

Winter 2-28-2016

Intra and Inter-Rater Reliability and Convergent Validity of FIT-HaNSA in Individuals with Grade II Whiplash Associated Disorder

Michael Pierrynowski

Colleen McPhee


Saurabh Mehta

Marshall University, mehtas@marshall.edu

Joy C. MacDermid

Anita Gross

Follow this and additional works at: http://mds.marshall.edu/physical_therapy_faculty

 Part of the [Medical Humanities Commons](#), [Orthopedics Commons](#), [Sports Medicine Commons](#), and the [Sports Studies Commons](#)

Recommended Citation

Pierrynowski M, McPhee C, Mehta SP, MacDermid JC, Gross A. Intra and Inter-Rater Reliability and Convergent Validity of FIT-HaNSA in Individuals with Grade II Whiplash Associated Disorder. *The Open Orthopaedics Journal*. 2016;10:179-189.

This Article is brought to you for free and open access by the Physical Therapy at Marshall Digital Scholar. It has been accepted for inclusion in Physical Therapy Faculty Research by an authorized administrator of Marshall Digital Scholar. For more information, please contact zhangj@marshall.edu, martj@marshall.edu.



The Open Orthopaedics Journal

Content list available at: www.benthamopen.com/TOORTHJ/

DOI: 10.2174/1874325001610010179



Intra and Inter-Rater Reliability and Convergent Validity of FIT-HaNSA in Individuals with Grade II Whiplash Associated Disorder

Michael Pierrynowski¹, Colleen McPhee², Saurabh P. Mehta^{3,*}, Joy C. MacDermid^{1,4} and Anita Gross¹

¹School of Rehabilitation Sciences, McMaster University, Hamilton, Ontario, Canada

²Stonechurch Family Health Centre, Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada

³School of Physical Therapy, Marshall University, Huntington, West Virginia, United States

⁴School of Physical Therapy, University of Western Ontario, London, Ontario, Canada

Received: September 16, 2015

Revised: February 25, 2016

Accepted: February 28, 2016

Abstract:

Background:

Whiplash-Associated Disorders (WAD) are common following a motor vehicle accident. The Functional Impairment Test - Hand, and Neck/Shoulder/Arm (FIT-HaNSA) assesses upper extremity physical performance. It has been validated in patients with shoulder pathology but not in those with WAD.

Objectives:

Establish the Intra and inter-rater reliability and the known-group and construct validity of the FIT-HaNSA in patients with Grade II WAD (WAD2).

Methods:

Twenty-five patients with WAD2 and 41 healthy controls were recruited. Numeric Pain Rating Scale (NPRS), Neck Disability Index (NDI), Disabilities of the Arm, Shoulder and Hand (DASH), cervical range of motion (CROM), and FIT-HaNSA were completed at two sessions conducted 2 to 7 days apart by two raters. Intraclass correlation coefficients (ICC) were used to describe Intra and inter-rater reliability. Spearman rank correlation coefficients (ρ) were used to quantify the associations between scores of the FIT-HaNSA and other measures in the WAD2 group (convergent construct validity).

Results:

The Intra and inter-ICCs for the FIT-HaNSA scores ranged from 0.88 to 0.89 in the control group and 0.78 to 0.85 in the WAD2 group. Statistically significant differences in FIT-HaNSA performance between the two groups suggested known group construct validity ($P < 0.001$). The correlations between the NPRS, NDI, DASH, CROM and FIT-HaNSA were generally poor ($\rho < 0.4$).

Conclusion:

The study results indicate that the total FIT-HaNSA score has good Intra and inter-rater reliability and the construct validity in WAD2 and healthy controls.

Keywords: FIT-HaNSA, Psychometrics, Shoulder functional ability, Shoulder performance, WAD2, Whiplash.

* Address correspondence to this author at the School of Physical Therapy, Marshall University, Huntington, United States; Email: mehtas@marshall.edu

INTRODUCTION

Whiplash-Associated Disorders (WAD) are the most common type of injuries following a motor vehicle accident [1]. Grade I and II injuries represent 90% of WAD claims [2]. Grade II WAD (WAD2) cases persisting beyond two to six months result in most of the financial burden and are warning signs of impending chronicity [3]. The incidence of WAD2 in Western countries is 300 per 100,000 inhabitants [4].

Evidence-based clinical practice guidelines suggest that the physical examination of WAD2 patients should include tests of inspection, range of motion, strength, palpation, provocation, muscular stability and cervical proprioception [5]. Research also suggests that apart from self-report ability measures (e.g. the Neck Disability Index (NDI) or the numeric pain rating scale (NPRS)), measures assessing physical performance should also be utilized while assessing WAD 2 patients [6]. In a clinical setting, physical performance can be assessed by testing a patient's ability to execute a standardized activity in a standardized environment [7]. Usually, time to complete the activity or number of repetitions performed are used to quantify the physical performance [8]. Conversely, self-report measures examine patients' perception and experience of their ability to perform functional tasks [7]. In patient groups with various musculoskeletal diagnoses, such as advanced knee or hip osteoarthritis and chronic low back pain, poor to fair concordance between physical performance and self-report measures of ability suggest that they assess different perspectives of function [8, 9]. As such both physical performance tests and self-report measures, are complementary, and should both be used [8 - 10]

The Functional Impairment Test - Hand, and Neck/Shoulder/Arm (FIT-HaNSA) is a relatively new physical performance test that measures the ability of a patient to perform upper limb reach, object grip and manipulation, and sustained overhead positioning [11]. FIT-HaNSA development and psychometric evaluation have been described elsewhere [11]. Excellent test-retest reliabilities (intraclass correlation coefficient (ICC) > 0.96) for the FIT-HaNSA were reported in controls and individuals with shoulder disorders [11, 12]. In addition, the FIT-HaNSA scores have also demonstrated good discriminant validity as well as expected convergent ($r = -0.73$ to -0.83 with the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire) and divergent relationships (strength ($r = 0.12$ to 0.66) and shoulder range of motion ($r = 0.45$ to 0.64)) in patients with shoulder pathology [11, 12].

