American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: Reliability, validity, and responsiveness

Lori A. Michener, PhD, PT, ATC,[a] Philip W. McClure, PhD, PT,[b] and Brian J. Sennett, MD,[c] *Richmond, Va, and Glenside and Philadelphia, Pa*

The purpose of this study was to examine the psychometric properties of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES), patient self-report section. Patients with shoulder dysfunction (n = 63) completed the ASES, The University of Pennsylvania Shoulder Score, and the Short Form–36 during the initial evaluation, 24 to 72 hours after the initial visit, and after 3 to 4 weeks of physical therapy. The test-retest reliability (intraclass correlation coefficient [1-way random-effects], 0.84; 95% CI lower limit, 0.75) and internal consistency (Cronbach α, 0.86) values were acceptable. The standard error of the measure was 6.7 ASES points (90% CI, 11.0). Construct and discriminant validity was demonstrated. Responsiveness was demonstrated with a standardized response mean of 1.5 and an effect size of 1.4. The minimal detectable change was 9.7 ASES points (90% CI, 16), and the minimal clinically important difference was 6.4 ASES points. The results indicate that the ASES is a reliable, valid, and responsive outcome tool. (J Shoulder Elbow Surg 2002;11: 587-94.)

The measurement of functional limitations and disability has become an important issue in health care. To address this need, self-report disease or condition-specific questionnaires have been developed. These scales are designed to assess a specific disease or region of the body. The American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES), published in 1994 by the Research Committee of the American Shoulder and Elbow Surgeons, contains a patient self-report section and a section used by medical professionals to record physical examination findings to assess patients with shoulder pathologies.[21]

The original publication of the ASES did not contain an assessment of the measurement properties. Measurement properties of a modified version of the patient self-report section of the ASES have been established[1,3]; however, these results cannot be extrapolated or directly used to interpret the results of the ASES. Because the ASES was modified with the addition of 5 questions, there is no assurance that the measurement properties of the modified version are similar to those of the original ASES. The measurement properties of the ASES need to be examined because this is the form commonly used by researchers and clinicians alike to assess patient outcomes.

In order for any type of measurement tool to be used with confidence when making decisions regarding patient care, the psychometric or measurement properties of reliability, validity, and responsiveness must be established. Moreover, these values must be calculated with a sample of patients who are representative of those for whom the measurement tool will be used, or the tool cannot be used with confidence.

A self-report measurement tool should demonstrate reliability by being both repeatable—yielding consistent results in a group of clinically stable patients—and internally consistent—demonstrating that the items on the questionnaire are strongly related to each other and thus measure a single characteristic.[7,19] In conjunction with the reliability indices, error values must be calculated to determine the error associated with a questionnaire's score, enabling the clinician or researcher to interpret scores with a level of confidence. The self-report tool must also demonstrate responsiveness—the ability of the scale to detect clinical change accurately when it has occurred.[2,7] Finally, the validity of a function or disability scale should be assessed to determine

From the Department of Physical Therapy, Virginia Commonwealth University–Medical College of Virginia, Richmond[a], Va; the Department of Physical Therapy, Arcadia University, Glenside, Pa[b]; and the Department of Orthopedic Surgery, University of Pennsylvania Health Systems, Philadelphia, Pa.[c]

Partially supported by a grant from the Foundation for Physical Therapy and the National Athletic Trainers Association Research and Education Foundation.

Reprint requests: Lori A. Michener, PhD, PT, ATC, Assistant Professor, Department of Physical Therapy, Virginia Commonwealth University, Medical College of Virginia Campus, PO Box 980224; Richmond, VA 23298 (E-mail: *lmichene@hsc.vcu.edu*).

