Scapular Muscle Tests in Subjects With Shoulder Pain and Functional Loss: Reliability and Construct Validity

Background and Purpose. Scapular muscle performance evaluated with a handheld dynamometer (HHD) has been investigated only in people without shoulder dysfunction for test-retest reliability of data obtained with a single scapular muscle test. The purpose of this study was to assess the reliability, error, and validity of data obtained with an HHD for 4 scapular muscle tests in subjects with shoulder pain and functional loss. Subjects and Methods. Subjects (N=40) with shoulder pain and functional loss were tested by measuring the kilograms applied with an HHD during 3 trials for muscle tests for the lower trapezius, upper trapezius, middle trapezius, and serratus anterior muscles. Concurrently, surface electromyography (sEMG) data were collected for the 4 muscles. The same procedures were performed 24 to 72 hours after the initial testing by the same tester. Muscle tests were performed 3 times, and the results were averaged for data analysis. Results. Intraclass correlation coefficients for intratester reliability of measurements of isometric force obtained using an HHD ranged from .89 to .96. The standard error of the measure (90% confidence interval [CI]) ranged from 1.3 to 2.7 kg; the minimal detectable change (90% CI) ranged from 1.8 to 3.6 kg. Construct validity assessment, done by comparing the amounts of isometric muscle activity (sEMG) for each muscle across the 4 muscle tests, revealed that the muscle activity of the upper trapezius and lower trapezius muscles was highest during their respective tests. Conversely, the isometric muscle activity of the middle trapezius and serratus anterior muscles was not highest during their respective tests. Discussion and Conclusion. In people with shoulder pain and functional loss, the intrarater reliability and error over 1 to 3 days were established using an HHD for measurement of isometric force for the assessment of scapular muscle performance. Error values can be used to make decisions regarding individual patients. Construct validity was established for the lower and upper trapezius muscle tests; therefore, these tests are advocated for use. However, construct validity was not demonstrated for the serratus anterior and middle trapezius muscle tests as performed in this study. Further investigation of these muscle tests is warranted. [Michener LA, Boardman ND, Pidcoe PE, Frith AM. Scapular muscle tests in subjects with shoulder pain and functional loss: reliability and construct validity. Phys Ther. 2005;85:1128–1138.]

Key Words: Measurement, Muscles, Reproducibility of results, Scapula, Shoulder.

Lori A Michener, N Douglas Boardman, Peter E Pidcoe, Angela M Frith
The scapula serves as the platform for humeral motion. The scapulothoracic articulation is stabilized and controlled, in part, by the muscles attached to the scapula. Therefore, if scapular muscle function is altered, then dysfunctional scapulothoracic kinematics may result. In construction workers with subacromial impingement syndrome, increased upper trapezius (UT) and lower trapezius (LT) muscle activity and decreased serratus anterior (SA) muscle activity were observed with concurrent alterations in scapular kinematics during glenohumeral elevation. In people without shoulder dysfunction, altered scapular kinematics have resulted from experimentally induced fatigue of the SA and UT muscles. Dysfunctional scapular muscle performance is a contributing factor for scapular dyskinesia.

Muscle performance assessments conducive to clinical practice include manual muscle tests (MMTs), which are commonly used by physical therapists. Manual muscle tests were designed to replicate the primary motion of a muscle while minimizing the contribution of the secondary mover muscles. Theoretically, during an individual MMT, the designated primary mover muscle should have the highest level of activity compared with the secondary mover or stabilizer muscles.

An MMT is performed by an examiner applying force to the individual in the direction opposite the action of the tested muscle’s action. The muscle then is graded with an ordinal grading scale of 0 to 5. The reliability of data obtained with this grading method has not been investigated for MMTs of scapular muscles. Moreover, it is unclear whether the grading categories are mutually exclusive.
exclusive (eg, the grading categories do not overlap). A handheld dynamometer (HHD) can be used to quantify muscle performance to address grading scale limitations.

The reliability of scapular muscle MMT scores obtained using an HHD has been investigated. In intrater test-retest reliability, defined as “the consistency of measurements when one person takes repeated measurements separated in time,” of a single scapular isometric muscle test using an HHD was investigated in 2 studies. The intrater test-retest reliability of measurements obtained with muscle testing of the middle trapezius (MT) muscle, as described by Hislop et al, was reported to be excellent (intraclass correlation coefficient [ICC(1,1)] = .96–.98). Because these studies included only subjects who were healthy and assessment by only a single scapular muscle test, the generalizability of these results to clinical practice is limited.