FIT-HaNSA has been used with patients with known neck and shoulder pathology [11, 12], however its psychometric properties have not been evaluated in the WAD2 population. Additionally, previous research studies have not examined the intra and inter-rater reliability of FIT-HaNSA. The purposes of this study are to estimate: 1) inter and Intrarater reliability, 2) known group construct validity, and 3) convergent validity with self-report measures of pain, ability, and impairment of FIT-HaNSA in samples of participants with and without WAD2.

MATERIAL AND METHODS

Participants

Forty-one control participants were recruited through public advertisement at a University and a Hospital. Control participants were included if they were over the age of 18, fluent in writing and speaking English and were not experiencing head, neck or upper extremity pain at the time of testing. Exclusion criteria for the control participants were: past history of a motor vehicle accident requiring rehabilitative treatment and concurrent medical concern that could significantly alter performance (e.g. past or present cervical disc herniation, cervical fracture or instability diagnosed through imaging techniques, previous neck or upper extremity surgery, rotator cuff tear diagnosed through ultrasound, neurological conditions affecting upper extremity, rheumatoid arthritis or fibromyalgia).

Twenty-five participants with WAD2 were consecutively recruited from a private outpatient physiotherapy practice during a one year period. Eligibility was determined by one of the clinic's two physiotherapists. All of the inclusion and exclusion criteria used to enroll the control participants were followed except that the WAD2 participants were experiencing neck pain (with or without head, face and arm pain) as the result of a motor vehicle accident greater than six weeks ago and were classified as WAD2 using the Spitzer criteria [3].

Raters

Two physical therapists, both Fellows of the Canadian Academy of Manual Therapists, administered all measures to the WAD2 patients. The physical therapists attended a 90 minute training session one month prior to the start of the study to become proficient in administering the FIT-HaNSA and to review documentation procedures [9]. During the

training session, the raters tested non-study volunteers. A second pair of raters assessed the control participants. These raters were instructed by the developer of FIT-HaNSA (JM) and a physical therapist (CM) regarding the administration of the pain, ability and impairment measures.

Measures

Measures for pain (NPRS), self-reported disability (NDI and DASH), movement impairment (active cervical range of motion (ROM)) and physical performance (FIT-HaNSA) were obtained.

NPRS: The NPRS is commonly used in assessing neck pain intensity [13, 14]. A patient is asked to rate his pain intensity over a 24 hour period on an 11 point scale where 0 indicates “no pain” and 10 indicates the “worst pain imaginable”. The NPRS demonstrates fair to good reliability (ICC = 0.64 to 0.86) when used in neck pain populations [14, 15]. The minimal detectable change (MDC₉₀) of the NPRS ranges from 1.3 to 2 points in a mixed orthopedic group including chronic neck pain [13, 15].

NDI: The NDI is a 10-item disease-specific self-report measure of function that captures perceived disability resulting from neck pain [16]. Each item is measured on a six point scale from zero (no disability) to five (complete disability) with total score between 0 to 50 [16]. The NDI total score can be interpreted, in regards to level of disability, as follows: 0 to 4 = none; 5 to 14 = mild; 15 to 24 = moderate; 25 to 34 = severe and over 34 is considered complete [16, 17]. The NDI has fair to good test retest reliability (ICC values between 0.50 to 0.98) in patient with different neck diagnoses and its MDC₉₀ is known to be between 5 to 10 points [18].

DASH: The DASH is 30-item region-specific outcome measure developed to evaluate upper extremity functional status in presence of musculoskeletal condition [19]. Each item is scored on a scale of one to five with the total score range of 0 and 100 with lower scores indicating greater ability. The reliability of the DASH is good (ICC of 0.90) and its MDC₉₀ is 10.2 points [20]. It has been validated for use for patients with neck pain [21].

CROM: Cervical ROM was measured using a mechanical protractor (CROM) (Performance Attainment Associates, St Paul, MN) that quantified cervical angular ROM, in degrees, in the sagittal (flexion-extension), frontal (right-left bend) and transverse (right-left axial rotation) planes. Measurement of cervical ROM is common in evaluation of patients with neck pain [22]. The reliability of the CROM is good (ICC = 0.80) in patients with neck pain [23] Fletcher and Bandy [17] reported an MDC₉₀ of 5° to 10° for each plane of motion for cervical ROM.

FIT-HaNSA: The FIT-HaNSA protocol consists of three timed tasks and each task is performed for a maximum of 300 seconds (s) with approximately 30 s pause between them (set-up time for next task). Task 1 (waist-up) requires the patient to alternately “grab, lift, move and place” three 1000 g containers located on waist level and 25 cm above waist level shelves, using their affected arm, at a metronome pace of 60 beats per minute for 300 s or until they felt unable to continue. The time to complete Task 1 is measured using a stopwatch. Task 2 (eye-down) is identical to Task 1 except that the two shelves are placed at eye-level and 25 cm below. Task 3 (overhead work) requires a patient to repeatedly screw and unscrew bolts in a sagittal plane oriented plate positioned at eye-level using both arms. The FIT-HaNSA tasks have demonstrated excellent test retest reliability (ICC > 0.89) in healthy controls and those with shoulder pathology [11, 12].