1058-2746/2002/$35.00 + 0 **32/1/127096**
doi:10.1067/mse.2002.127096

Table I Patient diagnoses

Diagnosis	No.
Impingement syndrome	25
Instability/dislocation	2
Rotator cuff syndrome	2
Adhesive capsulitis	5
Hemiarthroplasty	1
Shoulder weakness	2
Humeral fracture	5
Rotator cuff and adhesive capsulitis	6
Status—post surgery	15

Table II Patient characteristics (n = 63)

Characteristic	Summary
Age (y)	
Mean ± SD	51.7 ± 15.5
Range	20-81
Height (cm)	
Mean ± SD	167.6 ± 9.9
Weight (kg)	
Mean ± SD	78.9±20.6
Sex	
Male	26
Female	37
Dominant shoulder	
Right	49
Left	11
Ambidextrous	3
Time since onset	
<1 mo	9
1-3 mo	22
3-6 mo	6
>6 mo	26
Work status	
Not working	20
Working	43
Involved in litigation	
Yes	3
No	60

whether the scale measures what it intends to measure. This is a complex entity to assess.[7,19]

The purpose of this study was to examine the psychometric properties of the patient self-report section of the ASES: test-retest reliability, internal consistency, responsiveness, and validity. By demonstrating acceptable properties, this scale can be used to make confident judgments regarding the status of an individual patient and groups of patients.

MATERIALS AND METHODS

The patient self-report section of the ASES is a condition-specific scale—that is, it is designed for a specific condition, which is intended to measure functional limitations and pain of the shoulder. The original ASES[21] consists of 2 portions, a medical professional assessment section and a patient self-report section. The patient self-report section, evaluated in this study, is a patient self-evaluation questionnaire that takes approximately 5 minutes to complete and consists of 2 dimensions: pain and activities of daily living (Appendix Figure 1). The pain score is calculated from the single pain question and the function score from the sum of the 10 questions addressing function. The pain score and function composite score are weighted equally (50 points each) and combined for a total score out of a possible 100 points.[21]

The University of Pennsylvania Shoulder Score (Penn)[14] is a 100-point self-report scale consisting of 3 sections: pain, satisfaction, and function. There are 3 questions regarding pain (30 points), 1 regarding satisfaction of shoulder function (10 points), and 20 regarding function (60 points). The pain and satisfaction scores are derived from 10-point numeric scales. For function, a 4-category Likert scale of level of difficulty (0, cannot do at all; 1, much difficulty; 2, some difficulty; 3, no difficulty) is used for each function question. Measurement properties have been established for the Penn scale in patients with various shoulder pathologies.[14]

The Short Form–36 (SF-36) is a generic measure of overall health-related quality-of-life status, with 8 underlying constructs and 2 component summary scores.[24] Previous research has demonstrated that the SF-36 is a reliable, valid, and responsive generic measure of overall quality of life in patients with musculoskeletal disorders.[17,24]

Sixty-three patients with various shoulder pathologies were recruited for participation at a variety of outpatient clinics, upon referral for physical therapy services for both operative and nonoperative diagnoses (Table I). Patients were included if the physical therapist's diagnosis was that of a shoulder dysfunction. Patient demographics are shown in Table II. Patients were excluded if they were unable to complete the form because they had cognitive impairment, were illiterate, or could not understand English. In addition, patients were excluded if it was determined by the physical therapist that the patient's functional loss and pain were primarily a result of a comorbidity or a diagnosis not related to the shoulder.

Twenty-two physical therapists (mean age, 33 years) from 11 physical therapy clinics participated in this study. The clinicians (9 women and 13 men) had a mean of 9.6 years of practice and reported that they spent from 15% to 50% (mean, 29%) of their day treating patients with shoulder pathologies. Each of the clinicians examined 1 to 8 subjects each, representing a 1.6% to 12.7% contribution each.

The clinics were given subject data packets, and the physical therapists were given specific written and verbal instructions for data collection. The specific data collection protocol is depicted in Figure 1. The global rating of function question (GRF-therapist), completed at the first collection of data (C1) at the initial evaluation, was as follows: "On a scale of 0 to 100, the use of their shoulder is ___, with 0 = no use and 100 = full use". The improvement question, completed at the second visit 24 to 72 hours after the initial evaluation (C2) and was administered to determine whether the patient had improved since the initial evaluation, was as follows: "Since your first visit to physical therapy, do you feel that your shoulder has: gotten better, stayed the same, gotten worse?" The global rating of change question (GRC), completed at the third data collection (C3) 3 to 4 weeks after the initial evaluation and physical therapy treatment, was as follows: "Since your first