The validity of data for scapular MMTs has not been investigated. It is unclear whether the designated muscle of a given scapular MMT has the highest contribution during its specified MMT compared with other contributing muscles. To determine the construct validity of data for an MMT, the results for the corresponding muscle can be cross-validated by comparison with another measure of muscle performance. Surface electromyography (sEMG) is 1 such measure that can serve as a comparison measure to assess muscle activity during an MMT. This comparison would allow for the assessment of the construct of muscle performance.

Further investigation is necessary to determine the psychometric properties of scapular muscle isometric force testing with an HHD. Specifically, the test-retest reliability and error values need to be determined to allow users to make judgments about measures of scapular MMTs with an HHD. Construct validity assessment is necessary to provide evidence to indicate the degree to which a meaningful interpretation can be inferred. The first purpose of this study was to determine the intrater test-retest reliability and error values of measurements of isometric muscle force obtained during scapular MMTs with an HHD in subjects with shoulder pain and functional loss. The second purpose was to determine the construct validity of data for scapular MMTs with sEMG as the construct measure for comparison in subjects with shoulder pain and functional loss.

**Method**

Subjects (N=40) were recruited for this study by use of flyers displayed in physician and physical therapist offices. All subjects had complaints of shoulder pain and functional loss and were actively involved in nonsurgical treatment for their shoulder problem in a home exercise program or in supervised physical therapy. The inclusion criteria included self-reported shoulder pain, self-reported loss of shoulder function, and diagnosis of shoulder pathology of the musculoskeletal system by a physician or physical therapist. Exclusion criteria included the inability to raise the arm to 140 degrees of shoulder elevation, because this amount of elevation was required for the performance of the scapular muscle tests. The diagnoses were recorded as summarized in Table 1.

Pain and functional loss were measured by the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment, patient self-report section (ASES). The ASES consists of 2 subscales: pain and functional loss. The pain section contains 1 question: “How bad is your pain today?” The response is indicated on a 10-cm visual analog scale. The function section contains questions regarding a person’s perceived ability to complete 10 activities, graded on a Likert scale of 0 to 3 (0=“unable to do,” 1=“very difficult to do,” 2=“somewhat difficult to do,” and 3=“not difficult to do”). The ASES pain and function subscales are weighted equally (50 points), with the total score ranging from 0 to 100 points (0=“severe pain and functional loss”). The ASES has demonstrated test-retest reliability (ICC[1-way random effects] = .84), construct and discriminate validity, and the ability to be responsive when used to examine outpatients with shoulder dysfunctions. The ASES scores are reported in Table 2.

Subjects were given an informed consent form to read and sign. They were given as much time as they needed to read the informed consent form and to ask questions regarding the study. They were given a copy of their signed informed consent form, and the original was retained in each subject’s record. An intake form with demographic information was completed next. Demographics are reported in Table 2.

### Table 1. Subject Diagnoses

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>No. of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impingement syndrome</td>
<td>14</td>
</tr>
<tr>
<td>Impingement syndrome and instability</td>
<td>2</td>
</tr>
<tr>
<td>Full-thickness rotator cuff tear</td>
<td>6</td>
</tr>
<tr>
<td>Status after rotator cuff repair</td>
<td>3</td>
</tr>
<tr>
<td>Adhesive capsulitis</td>
<td>2</td>
</tr>
<tr>
<td>Total shoulder arthroplasty</td>
<td>1</td>
</tr>
<tr>
<td>Multidirectional instability</td>
<td>6</td>
</tr>
<tr>
<td>Anterior instability</td>
<td>1</td>
</tr>
<tr>
<td>Status after acromioplasty</td>
<td>1</td>
</tr>
<tr>
<td>Status after stabilization procedure</td>
<td>2</td>
</tr>
<tr>
<td>Sternocavicular joint separation</td>
<td>1</td>
</tr>
<tr>
<td>Muscle strain</td>
<td>1</td>
</tr>
</tbody>
</table>
The MMTs investigated in this study were the tests for the LT, SA, MT, and UT muscles. These muscles were selected on the basis of prior evidence indicating that alterations in the performance of these muscles were associated with glenohumeral pathology and dysfunctionsional shoulder kinematics. The methods of Hislop et al were selected for testing of the 3 sections of the trapezius muscle. For the MT and LT muscle tests, the application of force was modified by applying force directly to the scapula instead of indirectly via the humerus. This was done to isolate the scapulothoracic joint and thus eliminate the involvement of other joints. The method of Kendall et al was selected for the SA muscle because this method allows for the application of resistance to scapular protraction.