Study Procedures

Four raters, two at each site, were randomized prior to the start of the study to determine who would assess each participant at each of the two test sessions. Approximately, half of the participants were assessed by the same rater at the two test sessions - the other half was assessed by different raters at the second test session. Testing occurred in a physiotherapy clinic or in a university laboratory using the JobSim System (JTECH Medical, Salt lake City, UT). The WAD2 participants completed the study protocol before the start of treatment on the test day. During each participant's first session, information regarding his/her age, sex, height, mass, and accident date (if a WAD2 participant) was recorded. Three self-report measures (NPRS, NDI, DASH) were completed by the participant. Shortly thereafter, the rater assessed cervical ROM using the CROM, and then FIT-HaNSA was administered. Finally, the same rater scored the self-report measures and placed the participants data in an envelope then sealed. These envelopes were coded such that WAD2 or control group identity was blinded.

The participant was then scheduled to attend a second session 2 to 7 days following the first session. The WAD2 participants continued their medications and prescribed therapy between the two test sessions. The control participants were requested to refrain from non-typical activities. During the second session, the NPRS, NDI, DASH, CROM and

FIT-HaNSA were administered. Again, the participant's data were placed in an envelope and sealed.

The research ethics board at McMaster University, Hamilton, Canada approved the study.

Data Analysis

Two authors (MP, CM) performed data entry, screening, and inspection of scatterplots and histograms. Descriptive statistics including means and standard deviations were calculated for age, height, mass and duration of symptoms to determine the baseline characteristics of the WAD2 and control groups. Similar descriptive statistics were also calculated for the scores on the outcome measures for the two groups at each session.

The ICC were calculated to examine the reproducibility of the FIT-HaNSA tasks. Separate analyses were conducted for the same raters, each performing a FIT-HaNSA assessment at two sessions (Intrarater reliability) and different raters, each performing a FIT-HaNSA assessment at two sessions (inter-rater reliability). ICCs were calculated for each of the three FIT-HaNSA tasks and the total score for both the WAD2 and control groups. ICC values of > 0.7 were considered suggestive of good reliability [24].

The standard error of measurement (SEM) quantifies the error associated with a single score [25]. SEM was determined by calculating the square root of the mean square error term from the analysis of variance (ANOVA) tables. The SEM value was used to calculate the 90% minimal detectable change ($MDC_{90} = 1.65 \times SEM \times \sqrt{2}$) which is a statistic used to assess whether the change in a participant's score over time is a true change *versus* random error [26].

Bland and Altman analyses were used to determine the limits of agreement associated with FIT-HaNSA scores. An ICC is influenced by the size of the between-subjects variance [27], whereas Bland and Altman analyses examine within-subject variability or random error. Bland and Altman analyses plot the differences in scores between test and retest against the mean of test and retest scores. The mean difference and the standard deviation of the differences are used to construct 95% limits of agreement (LOA_{95}) [28].

To construct validity for a human group, a repeated measures mixed design ANOVA was used to determine if there were significant differences in performance between the WAD2 and control groups across sessions for the three tasks and total FIT-HaNSA scores. The main effect for the between-group factor was deemed statistically significant at the $P < 0.05$ level.

To measure convergent construct validity Spearman rank correlation coefficients (ρ) were used to quantify the magnitude of the association between FIT-HaNSA and the NPRS, NDI, DASH and CROM in the WAD2 group on each occasion. The correlations were classified as poor if $|\rho| < 0.40$, fair ($0.4 \leq |\rho| \leq 0.75$) and good ($|\rho| > 0.75$) [29].

SPSS Version 18.0 (SPSS Inc, Chicago, IL) was used for all analyses.

RESULTS

In the WAD2 group, there were nineteen females (76%) and 6 males (24%) and in the control group there were twenty-nine females (71%) and twelve males (29%). The WAD2 and control groups were similar in age, height, and mass (Table 1). The mean duration of symptoms in the WAD2 group was 12.4 months suggesting chronicity of the neck pain. Examination of the NPRS, NDI, DASH and CROM scores further characterize the groups (Table 2). The NPRS, NDI and DASH scores for the WAD2 compared to the control group suggests greater variability as indicated by the larger standard deviations. The mean CROM scores were relatively stable between sessions for each group.

Table 1. Demographics (mean, standard deviation) of the WAD2 and control groups.

	WAD2 (n = 25)	Control (n = 41)	Student t-test
Age (years)	36.4 (13.8)	34.0 (14.2)	$p = 0.58$
Height (centimeters)	165.9 (11.4)	169.6 (9.7)	$p = 0.19$
Mass (kilograms)	71.1 (20.5)	75.7 (26.6)	$p = 0.27$
Duration of Symptoms (months)	12.4 (13.5)		

The mean scores for three tasks and the total score were lower for the WAD2 group compared to the controls (see Table 2). The average FIT-HaNSA scores between two sessions in the WAD2 group differed by 12 s for Task 1, 5 s for

Task 2, and 0 s for Task 3. WAD2 patients and control group performed best on Task 1 followed by Task 3 and scored poorest on Task 2. Two WAD2 and 22 control participants demonstrated ceiling effects (achieved scores of 300 s). No WAD2 or control participant scored 0 s.

Table 2. Descriptive statistics (mean, standard deviation) for the self-report, cervical motion and FIT-HaNSA scores for the WAD2 and control groups at two test sessions.