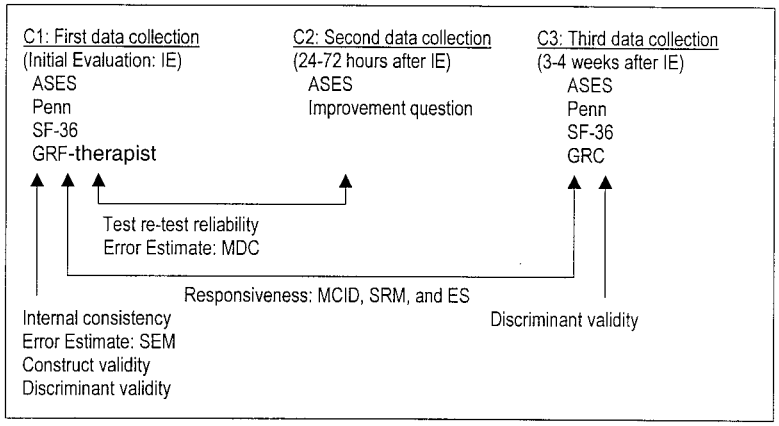

Figure 1 Data collection and data analysis of the patient self-report section of the ASES.

visit to physical therapy, do you feel that your shoulder has: gotten much better, gotten slightly better, stayed the same, gotten slightly worse, or gotten much worse?" Data analysis of the ASES patient self-report section is depicted in Figure 1.

Reliability

Test-retest reliability was examined with the use of a 1-way random-effects intraclass correlation coefficient (ICC) to determine the relationship between the scores from C1 to those from C2. A 1-way random-effects model was chosen because the only systematic source of variance was the subjects.[16] Only data for those patients who stated on C2 that their shoulder did not change since C1 were used to examine reliability. Homogeneity, or internal consistency of the scale, was determined by calculating a Cronbach α coefficient from the scores from C1.[19]

Error estimates

The error associated with the completion of the scale on a single occasion was determined by calculating the standard error of the measure (SEM). The SEM was calculated with the use of the Cronbach α coefficient, as this is a measure of the stability of the scale on a single occasion.[19] The formula for the SEM is as follows: SEM = SD $\times$ [square root of (1 $-$ Cronbach α)]. The SEM carries with it only a 68% CI; therefore, the SEM was multiplied by the *z* value associated with the 90% CI (1.65) to obtain the 90% CI for the error of a single application of the ASES, with SEM used as the estimated error.

The error associated with multiple applications of the patient self-report section of the ASES was determined by calculating the minimal detectable change (MDC), also known as the smallest real difference (SRD).[2] It was calculated with the following formula: MDC or SRD = (SD $\times$ [square root of (1 $-$ ICC value)]) $\times$ square root of 2.[2] Multiplication by the square root of 2 was performed to estimate the error in completing the questionnaire twice. The 90% CI was calculated for the MDC, which is the statistically minimal amount of change required in an ASES score to be 90% confident that "true change" has occurred.

Validity

Construct validity was determined by assessing both convergent and divergent validity. Convergent validity—correlation between similar scales or dimensions of a scale—was assessed with Pearson product moment correlation coefficients to examine the relationship between the C1 scores of the patient self-report section of the ASES and those of the Penn scale, the SF-36 physical function score, the SF-36 role physical score, and the SF-36 physical component summary score. Divergent validity—lack of correlation between dissimilar scales or dimensions of a scale—was assessed with the use of Pearson product moment coefficients to examine the relationship between the C1 scores of the patient self-report section of the ASES and the SF-36 mental health score, the SF-36 role emotional score, and the SF-36 mental component summary score. The α level was set at .05.