Table 2.
Subject Characteristics (N=40)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age [y]</td>
<td>42.9 (15.4)</td>
</tr>
<tr>
<td>Height [cm]</td>
<td>167.5 (9.1)</td>
</tr>
<tr>
<td>Weight [kg]</td>
<td>78.9 (21.9)</td>
</tr>
<tr>
<td>Sex [n]</td>
<td>15 Male, 25 Female</td>
</tr>
<tr>
<td>Dominant shoulder [n]</td>
<td>36 Right, 4 Left</td>
</tr>
<tr>
<td>Shoulder tested [n]</td>
<td>21 Dominant shoulder, 19 Nondominant shoulder</td>
</tr>
<tr>
<td>Time since onset of symptoms [n]</td>
<td>1 &lt;1 mo, 6 1–3 mo, 3 3–6 mo, 30 &gt;6 mo</td>
</tr>
<tr>
<td>Pain [n]</td>
<td>9 Constant, 31 Intermittent</td>
</tr>
<tr>
<td>ASES*: pain subscale</td>
<td>36.2 (12.8) Minimum–maximum 12.5–50</td>
</tr>
<tr>
<td>ASES: functional loss subscale</td>
<td>34.6 (10.1) Minimum–maximum 10–50</td>
</tr>
<tr>
<td>ASES: total score</td>
<td>70.8 (20.9) Minimum–maximum 26.5–100</td>
</tr>
</tbody>
</table>

* ASES = American Shoulder and Elbow Surgeons Standardized Shoulder Assessment.

Testing Procedures
On the first day of data collection (day 1), HHD and sEMG measurements were collected during 4 scapular MMT procedures. The subjects’ affected shoulder was tested. Subjects returned 24 to 72 hours after day 1 for a second day of data collection (day 2). Subjects received no form of intervention between testing days. Additionally, subjects were asked the following question: “Since the last time you were here, do you feel that your shoulder has gotten better, stayed the same, or gotten worse?” This question was asked to determine whether the subject remained stable since the first day of testing. On day 2, all subjects (N=40) answered “stayed the same.” On day 2, the subjects performed the 4 tests in the same order of testing as on day 1, but only HHD data were collected.

Surface Electromyography
Testing procedures were initiated with the application of the electrodes for the collection of sEMG data. Hair was removed with a razor as needed, and skin was cleansed with alcohol. Prefabricated, preamplified bipolar surface silver chloride electrodes were placed with conductive gel over the skin. The bipolar surface electrodes contained 2 electrodes, 8 mm in diameter, with an interelectrode distance of 22 mm.

The placements of the 4 electrodes are depicted in Figure 1. With the shoulder in maximum flexion, the LT muscle electrode was placed on a line parallel to the spinal column and approximately 5 cm lateral to the spinous process, at the level of the inferior angle of the scapula. The method of Hislop et al was selected for the SA muscle because this method allows for the application of resistance to scapular protraction.
the arm at 90 degrees of abduction, midway between the spine of the scapula and the spinous process in a position perpendicular to the spine.\(^{16,17}\) Finally, the UT muscle electrode was placed with the arm resting at the side of the body, on a straight line midway between the spinous process of the seventh cervical vertebra and the lateral edge of the acromial process.\(^{17}\) The electrodes were attached to the skin with prefabricated, double-sided adhesive. The ground electrode was affixed with paper tape over the posterior distal ulna at the wrist of the extremity not being tested.

The sEMG data from the 4 electrodes were collected with a Therapeutics Unlimited Model 544 System\(^*\) concurrently during the 4 muscle tests on day 1 only. Data were collected at a sampling rate of 1,000 Hz with a 12-bit, analog-to-digital band-pass filter at a low frequency of 20 Hz and a high frequency of 400 Hz. The sEMG data then were passed through an analog-to-digital converter into a personal computer and stored. The sEMG data were collected for 4 seconds, and the middle 2 seconds were used for data analysis. The raw data were reduced using the root-mean-square (RMS) method.\(^{18}\) Because the sEMG signal oscillates between positive and negative voltages, the signal cannot be summed. The RMS method quantifies the signal by squaring the raw data, summing the squares, dividing this sum by the number of observations, and then taking the square root of that sum. The RMSs of the middle 2 seconds for each trial were calculated, and the average of 3 trials for each muscle during each test was used for data analysis.