Measure	WAD2 (n = 25)		Control (n = 41)	
	Session 1	Session 2	Session 1	Session 2
NPRS (%)	51 (21)	54 (26)	1 (4)	2 (4)
NDI (%)	43.8 (12.6)	42.9 (14.6)	3.4 (4.2)	2.4 (4.1)
DASH (%)	38.9 (15.7)	38.3 (16.0)	1.7 (2.3)	1.2 (2.0)
CROMrr (°)	61 (12)	62 (14)	72 (10)	74 (7)
CROMlr (°)	63 (13)	63 (12)	74 (8)	74 (8)
CROMf (°)	45 (12)	44 (11)	54 (8)	57 (9)
CROME (°)	50 (17)	51 (18)	78 (15)	73 (15)
CROMrb (°)	33 (10)	34 (12)	47 (13)	47 (9)
CROMlb (°)	35 (9)	35 (9)	50 (12)	50 (12)
FIT-HaNSA Task 1 (s)	201 (80)	213 (75)	296 (20)	294 (36)
FIT-HaNSA Task 2 (s)	117 (75)	122 (66)	232 (89)	251 (78)
FIT-HaNSA Task 3 (s)	170 (79)	170 (75)	281 (49)	277 (59)
FIT-HaNSA Total (s)	488 (208)	506 (189)	809 (132)	820 (141)

NPRS = Numeric Pain Rating Scale; NDI = Neck Disability Index; DASH = Disability of the Arm, Shoulder and Hand, CROMrr = cervical right axial range of motion, CROMlr = cervical left axial range of motion, CROMf = cervical flexion range of motion, CROME = cervical extension range of motion, CROMrb = cervical right bend range of motion, CROMlb = cervical left bend range of motion

The intra and inter-rater reliabilities were fair to excellent for Tasks 1, 2 and 3 and the total score (see Table 3) when the raters assessed the control participants. The reliability coefficients were lower (ICC of 0.54 to 0.80) when the WAD2 participants were tested. The total FIT-HaNSA scores for the WAD2 had larger SEM and MDC₉₀ values compared to the control group. The WAD2 group’s SEM and MDC₉₀ were 76 and 176 s, respectively for the inter-rater testing compared to the control group’s 41 and 95 s, respectively.

Table 3. Intra and inter-rater reliabilities of Task 1, 2, 3, and total FIT-HaNSA scores for the WAD2 (n=18) and control (n=41) participants. Different numbers of participants (indicated below) were included in the Intra and inter-rater reliability calculations. Intraclass correlation coefficients (ICC_{2,1}) with 95% confidence intervals (in parentheses) are presented.

		Task 1	Task 2	Task 3	Total FIT-HaNSA
WAD2 (n = 25)	Intrarater (n = 11)	0.72 (0.27 - 0.91)	0.70 (0.21 - 0.91)	0.70 (0.20 - 0.90)	0.78 (0.37 - 0.94)
	Inter-rater (n = 14)	0.54 (0.01 - 0.82)	0.75 (0.39 - 0.91)	0.80 (0.51 - 0.93)	0.84 (0.59 - 0.95)
Control (n = 41)	Intrarater (n=18)	0.76 (0.48 - 0.90)	0.74 (0.44 - 0.73)	0.97 (0.92 - 0.99)	0.88 (0.72 - 0.95)
	Inter-rater (n=23)	NA	0.76 (0.52 - 0.89)	0.93 (0.85 - 0.97)	0.89 (0.77 - 0.95)

The Bland and Altman plot for the WAD2 group’s total FIT-HaNSA scores (Fig. 1) indicated a 26 s bias. The standard deviation of the difference was 124 s for the WAD2 group and the 95% LOA was 248 s. The Bland and Altman plot for the control groups’ total FIT-HaNSA score indicated some systematic positive improvement in performance (bias) as the mean Session 2 compared to Session 1 difference was 13 s higher (Fig. 2). These scores were evenly distributed above and below the bias line indicating that the variance was not influenced by the size of the mean. The standard deviation of the differences was 64 s for the control group and the 95% LOA was 128s.