Discriminant validity was assessed by comparing the final patient self-report ASES scores (C3) of those patients who stated on the GRC questionnaire that they had "gotten slightly better" with the scores of those who stated that they had "gotten much better." Analysis of covariance (ANCOVA), with C1 ASES scores as the covariate, was used to make this comparison of C3 scores. In addition, discriminant validity was assessed by comparing C1 ASES scores between patients of varying levels of the therapist's rating of global rating of function (GRF-therapist). To do so, the global rating values assigned by physical therapists were recoded to 1 of 3 categories: patients who were globally rated by their therapist to have a rating of 61 to 90 out of 100 were considered to be minimally functionally limited, those with a rating of 31 to 60 out of 100 were moderately functionally limited, and those with a global rating of 0 to 30 out of 100 were maximally functionally limited. No patient was rated above 90 global rating points by their physical therapist at C1. Analysis of variance (ANOVA) was used to compare the C1 ASES scores between these ratings of minimally functionally limited, moderately functionally limited, or maximally limited. For both ANCOVA and ANOVA, the α level was set at .05.

Table III Reliability and error estimates, means, and standard deviations for ASES at C1 (initial evaluation) and C2 (24 to 72 hours after initial evaluation) (n = 63)

	Mean ± SD		Reliability	Error estimates	
ASES score	**C1**	**C2**	**ICC (95% CI)***	**SEM (90% CI)**	**MDC (90% CI)**
Pain (50 points possible)	29.6 ± 11.3	31.0 ± 11.2	0.79 (0.69-0.88)	5.1 ± 8.4	7.2 ± 11.8
Function (50 points possible)	22.2 ± 9.2	24.6 ± 9.7	0.82 (0.72-0.89)	4.1 ± 6.7	5.8 ± 9.5
Total (100 points possible)	51.8 ± 16.9	55.6 ± 17.3	0.84 (0.75-0.91)	6.7 ± 11.0	9.4 ± 15.5

*One-way random-effects.

Responsiveness

The first step in determining responsiveness was to distinguish between those patients who had improved and those who had not, by identifying an external criterion indicative of clinically meaningfully change.[2] The external criterion of improvement chosen was the GRC. The use of the GRC as an indicator of change has been questioned.[18] The measurement properties of the GRC have not been investigated, and therefore it is uncertain if this measure can be collected consistently and accurately. Both functional limitation scales and the GRC are measures that involve a patient's judgment; therefore, these scores and their errors are most likely correlated. This potentially makes the GRC a biased measure of change.[18] An unbiased impairment measure that has been demonstrated to be reliable and valid may be a better external criterion. Recently, the relationship between an impairment index (aggregate of impairments) and the GRC has been investigated.[5] In patients with low back pain, it was demonstrated that patients identified by the GRC as "stable" showed little change on a physical impairment index, whereas those identified by the GRC as "improved" demonstrated a reduced impairment index.[5] This finding provides evidence for the use of the GRC as a determinant of clinically meaningful change.

There are various aspects of responsiveness that need to be examined. In this study, the GRC was used, as the external criterion for determination of those patients who improved, remained stable, and deteriorated was determined at C3. For making decisions regarding group changes, the standardized response mean (SRM)[15] and effect size (ES)[12] statistics were calculated. The SRM was calculated as: SRM = (C3-C1)/SD of change scores.[15] The ES is calculated as: ES = (C3-C1)/SD of the C1 scores.[12] A higher SRM or ES indicates greater responsiveness.[4,8] The responsiveness indices are interpreted based upon their magnitude. According to Cohen, an ES of 0.2 represents a small effect, 0.5 a moderate effect, and 0.8 or greater a large effect.[26]

The MDC, or SRD, is that change which is statistically meaningful but not necessarily clinically meaningful to the patient. Change that is clinically meaningful to the patient is called the minimal clinically important difference (MCID).[10] To determine the amount of change that differentiates those patients who have improved from those who have remained stable or deteriorated, receiver operating characteristics (ROC) curves were plotted.[4] The ROC curve was constructed by plotting sensitivity versus 1 − specificity for the patient self-report section of the ASES change scores (C3 − C1), as the cutoff point varied for differentiation between those patients who stated on the GRC that they had improved and those who remained stable or had deteriorated. The area under the curve was evaluated for significance, with the α level was set at .05.

RESULTS

Reliability

Test-retest reliability was calculated with the use of 1-way ICCs, the results of which are shown in Table III. The internal consistency estimate of Cronbach α was .86.[19]

Error estimates

The error associated with a single application (SEM) and multiple applications (MDC/SRD) of the patient self-report section of the ASES was calculated; the results are shown in Table III.