**Scapular Muscle Tests**

Subjects performed the 4 scapular muscle tests depicted in Figures 2 to 5.\(^{8,9}\) A Nicolas HHD\(^†\) was used to record the amount of resistance (in kilograms) applied by the examiner during the muscle test. The HHD was calibrated before testing by calculating the difference between a known load applied to a digital balance scale and the known load applied to the HHD. Absolute differences between the digital balance scale readout and the HHD readout for the known loads of 2, 5, 10, 15, 20, 25, and 30 kg were calculated. The resultant error was less than 0.1 kg.

Muscle testing was performed by first prepositioning the scapula in the midrange position of scapular motion for the specific muscle test. The midrange position was located by having the subject go through the available scapular range of motion, and then the midrange was estimated as the midpoint of the motion. This midrange position was selected to optimize the length-tension relationship of the tested muscle and, therefore, the generation of a maximum isometric contraction.\(^9\) A “make test” was performed for the muscle tests as described by Bohannon.\(^{19}\) The subjects were instructed to maintain the midrange position during each muscle test as resistance was gradually applied via the HHD until the examiner matched the subject’s effort.\(^9\)

The LT muscle test was performed as described by Hislop et al,\(^8\) with the resistance force from the HHD being applied to the spine of the scapula midway between the acromial process and the root of the spine, as depicted in Figure 2. The force on the scapula was applied in the superior and lateral direction parallel to the long axis of the humerus, which was at 140 degrees of elevation. The scapular motion for this test was scapular adduction and depression.

\(^*\) Therapeutics Unlimited Inc, 2835 Friendship St, Iowa City, IA 52245.

\(^†\) Lafayette Instruments, PO Box 57295700, Sagamore Parkway North, Lafayette, IN 47903.
The SA muscle test was performed as described by Kendall et al.9 The original test described resistance applied against the subject’s closed fist. For this study, the elbow was placed in 90 degrees of flexion, and resistance was applied to the ulna at the olecranon process along the long axis of the humerus. This application of resistance was modified for the SA muscle test because, during pilot testing, an HHD could not be applied in a stable or consistent manner over the subject’s hand, as depicted in the original description of the test by Kendall et al.9 Moreover, this modification decreased the number of joints that had to be crossed as force was applied to the scapulothoracic joint. The triceps muscle was monitored visually and by palpation to ensure that it did not contribute to force production during the SA muscle test. The scapular motion for this test was scapular protraction.

The MT muscle test was performed as described by Hislop et al.8 The HHD resistance force was applied to the spine of the scapula midway between the acromial process and the root of the spine, as depicted in Figure 4. The force was applied in the lateral direction parallel to the long axis of the humerus, which was placed in 90 degrees of abduction. Scapular retraction was the scapular motion for the MT muscle test.

The UT muscle test was performed as described by Hislop et al.8 The HHD was placed over the superior scapula. Force was applied directly downward (inferior) through the HHD in the direction of scapular depression. Scapular elevation was the scapular motion for the UT muscle test.

All subjects performed the muscle tests in the same order: LT, SA, MT, and UT muscle tests. The examiner did not view the digital readout on the HHD during performance of the muscle test. After the muscle test was completed, the examiner then viewed the kilograms and recorded the information on a data recording sheet. Each muscle test was performed 3 times consecutively, and the average was used for data analysis.

The same examiner performed all of the muscle tests on both days of testing. This physical therapist had approximately 1 year of experience with MMTs and the use of an HHD. Before the start of the study, reliability was assessed for a group of subjects who were healthy and free from shoulder dysfunctions (N=10). The muscle tests were performed in the standardized sequence (LT, SA, MT, and last UT). The amount of force applied (in kilograms) was measured with an HHD; 3 consecutive trials of each test were performed using a “make test”19 and averaged for data analysis. Excellent intrarater test-retest reliability (ICC[3,k]=.89–.95) was demonstrated for the LT, SA, MT, and UT muscle tests with an HHD.

**Data Analysis**

Descriptive data were calculated for all variables in the study. The average scores for day 1 and day 2 for each HHD scapular muscle test were used to calculate the between-day intratester test-retest reliability with a 2-way random-model ICC(2,k). Confidence intervals (CIs) were calculated at the 95% confidence level for the reliability coefficients. Confidence intervals or bounds allow for the expression of a level of certainty of point estimates. For example, if the correlation coefficient is .78 and the 95% CI is .70 to .85, then it can be said with 95% confidence that the true range of correlation coefficients is between .70 and .85.