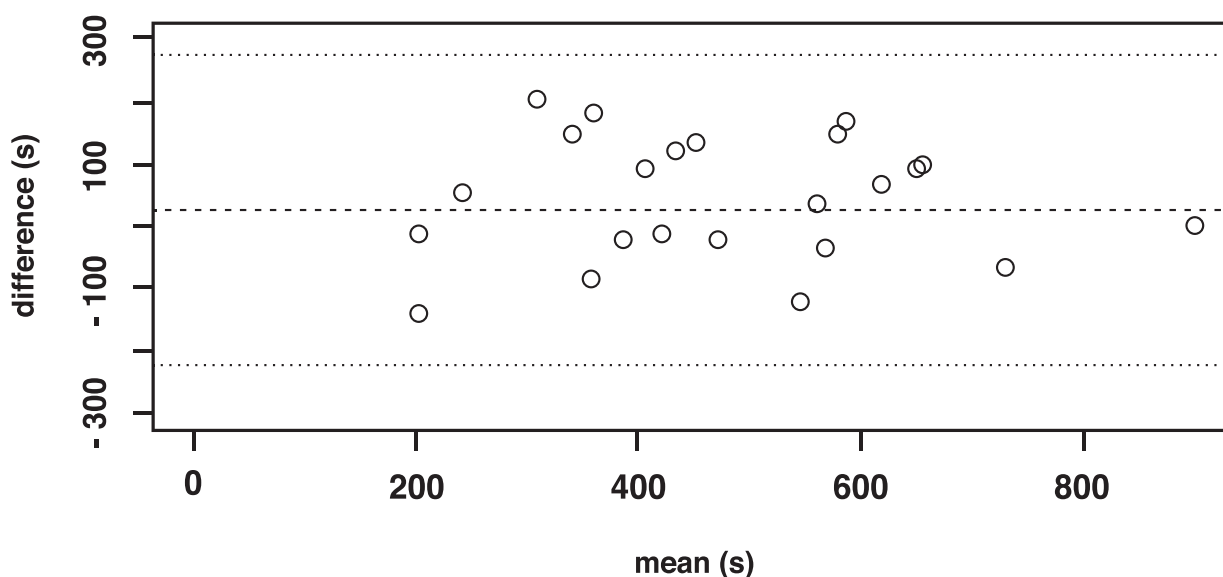


Fig. (1). The difference between the Session 2 and Session 1 total FIT-HaNSA scores (vertical axis) and the mean of the Session 1 and Session 2 total FIT-HaNSA scores (horizontal axis) for the 25 WAD2 participants. The mean difference (26 s) is the heavy dotted line and the limits of agreement (-223, 275) are the lighter dotted lines.

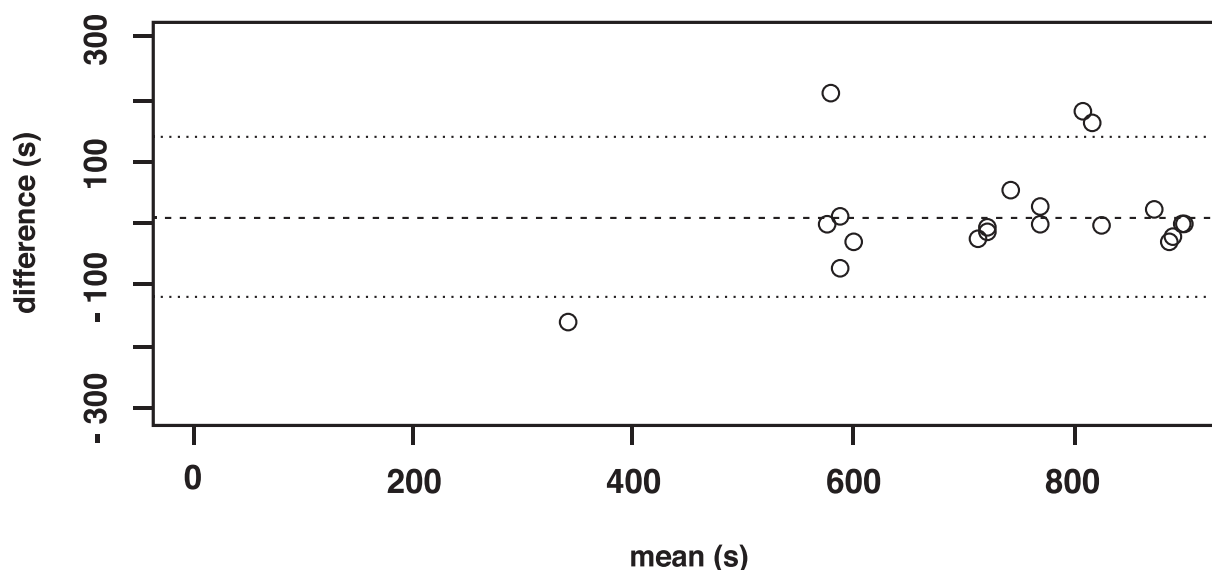


Fig. (2). The difference between the Session 2 and Session 1 total FIT-HaNSA scores (vertical axis) and the mean of the Session 1 and Session 2 total FIT-HaNSA scores (horizontal axis) for the control participants. The mean difference (12 s) is the heavy dotted line and the limits of agreement (-117, 141) are the lighter dotted lines.

The FIT-HaNSA performance differed significantly between the WAD2 and control groups (Task 1 ($F = 53.3, df = 1, 64, P < 0.001$), Task 2 ($F = 42.0, df = 1, 64, P < 0.001$), Task 3 ($F = 49.8, df = 1, 64, P < 0.001$)) and the total FIT-HANSA scores ($F = 62.6, df = 1, 64, P < 0.001$). Based on these findings the FIT-HANSA total score can be considered to have good known group construct validity.

Spearman rank correlations between the FIT-HaNSA, NPRS, NDI, DASH, and CROM scores for the WAD2 group for Session 1 are presented in Fig. (3). Of the 78 correlations, most (59) were poor ($p < 0.4$). The NDI - DASH and CROMrr - CROMlr scores had good correlations ($p > 0.75$) as did the correlations between the total and three Task FIT-HANSA scores (except Task1 - Task 3, which was fair).

