria.

^[4]One shoulder pain patient, one client bilateral ulnar nerve release, one cervical spine patient, one client lumbar fusion, one client microdiskectomy.

^[5]One client ulnar nerve release, one client lumbar fusion, one client lumbar diskectomy.

^[6]One lumbar fusion, one wrist client internal fixation who presented using a walker, one sacro-iliac joint fusion patient, one client lumbar diskectomy, one client lumbar diskectomy.

^[7]Four patients with rotator cuff repair, one subacromial decompression patient, one client shoulder pain, one client cervical spine and shoulder pain, one client wrist arthroscopy, one client ulnar shortening, one client cervical spine degenerative disk disease, one client back and knee pain, unable to perform complete lifts on lever arm, one client lumbar laminectomy, one back pain patient.

^[8]One client bilateral upper extremity pain client whose floor length dress was so tight that the client was unable to sufficiently bend at the knees, one low back pain client who refused to lift, one client resection of 9th and 10th ribs who refused to lift lever arm, one fibromyalgia client who insisted on lifting with the hips in maximum abduction (no lifting assessment secondary to safety concerns), one ulnar nerve release client who insisted on standing with one ankle inverted, one fibromyalgia client complaints of frequent falls and demonstrating unstable gait pattern, one wrist fracture client and one bilateral median nerve release client who essentially refused to lift lever arm, one cervical fusion client demonstrating a loss of balance on multiple occasions, one fibromyalgia client complaining of loss of equilibrium, one client unable to lift an empty milk crate 1.29 kg (2.85 lbs.) and complaining of intermittent blindness secondary to low back injury (unconfirmed subjective report).

these presentations lack credibility. Again, no such behaviors were present in the group of clients passing all hand strength validity criteria.

Table 2 reports the results of the lifting evaluation.

The percentage change for each set of comparative lifts was calculated in the manner described beneath the tables (lever arm values being the numerator). Unilateral lever arm lifts were performed only on the symptomatic limb or on the symptomatic side (if a back or lower extremity client) because it is assumed that there is no incentive to under-perform during a test of an asymptomatic part of the body. Persons having 'equivocal' results during the hand strength assessment are omitted from this table due to smallness of sample size, with only seven persons from this group being tested on the lever arm. Only one of eight (12.5%) of the subjects who failed one hand strength criterion performed with 'acceptable consistency' during the lifting evaluation. Another subject from this group lifted weight equal to the amount of weight lifted on the job. Due to sample size, these data are omitted from the table.

Two subjects classified as having 'unacceptably high variability' during the repeated measures lifting protocol completed three baseline lifts, but only two lifts on the lever arm. Average variability for the two lifts was 59.1% for one client and 70% for the other. For all other clients whose data are shown in Table 3, at least three sets, and no more than five sets of comparative lifts were performed.

In Table 3, without exception, for all bilateral and unilateral lifts, the average percent change between baseline and lever arm lifts is lowest for the clients passing all hand strength validity criteria and highest for clients failing two or more validity criteria. In comparing these two groups of clients, there are significant differences in variability between the repeated measures for all bilateral lifts, $p < 0.0005$ in all three cases. P values showing statistically significant group differences during unilateral lifting were seen for the right unilateral lift from 10" ($p = 0.010$) and for the left unilateral lift from 10" ($p = 0.014$).

Table 2
Baseline and lever arm Lifts^[1] per test outcome of simultaneous bilateral hand strength assessment