Measurement error was calculated in 2 ways. Error associated with a 1-time HHD measure for each muscle test was calculated by use of the standard error of
measure (SEM) as follows: SEM = SD \times \sqrt{1-ICC} \].20 The SEM then was multiplied by the z score (1.64) for the 90% CI for the true score about the observed score (SEM90). The intrarater test-retest reliability coefficient used for determining the SEM was calculated using the 3 trials on day 1 with a 2-way random-model ICC(2,1). This within-day (intratrial) ICC was used for the calculation of the SEM because the SEM is a reflection of error with a measurement collected at a single point in time.

The error associated with multiple measures of each muscle test (ie, measures over the course of time) was calculated and is referred to as the “minimal detectable change” (MDC) or the smallest detectable difference. The MDC is the smallest amount of change that can be considered actual change that exceeds error in the measurement. The MDC was calculated for each of the 4 muscle tests individually with the following formula: 

$\text{MDC} = \text{SEM} \times \sqrt{1-ICC} \times \sqrt{2}$.21 The MDC then was multiplied by the z score (1.64) for the 90% CI for the true score about the observed score (MDC90). The between-day ICC using the average scores for day 1 and day 2 was used in the MDC calculation. This ICC was used because the MDC is an index of error across multiple measurement points over time. Only the upper limit of the 90% CI was reported for the error values because that value is the maximum amount of error in the measurement tool.

To assess the construct validity of data for the scapular muscle tests, the RMSs of the sEMG data for each muscle were compared across the 4 muscle tests. Four repeated-measures analyses of variance (ANOVAs) were performed to compare the average RMSs for an individual muscle across the 4 muscle tests. If a significant main effect was found, then post hoc testing of simple planned contrasts was used to determine whether the RMS for the muscle was significantly greater during the corresponding muscle test than during the other 3 muscle tests. The SPSS version 11.0.1 package2 was used for all data analyses.

Table 3.
Within-Day (Intratrial) Reliability and Error Estimates: Mean, Standard Deviation, and Minimum and Maximum Scores for Handheld Dynamometer Scapular Muscle Tests on First Day (Day 1) and Repeat Day (Day 2) of Testing (N=40)‡

<table>
<thead>
<tr>
<th>Test</th>
<th>ICC (95% CI)</th>
<th>X (SD) Minimum–Maximum (kg)</th>
<th>Error Estimates (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Day 1</td>
<td>Day 2</td>
</tr>
<tr>
<td>Lower trapezius muscle</td>
<td>.93 [.89–.96]</td>
<td>9.2 [3.4], 1.5–15.5</td>
<td>10.5 [4.0], 1.8–18</td>
</tr>
<tr>
<td>Serratus anterior muscle</td>
<td>.93 [.88–.96]</td>
<td>15.3 [6.3], 3.0–27.2</td>
<td>15.2 [6.0], 2.5–25.1</td>
</tr>
<tr>
<td>Middle trapezius muscle</td>
<td>.94 [.90–.97]</td>
<td>11.1 [3.2], 3.1–17</td>
<td>11.9 [3.1], 4.4–17.5</td>
</tr>
<tr>
<td>Upper trapezius muscle</td>
<td>.95 [.92–.97]</td>
<td>16.1 [7.1], 2.4–29.2</td>
<td>17.2 [7.1], 4.7–28.1</td>
</tr>
</tbody>
</table>

* ICC=intraclass correlation coefficient (2-way random model). CI=confidence interval, SEM=standard error of measure, SEM90=90% CI for standard error of measure.

The results are summarized in Table 3 for both days of testing. For test-retest reliability across days of testing, the ICC(2,k) for the intrarater reliability of isometric force data for scapular muscle testing ranged from .89 to .96 (Tab. 4). The single-measure error estimate (SEM90) ranged from 1.8 to 3.6 kg (Tab. 3). The error estimate associated with multiple measures (MDC90) ranged from 1.3 to 2.7 kg (Tab. 3). The error estimate associated with multiple measures (MDC90) ranged from 1.8 to 3.6 kg (Tab. 4).

Results

Means and standard deviations for HHD measures are summarized in Table 3 for both days of testing. For test-retest reliability across days of testing, the ICC(2,k) for the intrarater reliability of isometric force data for scapular muscle testing ranged from .89 to .96 (Tab. 4). The single-measure error estimate (SEM90) ranged from 1.3 to 2.7 kg (Tab. 3). The error estimate associated with multiple measures (MDC90) ranged from 1.8 to 3.6 kg (Tab. 4).