Validity

The correlations (95% CI) for the assessment of convergent validity were significant between the patient self-report section of the ASES and the Penn scores ($r = 0.78$; $P < .01$; CI, 0.86-0.66), SF-36 physical function score ($r = 0.41$; $P = .001$; CI, 0.18-0.69), SF-36 role physical score ($r = 0.33$; $P = .008$; CI, 0.09-0.53), and SF-36 physical component summary score ($r = 0.40$; $P = .001$; CI, 0.17-0.59). The correlations for the assessment of divergent validity were not significant between the ASES and the SF-36 role emotional score ($r = 0.24$; $P = .21$; CI, −0.01 to 0.46), SF-36 mental health score ($r = 0.05$; $P = .70$; CI, −0.20 to 0.29), or SF-36 mental component summary score ($r = 0.15$; $P = .25$; CI, −0.10 to 0.38).

Discriminant validity of the patient self-report section of the ASES was demonstrated with significantly higher C3 ASES scores for the patients who stated that they had "gotten much better" (mean, 80.4 ± 13.0 ASES points) versus those who had "gotten slightly better" (mean, 67.0 ± 18.6 ASES points) (ANCOVA: $F[1,54] = 21.5$; $P < .001$).

Discriminant validity of the patient self-report section of the ASES at the initial evaluation was demonstrated with significantly different ASES scores at C1 between groups who were rated by a therapist as

Table IV Responsiveness indices, means, and standard deviations for ASES at C1 (initial evaluation) and C3 (3 to 4 weeks after initial evaluation) (n = 55)

ASES score	Mean (SD)		Change scores (mean ± SD)	SRM	ES
	C1	C3			
Pain (50 points possible)	30.1 ± 11.0	42.0 ± 9.2	11.8 ± 10.9	1.08	1.07
Function (50 points possible)	22.0 ± 9.6	33.9 ± 9.7	11.9 ± 8.9	1.34	1.24
Total (100 points possible)	52.1 ± 17.1	75.8 ± 16.3	23.7 ± 15.4	1.54	1.39

minimally, moderately, or maximally functionally limited (ANOVA: F[2,60] = 23.50; $P < .001$). Scheffé post hoc testing revealed that the minimally functionally limited group (mean, 66.1 ± 11.6 ASES points) had significantly higher ASES scores than the maximally functionally limited group (mean, 40.7 ± 15.1 ASES points) and the moderately functionally limited group (mean, 45.0 ± 12.5 ASES points).

Responsiveness

Of the original 63 patients, 59 (94%) completed both the initial (C1) and final (C3) data collection. Of those 59 patients, 4 reported on the GRC question that they had gotten worse or stayed the same since the initial data collection (C1). Thus, 55 patients (87%) remained for calculation of the responsiveness indices of SRM and ES (Table IV). All 59 who completed forms at C1 and C3 were used to calculate the MCID. An ROC curve for the change scores was constructed to determine the MCID (Figure 2). The area under the curve was 0.818 (SE, 0.106), which was significant ($P = .035$). The cutoff point, which was the point nearest the upper left-hand corner of the graph, was 6.4 ASES points; the sensitivity was 91% and the specificity 75%.

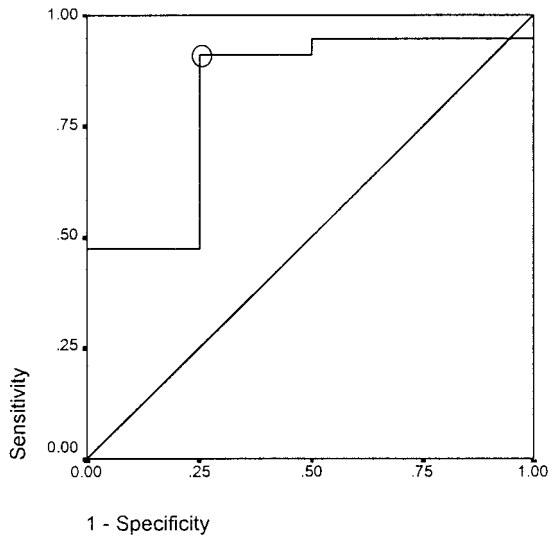

Figure 2 ROC curve for the ASES. It is constructed with data points of sensitivity and 1 − specificity values, as the cutoff point varied for differentiation between those patients who stated that they had improved and those who did not. The point closest to the far left corner (circle) is the cutoff point of 6.4 ASES points, with a sensitivity of 91% and a specificity of 75%. The greater the area under the curve, the greater a measure's ability to distinguish between patients who underwent a meaningful change and those who did not.