	Clients passing all hand strength criteria	Clients failing two or more hand strength criteria	t-test results (group differences)
Bilateral Lift from 10" (0.25 m)	N = 31 17.37 kg, 6.40 SD ^[2] 19.23 kg, 6.58 SD ^[3] 22.3% Mean Change, 25.4 SD ^[4] Range = 1.6–107.9%	N = 48 8.80 kg, 5.08 SD ^[2] 13.02 kg, 6.12 SD ^[3] 60.9% Mean Change, 49.5 SD ^[4] Range = 0.9–336.1%	t = 4.01 (77), p < 0.0005
Bilateral Lift from 15" (0.38 m)	N = 37 18.46 kg, 0.93 SD ^[2] 20.94 kg, 8.07 SD ^[3] 20.7% Mean Change, 24.9 SD ^[4] Range = 0.1–107.9%	N = 54 9.30 kg, 4.81 SD ^[2] 13.38 kg, 5.67 SD ^[3] 55.4% Mean Change, 45.8 SD ^[4] Range = 3.6–336.1%	t = 4.20 (89), p < 0.0005
Bilateral Lift from 20" (0.51m)	N = 38 19.05 kg, 7.93 SD ^[2] 20.77 kg, 8.39 SD ^[3] 19.6% Mean Change, 17.6 SD ^[4] Range = 0.9–68.7%	N = 62 8.75 kg, SD 4.76 ^[2] 12.70 kg, 4.99 SD ^[3] 56.0% Mean Change, 37.1SD ^[4] Range = 1.6–156.7%	t = 6.49 (98), p < 0.0005
Right Unilateral Lift from 10" (0.25 m)	N = 10 12.34 kg, 7.94 SD ^[2] 15.15 kg, 6.21 SD ^[3] 49.5% Mean Change, 51.9 SD ^[4] Range = 4.5–128.7%	N = 29 5.90 kg, 3.95 SD ^[2] 10.25 kg, 3.76 SD ^[3] 97.9% Mean Change, 47.4 SD ^[4] Range = 3.0–205.6%	t = 2.72 (37), p = 0.010
Left Unilateral Lift from 10" (0.25 m)	N = 17 12.34 kg, 6.35 SD ^[2] 14.01 kg, 6.35 SD ^[3] 50.2% Mean Change, 38.6 SD ^[4] Range = 0.1–129.1%	N = 28 6.03 kg, 3.18 SD ^[2] 10.34 kg, 3.90 SD ^[3] 86.2% Mean Change, 49.5 SD ^[4] Range = 6.8%–205.8%	t = 2.56 (43), p = 0.014
Right Unilateral Lift from 20" (0.51 m)	N = 2 13.20 kg, 1.68 SD ^[2] 15.15 kg, 31.13 SD ^[3] 13.9% Mean Change, 9.2 SD ^[4] Range = 4.6–23.2%	N = 7 9.03 kg, 4.58 SD ^[2] 11.20 kg, 3.49 SD ^[3] 43.6% Mean Change, 18.8 SD ^[4] Range = 9.7–115.8%	Not applicable, sample sizes too small
Left Unilateral Lift from 20" (0.51 m)	N = 2 17.19 kg, 9.34 SD ^[2] 17.60 kg, 6.12 SD ^[3] 17.0% Mean Change, 0.30 SD ^[4] Range = 16.7–17.4%	N = 8 8.48 kg, 3.18 SD ^[2] 11.29 kg, 3.99 SD ^[3] 42.6% Mean Change, 18.2 SD ^[4] Range = 0.6–91.8%	Not applicable, sample sizes too small

^[1]Includes only clients undergoing both baseline and lever arm testing. Subjects lifting the amount of weight required on the job were not tested on the lever arm.

^[2]Baseline lifts (unmarked steel bars). All clients lifting less weight than required on the job and who were also tested on the lever arm.

^[3]Lever arm lifts.

^[4]The average of all changes for each lift for all subjects, each change calculated with: [(Lever Arm lift/Baseline lift) * (100)]–100.

Table 3 reports the agreement between the classification of validity of effort for the hand strength assessment and the test behavior or presentation during the lifting protocol. Clients who passed all hand strength validity criteria had a statistically higher ($p < 0.0005$) frequency of performing with ‘acceptable consistency’ during the repeated measures testing, as defined in the Methods section, than did clients who failed two or more hand strength criteria. Similarly, clients who failed two or more hand strength criteria had lifting assessment results that were classified as having ‘unacceptably high variability’ at a rate that was significantly

higher than those clients who passed all hand strength validity criteria ($p < 0.0005$).

In Table 3, regarding the degree of consistency between the results of the hand strength assessment and behavior during the lifting assessment, consideration is given not only to the results of the clients who were tested on the lever arm, but also the various presentations that were observed during the test. Clients were classified and grouped, based on behavior. For example, clients who lifted the amount of weight required on the job were considered to be similar to the clients who performed with ‘acceptable consistency’, as defined in the

Table 3
Concurrent validity between hand strength assessment classification of validity and client behavior during the lifting assessment

	Passed all validity criteria during hand strength assessment Number/N (Percentage)	Failed two or more criteria during hand strength assessment Number/N (Percentage)	Group Differences per χ^2 , p values
Met criteria for 'acceptable consistency' during repeated measures lifting evaluation	25/38 (65.8%)	5/61 (8.1%)	$\chi^2 (1) = 36.77, p < 0.0005$
Met criteria for 'unacceptably high variability' during lifting evaluation	12/38 (31.6%)	55 ^[3] /61 (90.1%)	$\chi^2 (1) = 37.74, p < 0.0005$
Convergence between test behaviors in hand strength assessment results and lifting assessment	65 ^[1] /77 ^[2] (84.4%)	77 ^[4] /92 ^[5] (83.7%)	$\chi^2 (1) = 0.0260, p = 0.899$

^[1]Includes all 25 subjects who passed the hand strength assessment and also had lifting results were classified as having 'acceptable consistency' between repeated measures and 40 subjects who demonstrated the ability to perform the heaviest lifting required on the job (no lever arm testing). One 'equivocal consistency' lifting evaluation not included.

^[2]Includes all 37 subjects who completed Baseline and Lever Arm testing (one subject having 'equivocal consistency' during the lifting assessment not included), plus 40 subjects who lifted the amount of weight required on the job (Category 2 clients in Table 1).

^[3]Includes one lifting assessment with results classified as atypical' as described in Methods and Results.