Means and standard deviations for the RMSs of the middle 2 seconds of the sEMG data are reported in Table 5. A repeated-measures ANOVA revealed a significant main effect for the LT muscle RMS across all 4 MMTs (F=17.26; df=3.37; P<.001; power=.10). Post hoc testing revealed that the LT muscle RMS was statistically significantly greater during the LT muscle test than during the other muscle tests. A significant main effect for the UT muscle RMS also was revealed by a repeated-measures ANOVA (F=24.48; df=3.37; P<.001; power=.10). Post hoc testing indicated that the UT muscle RMS was statistically significantly greater during the UT muscle test than during the other 3 muscle tests. A repeated-measures ANOVA for the MT muscle RMS revealed a significant main effect (F=23.26; df=3.37; P<.001; power=.01], and the MT muscle RMS was shown to be statistically significantly greater during the LT muscle test than during the other 3 muscle tests. A significant main effect (F=6.21; df=3.37; P=.002; power=.10) was found for the SA muscle test. Post hoc testing
revealed that the SA muscle RMS was statistically significantly greater during the SA muscle test than during the LT and UT muscle tests but not during the MT muscle test ($P=0.48$, power $=0.11$). We were unable to perform an a priori power analysis to determine the sample size because sEMG data from isometric force testing in subjects with shoulder dysfunctions were not available.

**Discussion**

The accuracy of a clinical examination is dependent upon the quality of the tests and measures used during the examination. Our study demonstrated excellent intrarater test-retest reliability and established error values for the measurement of isometric force with an HHD during 4 scapular muscle performance tests in subjects with shoulder pain and functional loss. Construct validity results indicated that the LT and UT muscles have the highest muscle activity during their respective muscle tests. Moreover, this study provided evidence to indicate that the SA and MT muscles do not yield the highest muscle activity during their respective muscle tests as performed in this study. Before this study, only the intrarater test-retest reliability of the MT muscle test in subjects without shoulder dysfunction had been assessed.

Reliable measurements of isometric muscle force with an HHD can be obtained for scapular muscle tests of the LT, MT, UT, and SA muscles in people with a variety of shoulder disorders. The intrarater test-retest reliability coefficients ranged from .89 to .96, suggesting excellent reliability. Earlier investigations of intrarater test-retest reliability indicated the same conclusion for the MT muscle test, but in subjects without shoulder dysfunctions only. During a single testing session with an HHD and the MT muscle test described by Hislop et al., an excellent intratester reliability was reported by DiVeta et al.10 (ICC[1,1] = .96) and Zmierski et al.11 (ICC[1,1] = .96–.98). Because these earlier studies were performed with subjects who were free from shoulder pathology, the results have limited generalizability to clinical practice.

The SEM$_{90}$ ranged from 1.3 to 2.7 kg for the 4 muscle tests. The SEM can be used to make decisions regarding a single scapular muscle test result for a given patient. For example, in a given subject, the LT muscle test yielded a score of 9.5 kg. The SEM$_{90}$ for the LT muscle test was 1.5 kg, indicating that the true score for this subject would be ±1.5 kg about the observed score of 9.5 kg. At present, the error value of a single-time HHD scapular muscle measurement is unclear without the establishment of normative values of scapular muscle measurements. Future research also is needed to determine the usefulness of these single measurements to predict a future event, such as treatment outcome and
the capacity to discriminate between patients who need scapular muscle strengthening and those who do not.

The SEM was calculated using the within-day (intertrial) ICC from day 1 of testing (Tab. 3). Data from the 3 trials were collected consecutively, without having the subject get off the table or reposition each time between trials. This procedure may have overestimated the reliability and, therefore, underestimated the SEM for the HHD with scapular muscle testing.

The MDC$_{90}$ ranged from 1.8 to 3.6 kg for the 4 muscle tests. The MDC indicates the amount of change that is necessary for a result to be considered greater than measurement error. The MDC can be used to make decisions about muscle test results performed consecutively over time in individual patients. For the subject described above, a repeat LT muscle test performed 3 weeks later yielded a score of 11.5 kg. Therefore, the subject’s LT muscle test score increased by 2.0 kg. Because the MDC$_{90}$ for the LT muscle test was 2.6 kg, we could conclude that the increase of 2.0 kg was a change that was within the range of variability of 90% of stable patients. That is, this subject’s change in LT muscle test results did not demonstrate a change greater than the random error (MDC$_{90}$=2.6 kg) associated with repeat testing. The MDC is a more useful error estimate than the SEM. The MDC can be used to determine whether change in an HHD muscle test score is likely due to error or is true change. However, the MDC values from this study should be applied to HHD measurements taken over 1 to 3 days only because the MDC calculation was based on the between-day reliability coefficients. Greater measurement error with poorer reliability is possible with repeat measurements over longer periods of time.