DISCUSSION

In the present health care climate, there is an ever-increasing demand to demonstrate the efficacy of therapeutic interventions. Because of this need self-report instruments have been developed to assess the impact of a disease on a patient's functional limitations and disability. However, in order for an outcome measure to be interpreted and used, these instruments must have documented psychometric properties.

To assess reliability, the patient self-report section of the ASES was administered 1 to 3 days apart in order to reduce the possibility of clinical change in the study population. The test-retest reliability of the patient self-report section of the ASES is considered to be excellent, with an ICC of 0.84.[20] Other shoulder self-report forms have demonstrated a higher level of test-retest reliability (range, 0.91-0.99),[1,13,14] whereas others have demonstrated a lower level (range, 0.64-0.71).[3,22] The self-report section of the ASES test-retest reliability value is above the cutoff value for acceptable, at 0.75.[20] The internal consistency value of 0.86 indicates that the self-report section of the ASES possesses an acceptable level of homogeneity. Theoretically, if this value is too large (≥0.95), the scale may contain too many items. This was not indicated with the Cronbach α value for the ASES.

Self-report condition-specific outcome measures aimed at the assessment of functional limitations should address areas of principal concern to the patient regarding his or her functional status. These include aspects of activities of daily living, as well as work and recreational activities. The patient self-report section of the ASES evaluates these aspects, which provides evidence for the face validity of this scale.

Theoretically, a discriminative instrument should be able to discriminate between patients of various functional levels.[7,19] It was demonstrated that the self-report section of the ASES was able to discriminate between different levels of patient-rated improvement and therapist-rated level of function.

Convergent construct validity for the self-report section of the ASES was supported by the moderate to strong significant relationships with both another shoulder condition-specific measure and a generic measure. In previous studies, as well as in this study, the correlations between shoulder questionnaires were strong.[3,25] Conversely, the relationship between the self-report section of the ASES and the generic measure of the SF-36 was only moderate to weak, which was also reported in previous studies of other shoulder self-report instruments.[3,9,25] The lower correlations with the generic measure can most likely be explained by the fact that only 4 of 10 questions regarding physical function on the SF-36 pertain to the upper extremity. These 10 questions are used separately and in combination with other items to calculate the physical function, role physical, and physical component summary scores.

Another aspect of construct validity, divergent validity, was established with the nonsignificant correlations of the self-report section of the ASES with the SF-36 mental health, role emotional, and mental health component scores. However, even though the correlations were not significant, there was overlap between the mental and emotional health variables and the ASES. This suggests that the self-report section of the ASES is basically measuring a different construct other than that of mental and emotional health. This also provides further evidence to suggest that a condition-specific measure should be done in conjunction with a generic health status measure, because they are indeed measuring different constructs.[6,9,11]

For making decisions regarding groups of patients, the patient self-report section of the ASES is responsive, with both the SRM and ES greater than 1.0.[20,23] These values are both considered to be large,[4,20,23] thus indicating good responsiveness. The SRMs of other shoulder condition–specific measures from previous studies ranged from 0.65 to 1.38,[1,9,13] in contrast to the larger SRM calculated for the total self-report ASES scores in this study (1.54). These results appear to indicate that the patient self-report section of the ASES is more responsive than other shoulder questionnaires. However, this can only be inferred until a head-to-head comparison is made of all shoulder condition–specific measures on a variety of patient diagnoses in a variety of settings.