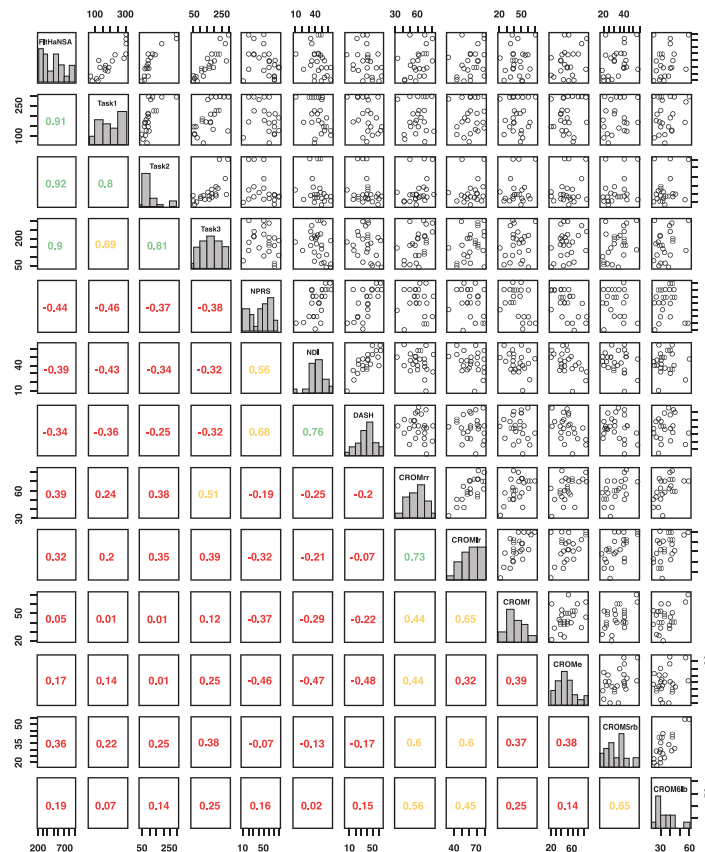


Fig. (3). Pairwise scatterplots (upper tableau) and Spearman rank correlations (lower tableau) for the total and Task 1, 2, and 3 FIT-HaNSA, NPRS, NDI, DASH, CROMr, CROMl, CROMf, CROMe, CROMrb, and CROMlb scores for the 25 WAD2 participants assessed at Session 1. The correlations are color coded as red = low ($|\rho| < 0.4$), amber = moderate ($0.4 \leq |\rho| \leq 0.7$) and green = high ($|\rho| > 0.7$). Each outcome’s distribution is also plotted as a histogram along the diagonal. Abbreviations defined in Table 2.

DISCUSSION

This study found that the FIT-HaNSA has fair to good within and between raters’ reliabilities, for both WAD2 and control participants and can discriminate between WAD2 and control participants. It showed poor concordance with pain, ability and impairment measures, suggesting that it measures a different aspect of outcome. The study provides further support to the preliminary literature regarding the reliability and validity of the FIT-HaNSA that were conducted in patients with shoulder pathology [11, 12].

The results of the study can be generalized to patients with chronic WAD2 attending an outpatient clinic. The WAD2 population was consecutively sampled which minimized volunteerism and other selection biases that may call into question representativeness [30]. Patients with WAD2 can exhibit substantial heterogeneity in clinical presentation [31]. Since only 25 individuals with WAD2 were recruited, variations in test results may have been further inflated. The control sample was recruited across two sites (hospital and university setting) which improves the generalizability of the study results.

The raters and the WAD2 and control participants were blinded to the FIT-HaNSA scores between sessions to minimize attempts to match or improve performance. The test-retest interval of 2 to 14 days used in this study is considered acceptable for patients with musculoskeletal diagnoses [32]. This interval allows the participant the time to recover from potential muscular fatigue and pain yet close enough to mitigate real change in upper extremity function. Examination of the NPRS, NDI, DASH and CROM values suggest that the participants did not change between the two sessions.

Clinicians are interested in how WAD2 participants change over time. The MDC_{90} provides a metric when a clinician is 90% confident that a true change occurred. The MDC_{90} associated with the WAD2 group was 176 s (20 %). This suggests that a WAD2 participant’s total FIT-HaNSA score must change 176 s between test sessions to reflect a

true change. This required change of 20% of the total score is consistent with the change values associated with the NPRS, NDI and DASH that range from 7 % to 30% [13, 15, 18, 20, 33, 34]

The control group performed better as reflected in the three tasks and total FIT-HaNSA scores. Both groups performed poorest on Task 2. This result is similar to what has been observed previously in shoulder pathology groups [11, 12]. It has been hypothesized that Task 2 is the most difficult for shoulder impingement patients because it moves the injured shoulder into the “impinged position” [11, 12]. In WAD2 and control groups, Task 2 may be difficult because Task 1 “pre-fatigues” the upper quadrant musculature. Task 1 was included within the FIT-HaNSA performance test to challenge the shoulder girdle musculature prior to subsequent testing and reduce potential floor effects [11]. This approach is effective in that WAD2, shoulder pathology, and control groups perform best on Task 1 and no floor scores were reported [11, 12].

Clinical observation suggests that patients with WAD2 often cite that over-head tasks are difficult due to pain in the neck region and that they have decreased cervical extension ROM. It is reasonable to assume that WAD2 participants would have the greatest difficulty with Task 3 due to the requirement of “looking up” for 300 s. However, the study results indicate that Task 2 (repetitive upper extremity activity) was more challenging than Task 3. A similar finding was reported for participants with shoulder pathology [12] and in patients with chronic neck pain [35] suggesting that the interaction between the upper fibers of trapezius and the serratus anterior muscles are compromised during the performance of Task 2 in the presence of chronic pain [35]. Additionally, our WAD2 patients were observed to bilaterally shrug the shoulders to alter the mechanics of the neck and shoulder musculature and decrease the amount of neck extension during Task 3.

The Correlations Between the FIT-HaNSA and the NPRS, NDI, DASH and CROM were Generally Poor ($\rho < 0.4$)

The relationships between the FIT-HaNSA, NPRS, NDI, DASH and CROM scores reinforce the theory that the relationship between physical performance, pain, ability and impairment, whether they are determined by self-perception or actual performance, are varied and complex. Previous research in shoulder pathology groups has presented similar findings [11, 12]. Patients may either over or under estimate functional ability. For example, back pain patients with depression demonstrate a tendency to underestimate their functional ability but it had no significant effect on treadmill walking ability [36]. The results of this study substantiate the concept that pain, ability and impairment measures should be used in conjunction with a physical performance test when evaluating patients with WAD2 as they all provide information about different aspects of human health secondary to injury.