The MDC indicates the amount of change that is statistically meaningful on the basis of calculations using the test-retest reliability statistic. Is this amount of change (MDC) also the amount of change that is important to the patient? That is, does it relate to an improvement in the use of the patient’s shoulder? This study and previous work have not provided the amount of change in HHD measurements of scapular muscle performance that is considered meaningful to the patient. The amount of change in a measurement that is considered clinically meaningful to the patient has been labeled the “minimal clinically important difference” (MCID). The design of this study did not allow for the calculation of the MCID. In order to determine an MCID, HHD measurements would have to be taken before and after a therapeutic intervention along with a measurement of patient-estimated improvement after the intervention. Future studies should determine the MCID for scapular muscle tests, because this value is potentially the more appropriate one for determining clinically meaningful change in a patient’s status.

Construct validity was demonstrated for the UT and LT muscle tests with an HHD. The assessment of construct validity was carried out by measuring the isometric muscle activity of all 4 muscles during the 4 muscle tests with sEMG as the construct measure of comparison. Specifically, the muscle activity of an individual muscle during its designated muscle test was compared with its muscle activity during the other 3 muscle tests. Higher muscle activity indicated increased electrical activity under the electrode overlying that muscle and thus potentially greater muscle recruitment. The LT muscle test yielded the highest LT muscle activity compared with the other 3 muscle tests. Similarly, the highest muscle activity for the UT muscle occurred during the UT muscle test. The UT and LT muscle tests appear to be the most favorable tests for generating the highest muscle activity for their respective muscles in the 4 muscle tests investigated. Our results are consistent with those of a study that investigated the optimal normalization procedure for the trapezius muscle in individuals without shoulder dysfunction; the UT and LT muscles generated the highest muscle activity during their respective tests as performed in this study compared with the other muscle tests performed. Direct comparisons cannot be made to our study because of the methodological differences between a “break test” and a “make test,” the application of resistance via the examiner’s hand, and the application of resistance to the distal humerus for the LT muscle test.

Construct validity results for the MT muscle test indicated that MT muscle activity was not highest during this test. Rather, the MT muscle demonstrated the highest RMS during the LT muscle test. The MT muscle RMS (0.025 mV) during the MT muscle test was significantly lower than the MT muscle RMS (0.034 mV) during the LT muscle test. The MT muscle test performed in this study does not appear to be the optimal test for the MT muscle. Potential explanations for the higher muscle activity of the MT muscle during the LT muscle test may involve similarities of placement of the HHD and the direction of resistance applied. Moreover, fatigue may have been a factor, because the LT muscle test always was performed before the MT muscle test. Our findings agree with those of the previously described normalization study performed on subjects without shoulder dysfunction; the muscle activity of the MT muscle was higher during the LT muscle test than during the MT muscle test as performed in our study. In the normalization study by Ekstrom et al., the highest muscle activity of the MT muscle was demonstrated during the LT muscle test and during the MT muscle test performed as described by Kendall et al. (patient positioned supine
with shoulder horizontally abducted to 90° and with maximum external rotation). Future work should investigate other muscle tests to determine the optimal isometric muscle test for the MT muscle in people with shoulder dysfunctions.

Construct validity was not demonstrated for the SA muscle during the SA muscle test performed in this study. Data analysis revealed a main effect; however, post hoc testing revealed no significant difference between the muscle activity of the SA muscle during the SA muscle test and the muscle activity of the SA muscle during the MT muscle test \((P=.48)\). Methodological issues may explain this finding. The upper portion of the SA muscle was not monitored with our sEMG electrodes because we placed electrodes over the lower portion of the SA muscle only. The primary action of the lower portion of the SA muscle is upward rotation, in contrast to the action of scapular protraction elicited during the SA muscle test performed in this study. Higher muscle activity of the lower portion of the SA muscle may be elicited during another described SA muscle test,\(^8,9\) in which resistance to shoulder flexion is applied with the arm elevated to 125 degrees.\(^24\) Future investigations should explore other previously described SA muscle tests\(^8,9\) while monitoring both the upper and the lower portions of the SA muscle.