For making decisions regarding individual patients, the error was calculated in various ways. The error (SEM) in the patient self-report section of the ASES with a single application was 11 ASES points, and with multiple applications, it was 16 ASES points, expressed with a 90% CI. This appears to be a logical finding, as the error would be expected to be greater with a measurement tool used twice as opposed to only once. The clinically important difference was calculated to be 6.4 ASES points, with the use of ROC curve analysis. The MCID is less than the MDC, which has been demonstrated with other self-report scores as well,[5] indicating that the amount of change a patient perceives as important change can be less than the amount deemed to be statistically significant change.

The SEM, MDC, and MCID values can be applied in the determination of patient status and change in a patient's functional limitations. For example, if a patient has an initial self-report ASES score of 40 points, the clinician can be 90% confident that the actual score falls within ± 11 points of that 40-point score. To determine whether change in the ASES score is meaningful, the MDC and MCID values can be used. For example, if that same patient who scored 40 points on the initial evaluation scores 62 points during reassessment 4 weeks later, the clinician will be able to state confidently that the patient has demonstrated statistically meaningful improvement because the change of 22 points is greater than the MDC value (90% CI, 16 points). However, consider the same patient who, on reassessment at only 2 weeks, had a change of only 10 points on the self-report section of the ASES. This change is interpreted as a nonstatistically meaningful change but is interpreted as a clinically meaningful change because it is greater than the MCID value of 6.4 ASES points. The MCID value may be the more important value to consider when evaluating change in a patient's status because it is the value that is associated with the patient's perception of meaningful change.

There are limitations with the error and responsiveness indices. First, there are factors that were not controlled for and, therefore, may have affected the ASES self-report section change scores. These factors include the quality and progression of therapy, amount of therapy per session, and the patient types treated.

The sample for calculation of the responsiveness statistics did not include subjects who had deteriorated; therefore, these results cannot be generalized to this type of population. In addition, the SRM and ES calculated in this study can only be used in the interpretation of group differences. Specifically, these values can only be applied to those studies that compare pretest scores with posttest scores because that was the design of this study.

The goal for the assessment of responsiveness was to determine the ability of the self-report section of the ASES to detect change when it occurred. Choosing the GRC as an external criterion for change has its

limitations, as detailed in the data analysis section. Other external criteria that have been suggested as indicators of patient improvement or deterioration are (1) treatment of known efficacy; (2) a measure that has been demonstrated to be reliable and valid, such as an impairment measure; and (3) the stage of the diagnosis, such as acute versus chronic.[2,5]

The results of this study can be generalized only to those individuals and groups of patients who possess characteristics similar to those of the sample in this study. In addition, with regard to use of responsiveness indices for group comparisons, they should be generalized only to studies with a within-group design, such as this study.

One criticism of the patient self-report section of the ASES is the instrument's combination of the measures of pain and function for a total score. Because pain is an impairment, we believe that pain should not be included to generate a score of functional limitations. Therefore, we calculated the measurement properties of only the function component so that a clinician or researcher can use the function component score separately to assess functional limitations of an individual patient or a group of patients. The measurement properties for the self-report section of the ASES function component are shown in Tables III and IV and indicate that it is reliable and responsive. In addition, an ROC curve analysis indicated that the area under the curve was significant (area, 0.90; $P = .008$; SE, 0.043). The cutoff point for clinically meaningful change for the ASES function scores was 3.4 points (sensitivity, 86%; specificity, 75%).

Future work is needed for the further investigation of the measurement properties of the patient self-report section of the ASES. The SRMs and ESs need to be determined for between-group comparisons, to aid in decision making with regard to treatment effectiveness in a design that involves a control group and treatment group, or two treatment groups.

This study has provided evidence for the use of the patient self-report section of the ASES as both an evaluative and discriminative instrument for patients with shoulder dysfunction. As a result, these measurement properties can be applied to studies or individual patients with characteristics similar to those in this study. The self-report section of the ASES can be used to assess both individual patients and groups of patients with confidence. As a measure of functional limitations, the self-report section of the ASES is short and easy to administer and interpret, with a minimal investment of time required for either the clinician or researcher. The self-report section of the ASES has been demonstrated to be reliable, valid, and responsive to clinical change, thereby supporting its use as a tool with which to assess functional limitations in patients with shoulder dysfunction.