There are some limitations associated with this study. It was intended to achieve a sample of 40 WAD2 and 40 control participants to achieve acceptable power [37]. Dividing the participants among inter-rater and Intrarater groups further reduced the required sample sizes. The number of sessions could have been increased, which would have reduced the required sample size, but due to clinician time constraints, this was not possible. A more reasonable solution may be to evaluate the reliability of the protocol amongst raters who might be more apt to administer the test, such as kinesiologists and physical therapy assistants. This study was also limited in that the relationships between FIT-HaNSA and NPRS, NDI, DASH and CROM were examined at two closely separated times.

CONCLUSION

The results of this study indicate that the total FIT-HaNSA score can be reliability measured when used with patients with WAD2. The scores of the three tasks of FIT-HaNSA have fair to good within and between raters’ reliabilities for patients with WAD2. A clinician can administer FIT-HaNSA with confidence and interpret scores in a meaningful manner. The total FIT-HaNSA score discriminates between WAD2 and control participants demonstrating known group construct validity. An interpretation of the FIT-HaNSA convergent construct validity when used with WAD2 patients shows that most relationships between pain, ability and impairment are poor suggesting the continued use of a variety of assessment techniques in patients with WAD2.

LIST OF ABBREVIATIONS

ANOVA	=	Analysis of variance
CROM	=	Cervical range of motion
DASH	=	Disabilities of Arm, Shoulder, and Hand

FIT-HaNSA	=	Functional Impairment Test - Hand, and Neck/Shoulder/Arm
ICC	=	Intraclass correlation coefficient
LOA	=	Limits of agreement
MDC₉₀	=	Minimal detectable change at 90% confidence level
NDI	=	Neck Disability Index
NPRS	=	Numeric Pain Index Scale
SEM	=	Standard error of measurement
WAD	=	Whiplash associated disorders

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Zuby DS, Lund AK. Preventing minor neck injuries in rear crashes-forty years of progress. *J Occup Environ Med* 2010; 52(4): 428-33.
- [2] Ferrari R, Russell AS, Carroll LJ, Cassidy JD. A re-examination of the whiplash associated disorders (WAD) as a systemic illness. *Ann Rheum Dis* 2005; 64(9): 1337-42.
[<http://dx.doi.org/10.1136/ard.2004.034447>] [PMID: 15731286]
- [3] Spitzer WO, Skovron ML, Salmi LR, *et al.* Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: redefining "whiplash" and its management. *Spine* 1995; 20(8)(Suppl.): 1S-73.
[PMID: 7604354]
- [4] Holm LW, Carroll LJ, Cassidy JD, *et al.* The burden and determinants of neck pain in whiplash-associated disorders after traffic collisions: results of the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *Spine* 2008; 33(4)(Suppl.): S52-9.
[<http://dx.doi.org/10.1097/BRS.0b013e3181643ece>] [PMID: 18204401]
- [5] Scholten-Peeters GG, Bekkering GE, Verhagen AP, *et al.* Clinical practice guideline for the physiotherapy of patients with whiplash-associated disorders. *Spine* 2002; 27(4): 412-22.
[<http://dx.doi.org/10.1097/00007632-200202150-00018>] [PMID: 11840109]
- [6] Childs JD, Cleland JA, Elliott JM, *et al.* Neck pain: Clinical practice guidelines linked to the international classification of functioning, disability, and health from the orthopedic section of the American physical therapy association. *J Orthop Sports Phys Ther* 2008; 38(9): A1-A34.
[<http://dx.doi.org/10.2519/jospt.2008.0303>] [PMID: 18758050]
- [7] Finch E, Mayo NE, Stratford PW. *Physical Rehabilitation Outcome Measures: A Guide to Enhanced Clinical Decision-making*. 2nd ed. USA: Lippincott Williams and Wilkins 2002.
- [8] Simmonds MJ, Olson SL, Jones S, *et al.* Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. *Spine* 1998; 23(22): 2412-21.
[<http://dx.doi.org/10.1097/00007632-199811150-00011>] [PMID: 9836355]
- [9] Stratford PW, Kennedy D, Pagura SM, Gollish JD. The relationship between self-report and performance-related measures: questioning the content validity of timed tests. *Arthritis Rheum* 2003; 49(4): 535-40.
[<http://dx.doi.org/10.1002/art.11196>] [PMID: 12910560]
- [10] Novy DM, Simmonds MJ, Lee CE. Physical performance tasks: what are the underlying constructs? *Arch Phys Med Rehabil* 2002; 83(1): 44-7.
[<http://dx.doi.org/10.1053/apmr.2002.27397>] [PMID: 11782832]
- [11] MacDermid JC, Ghobrial M, Quirion KB, *et al.* Validation of a new test that assesses functional performance of the upper extremity and neck (FIT-HaNSA) in patients with shoulder pathology. *BMC Musculoskelet Disord* 2007; 8: 42.
[<http://dx.doi.org/10.1186/1471-2474-8-42>] [PMID: 17509150]
- [12] Kumta P, MacDermid JC, Mehta SP, Stratford PW. The FIT-HaNSA demonstrates reliability and convergent validity of functional performance in patients with shoulder disorders. *J Orthop Sports Phys Ther* 2012; 42(5): 455-64.
[<http://dx.doi.org/10.2519/jospt.2012.3796>] [PMID: 22281818]
- [13] Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC. Minimal clinically important change of the neck disability index and the numerical rating scale for patients with neck pain. *Spine* 2007; 32(26): 3047-51.
[<http://dx.doi.org/10.1097/BRS.0b013e31815cf75b>] [PMID: 18091500]
- [14] Cleland JA, Childs JD, Whitman JM. Psychometric properties of the neck disability index and numeric pain rating scale in patients with mechanical neck pain. *Arch Phys Med Rehabil* 2008; 89(1): 69-74.