Another possible explanation for the lower muscle activity of the SA muscle during the SA muscle test is electrode placement. The method used for SA muscle electrode placement may not have been the most accurate because of the difficulty in differentiating SA muscle activity from the activity of the latissimus dorsi muscle. In an alternate method, the latissimus dorsi muscle is palpated along the lateral chest wall at the level of the inferior scapular angle, and then the electrode is placed anterior to this location.\(^18\) We were unaware of this alternate technique before the initiation of this study.

The use of sEMG as a construct measure in this study has limitations. Measuring muscle activity on the skin surface allows for recording concurrent activity from any other muscles in the area surrounding the electrode. Thus, a plausible explanation for the electrical activity detected under an electrode may involve the recruitment of the underlying muscle and surrounding muscles. We used established protocols for electrode placement to optimize individual muscle activity; however, it is not clear whether the muscle activity of other, surrounding muscles was measured as well.

In this study, we monitored only the muscle activity of the 3 sections of the trapezius muscle and the lower portion of the SA muscle. Thus, it is unknown whether other, unmonitored muscles also may have been highly active during these tests. Future investigations of muscle tests should include surface or indwelling electromyographic measurements of all muscles that may be active during a muscle test.

Construct validity was assessed with the use of sEMG as the comparison measure of muscle activity. Criterion validity was not assessed in this study because there was no direct measure of muscle activity. A direct measure of muscle activity would involve measurement of muscle activity in vivo. Future studies should use a direct measure, such as an electromyographic needle electrode placed directly into the muscle, to determine the criterion validity of measurements of muscle activity during muscle testing.

Manual muscle tests were developed with the theory that the muscle targeted by an individual muscle test would produce the maximum muscle activity.\(^8,9\) The sEMG measure detects the electrical activity of a muscle under the electrode, not the production of muscle force. Because sEMG muscle activity is not equal to muscle force production, this study cannot provide conclusive evidence to support or refute the theoretical postulate of manual muscle testing for the scapular MMTs investigated.

Subjects in this study were subjects with shoulder pain and functional loss. All subjects were able to complete all 4 muscle tests; however, the effect of pain on their ability to generate muscle activity is unknown. Pain levels were not recorded during the testing procedures. Before the start of testing on day 2, subjects were asked whether they had “gotten better, gotten worse, or stayed the same” since day 1. This question does monitor their pain and functional loss in a global sense, but not their muscle performance stability.

The 4 muscle tests were performed in a standardized order. The test order was not randomized to ensure that the MT and LT muscle tests were not performed in succession. This was done to reduce the possibility of fatigue, because the actions performed during these muscle tests were very similar. However, the lack of randomization may have contributed to fatigue in the muscle tests performed third and fourth. Therefore, our efforts to reduce fatigue actually may have led to fatigue during the MT and UT muscle tests.

The generalizability of the results is limited. Only 1 examiner performed the tests in a single clinical setting. Additionally, the diagnoses of the subjects in this study were diverse but were weighted toward subacromial impingement syndrome. The small number of subjects also reduces the variety of diagnoses and limits the generalizability. Future research should assess these mus-
cule tests with an HHD and with a variety of diagnoses and examiners.

The collection of any type of data during an examination or reexamination of a patient should be performed for use in differential diagnosis, guiding of treatment interventions, or assessment of the outcome of interventions. The usefulness of scapular muscle testing with an HHD for outcome assessment was established in this study. Future investigations of the predictive and prescriptive validity of data for scapular muscle tests with an HHD are necessary to address these issues.

**Conclusion**

The delivery of health care in an evidence-based medicine environment requires the use of tools that possess acceptable measurement properties. This study has contributed to the literature by establishing the intrarater test-retest reliability and error values over 1 to 3 days for the LT, SA, MT, and UT muscle tests with an HHD in subjects with shoulder pain and functional loss in an outpatient setting. The LT, SA, MT, and UT muscle tests can be reliably performed with an HHD. Individual subject decision making based on HHD results from scapular muscle tests can be improved by use of the presented MDC error values for patients with a variety of shoulder diagnoses. However, only the UT and LT scapular muscle tests demonstrated that the UT and LT muscles were maximally challenged by the respective MMTs. Construct validity was not established for the SA and MT muscle tests, because the designated muscle did not have the highest muscle activity of the monitored muscles during these tests. Therefore, on the basis of the results of this study, the SA and MT muscle tests as performed in this study are not recommended to produce the highest muscle activity for the SA and MT muscles.

**References**