^[4]Includes 55 subjects who failed the hand strength assessment and also had 'unacceptably high variability' during the lifting assessment, 12 subjects from Category 6 in Table 1, and 10 subjects from Category 7 in Table 1. One 'equivocal consistency' lifting evaluation not included.

^[5]Includes 60 subjects who completed Baseline and Lever Arm testing (one 'equivocal consistency' lifting evaluation not included), 10 subjects who lifted the amount of weight required on the job and had no lever arm testing (Category 2 clients in Table 1), and all 22 clients in Categories 6 and 7 in Table 1.

Table 4
Frequency of surgery per test classification for 'passed both', 'failed both', and 'failed one' test for validity of effort

	Passed validity criteria for both tests Number/N (percentage)	Failed validity criteria for both tests Number/N (percentage)	Group differences per χ^2 , p value
History of a Relevant Surgery or Fracture ^[1]	20/25 (80.0%)	32/55 (58.2%)	$\chi^2 (1) = 3.60, p = 0.056$ ^[2]
History of a Relevant Surgery or Fracture ^[1]	Passed Validity Criteria for Both Tests 20/25 (80.0%)	Failed Validity Criteria for One Test 13/18 (72.2%)	Group differences per χ^2 , p value $\chi^2 (1) = 0.36, p = 0.551$ ^[3]

^[1]Relevant Surgery or Fracture denotes a surgery or fracture involving the cervical spine and/or at least one upper extremity, including the shoulder.

^[2]Compares 'Passed All' and 'Failed Both'.

^[3]Compares 'Passed All' and 'Failed One'.

Methods section. Likewise, the clients whose unusual presentations precluded participation in a lifting assessment were considered to have test behavior similar to the clients whose lifting results revealed 'unacceptably high variability'. These tallies were then compared to the results obtained for each of these groups, per hand strength assessment classification (passed all criteria versus failed two or more criteria). In these comparisons, for clients who passed all hand strength validity criteria, 65/77 (84.4%), test behaviors during the lifting assessment were consistent with the hand strength test results. Similarly, for clients failing two or more criteria, 77/92 (83.7%) demonstrated test behaviors that were consistent with the abnormal test behaviors seen in the hand strength test. 'Gray zone' hand strength assessments and lifting evaluations are not included in these percentages.

Table 4 compares the diagnostic status of the subjects per test classification, with reference to whether the subjects had a relevant surgical history or history of a fracture. A total of 25 subjects passed the validity criteria for both tests. Of this number, 20 (80%) had undergone surgery involving the cervical spine and/or at least one upper extremity, including the shoulder. In contrast, there were 55 subjects who failed both tests for validity of effort. Of this number, 32 (58.2%) had undergone surgery involving the cervical spine and/or at least one upper extremity, including the shoulder. Chi-square analysis shows that these differences approach statistical significance, ($p = 0.056$), with regard to frequency of surgery. There were no significant differences between those who passed both tests compared to those who failed one validity test, with regard to frequency of surgery ($p = 0.0557$).

12. Discussion

Test behavior in one of the distraction-based tests in this study can predict behavior in the other with relatively high accuracy. Thus, there is good concurrent validity between the two tests. The classification of test behavior during the hand strength assessment and during the repeated measures lifting assessment were based on uniformly applied statistical analyses; applied in the same manner for all clients. Clients who fail the validity criteria for one of these distraction-based tests tended to either fail the other, or presented with behaviors which were readily judged, by any reasonable standard, as a likely misrepresentation of functional status. However, the correlation between test outcomes is not perfect. Therefore, it is advised that more than one test which relies on empirical data to classify effort be used to assess persons presenting for functional assessment in medical-legal cases.

13. Conclusions

The research hypothesis is validated. We conclude:

1. This study reveals a pattern of performance related to the degree of variability in repeated mea-

asures protocols for these two distraction-based protocols administered to a population of insurance claimants.

2. Passing or failing the hand strength assessment are each equally predictive of test outcome during the distraction-based lifting assessment.
3. The failure of the validity criteria in these two distraction-based tests can not be

attributed to a history of surgery but, rather, it is the result of abnormal test behavior.

References

- [1] D. Schapmire, J.D. St. James, R. Townsend et al., Simultaneous bilateral testing: validation of a new protocol to detect insincere effort during grip and pinch strength testing, *J Hand Ther* **15** (2002), 242–250.
- [2] D. Schapmire, J.D. St. James, L. Feeler and J. Kleinkort, *Simultaneous Bilateral Hand Strength Testing in a Client Population I: Diagnostic, Observational and Subjective Complaint Correlates to Consistency of Effort*, accepted for publication as part of the present study.
- [3] B. Sherin, Common Sense Clarified: The Role of Intuitive Knowledge in Physics Problem Solving, *J Research Science Teaching* **43** (2006), 535–555.
- [4] G. Waddell, J. McCulloch, E. Kummel and R. Venner. Nonorganic physical signs in low- back pain, *Spine* **5** (1980), 117–125.