We would like to express our sincere appreciation to the clinicians from the following clinics who participated in the data collection process, because without their assistance this project could not have been completed: Penn Therapy and Fitness, University of Pennsylvania Health Systems, Philadelphia, Pa; LifeCare Medical Center, Glassboro, NJ; Rehabilitation and Fitness Services, Philadelphia, Pa; NovaCare, Inc, Cinnamenson, NJ; Healthsouth, Philadelphia and Levittown, Pa; Temple Sports Medicine, Fort Washington, Pa; Bryn Mayr Sports Medicine, Bryn Mayr, Pa; Lycoming Physical Therapy, Montoursville, Pa; and Crozer-Chester Medical Center, Springfield, Pa. In addition, we would like to thank Dan Riddle, PhD, PT, Virginia Commonwealth University-Medical College of Virginia, Richmond, for his helpful suggestions and comments in reviewing this manuscript.

REFERENCES

1. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg 1998;7:565-72.
2. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. Spine 2000;25:3192-9.
3. Beaton DE, Richards RR. Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. J Bone Joint Surg Am 1996;78:882-90.
4. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986;39:897-906.
5. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. Phys Ther 2001;81:776-88.
6. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622-9.
7. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? J Clin Epidemiol 1992;45:1341-5.
8. Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171-8.
9. Heald SL, Riddle DL, Lamb RL. The shoulder pain and disability index: the construct validity and responsiveness of a region-specific disability measure. Phys Ther 1997;77:1079-89.
10. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
11. Kantz ME, Harris WJ, Levitsky K, Ware JE Jr, Davies AR. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. Med Care 1992;30(Suppl 5):MS240-52.
12. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(Suppl 3):S178-89.
13. L'Insalata JC, Warren RF, Cohen SB, Altchek DW, Peterson MG. A self-administered questionnaire for assessment of symptoms and function of the shoulder. J Bone Joint Surg Am 1997;79:738-48.
14. Leggin BG, Neuman RM, Shaffer MA, Iannotti JP, Williams GR, Brenneman SK. Reliability and validity of a shoulder outcome scoring system [abstract]. J Orthop Sports Phys Ther 1996;23:65.
15. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990; 28:632-42.
16. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996;1:30-46.
17. McHorney CA, Ware JE Jr, Rogers W, Raczek AE, Lu JF. The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts. Med Care 1992; 30(5 Suppl):MS253-65.
18. Norman GR, Stratford P, Regehr G. Methodological problems in

the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 1997;50:869-79.
19. Nunnally JC, Bernstein IH. Psychometric theory. New York: McGraw-Hill, Inc; 1994. p. 83-101, 251-92.
20. Portney LG, Watkins MP. Foundations of clinical research: applications to practice, 2nd ed. Upper Saddle River (NJ): Prentice Hall Health; 2000. p. 565, 705-11.
21. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg:347-52.
22. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res 1991;4:143-9.
23. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol 1995; 48:1369-78.
24. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
25. Williams JW Jr, Holleman DR Jr, Simel DL. Measuring shoulder function with the Shoulder Pain and Disability Index. J Rheumatol 1995;22:727-32.
26. Cohen J. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

Appendix: ASES

Patient Self-Evaluation

Pain:

How bad is your pain today (mark line)?

No pain at all — Pain as bad as it can be

Function: Circle the number in the box that indicates your ability to do the following activities:

0 = Unable to do; 1 = Very difficult to do; 2 = Somewhat difficult; 3 = Not difficult

Activity	Right arm	Left arm
1. Put on a coat	0 1 2 3	0 1 2 3
2. Sleep on your painful or affected side	0 1 2 3	0 1 2 3
3. Wash back/do up bra in back	0 1 2 3	0 1 2 3
4. Manage toiletting	0 1 2 3	0 1 2 3
5. Comb hair	0 1 2 3	0 1 2 3
6. Reach a high shelf	0 1 2 3	0 1 2 3
7. Lift 10 lbs above shoulder	0 1 2 3	0 1 2 3
8. Throw a ball overhand	0 1 2 3	0 1 2 3
9. Do usual work – List: _____	0 1 2 3	0 1 2 3
10. Do usual sport – List: _____	0 1 2 3	0 1 2 3