- [http://dx.doi.org/10.1016/j.apmr.2007.08.126] [PMID: 18164333]
- [15] Stratford PW, Spadoni G. Feature articles: The reliability, consistency, and clinical application of a numeric pain rating scale. *Physiother Can* 2001; 53(2): 88-91.
- [16] Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther* 1991; 14(7): 409-15. [PMID: 1834753]
- [17] Fletcher JP, Bandy WD. Intrarater reliability of CROM measurement of cervical spine active range of motion in persons with and without neck pain. *J Orthop Sports Phys Ther* 2008; 38(10): 640-5. [http://dx.doi.org/10.2519/jospt.2008.2680] [PMID: 18827326]
- [18] MacDermid JC, Walton DM, Avery S, et al. Measurement properties of the neck disability index: a systematic review. *J Orthop Sports Phys Ther* 2009; 39(5): 400-17. [http://dx.doi.org/10.2519/jospt.2009.2930] [PMID: 19521015]
- [19] Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand. *Am J Ind Med* 1996; 29(6): 602-8. [http://dx.doi.org/10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L] [PMID: 8773720]
- [20] Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum* 2009; 61(5): 623-32. [http://dx.doi.org/10.1002/art.24396] [PMID: 19405008]
- [21] Mehta S, Macdermid JC, Carlesso LC, McPhee C. Concurrent validation of the DASH and the QuickDASH in comparison to neck-specific scales in patients with neck pain. *Spine* 2010; 35(24): 2150-6. [http://dx.doi.org/10.1097/BRS.0b013e3181c85151] [PMID: 20453724]
- [22] Prushansky T, Dvir Z. Cervical motion testing: methodology and clinical implications. *J Manipulative Physiol Ther* 2008; 31(7): 503-8. [http://dx.doi.org/10.1016/j.jmpt.2008.08.004] [PMID: 18804000]
- [23] Youdas JW, Carey JR, Garrett TR. Reliability of measurements of cervical spine range of motion-comparison of three methods. *Phys Ther* 1991; 71(2): 98-104. [PMID: 1989013]
- [24] Rosner B. *Fundamentals of Biostatistics*. Belmont, CA: Brooks. Cole 2005.
- [25] Stratford PW. Getting more from the literature: estimating the standard error of measurement from reliability studies. *Physiother Can* 2004; 56(1): 27-30. [http://dx.doi.org/10.2310/6640.2004.15377]
- [26] Portney L, Watkins M. *Foundations of clinical research: applications to clinical practice*. 3rd. Upper Saddle River, NJ: Prentice Hall 2009.
- [27] Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 1998; 12(3): 187-99. [http://dx.doi.org/10.1191/026921598672178340] [PMID: 9688034]
- [28] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1(8476): 307-10. [http://dx.doi.org/10.1016/S0140-6736(86)90837-8] [PMID: 2868172]
- [29] Fleiss JL. Reliability of Measurement. In: Fleiss JL, Ed. *The Design and Analysis of Clinical Experiments*. Toronto: John Wiley and Son 1986; pp. 1-32.
- [30] Hulley S, Cummings S, Browner W, Grady D, Hearst N, Newman T. *Designing clinical research*. USA: Lippincott Williams and Wilkins 2001.
- [31] Elliott JM, Noteboom JT, Flynn TW, Sterling M. Characterization of acute and chronic whiplash-associated disorders. *J Orthop Sports Phys Ther* 2009; 39(5): 312-23. [http://dx.doi.org/10.2519/jospt.2009.2826] [PMID: 19411765]
- [32] Streiner D, Norman G. *Health measurement scales*. 2nd ed. Oxford: Oxford University Press 1995.
- [33] Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001; 94(2): 149-58. [http://dx.doi.org/10.1016/S0304-3959(01)00349-9] [PMID: 11690728]
- [34] Cleland JA, Childs JD, Whitman JM. Psychometric properties of the neck disability index and numeric pain rating scale in patients with mechanical neck pain. *Arch Phys Med Rehabil* 2008; 89(1): 69-74. [http://dx.doi.org/10.1016/j.apmr.2007.08.126] [PMID: 18164333]
- [35] Hawkes DH, Alizadehkhayat O, Fisher AC, Kemp GJ, Roebuck MM, Frostick SP. Normal shoulder muscular activation and co-ordination during a shoulder elevation task based on activities of daily living: an electromyographic study. *J Ortho Res* 2012; 30(1): 53-60. [http://dx.doi.org/10.1002/jor.21482]
- [36] Wittink H, Rogers W, Sukiennik A, Carr DB. Physical functioning: self-report and performance measures are related but distinct. *Spine* 2003; 28(20): 2407-13. [http://dx.doi.org/10.1097/01.BRS.0000085304.01483.17] [PMID: 14560092]

- [37] Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther* 1997; 77(7): 745-50.
[PMID: 9225846]

© Pierrynowski *et al.*; Licensee *Bentham Open*

This is an open access article licensed under the terms of the Creative Commons Attribution-Non-Commercial 4.0 International Public License (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.