Arch Physioter 2025; 15: 148-157
ISSN 2057-0082 | DOI: 10.33393/aop.2025.3476
ORIGINAL RESEARCH ARTICLE

Validation of a Novel device for Assessing Neck Muscle Strength

Michail Arvanitidis[1], Hon Hin Ken Mak[1], Eduardo Martinez-Valdes[1], Marco Barbero[2], Deborah Falla[1]

[1]Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham - UK
[2]Rehabilitation Research Laboratory 2rLab, Department of Business Economics, Health and Social Care (DEASS), University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Manno - Switzerland

ABSTRACT

Background: The Neuromuscular Cranio-Cervical Device (NOD) was originally designed to evaluate Cranio-Cervical Flexion Test performance but can also be used as a handheld dynamometer for testing other muscle groups, including neck muscle strength. It offers a potential alternative to the Multi-Cervical Unit (MCU), a fixed dynamometer, more closely aligned with isokinetic dynamometry, the gold standard. However, its validity and reliability need to be established. This study aimed to evaluate concurrent validity compared to the MCU and inter- and intra-rater reliability of the NOD for measuring neck flexion and extension muscle strength.
Methods: Twenty participants were assessed for neck flexion/extension strength whilst in a seated position, with the measurements repeated over three sessions. Concurrent validity was assessed by comparing NOD measurements to the MCU using Pearson correlation coefficients, and reliability was determined using Intraclass Correlation Coefficients (ICCs).
Results: Concurrent validity was strong for extension (r = 0.954) but lower for flexion (r = 0.705), indicating some variability in flexion measurements. Inter-rater reliability was good to excellent for both flexion (ICC = 0.931) and extension (ICC = 0.896). Intra-rater reliability for extension was good to excellent (ICC = 0.893), while flexion ranged from moderate to excellent (ICC = 0.844).
Conclusions: The NOD is a valid tool, particularly for extension measurements, although further refinement of testing is needed to improve the accuracy for flexion strength measurements. It is also reliable for both extension and flexion, showing promise as a practical, affordable, portable tool with real-time feedback for the assessment of neck muscle strength in clinical settings.

Keywords: Handheld Dynamometer, Multi-Cervical Unit, Neck Muscle Strength, Concurrent Validity, Inter-rater Reliability, Intra-rater Reliability

What is already known about this topic:

- *The NOD, initially designed for use during the Cranio-Cervical Flexion Test, can also be used as a handheld dynamometer for neck strength testing, though its validity and reliability remain unclear.*

What does the study add:

- *This study establishes the NOD as a valid and reliable tool for assessing neck strength, particularly neck extension strength, providing a portable, practical alternative to fixed devices for clinical and research applications.*

Introduction

Muscle strength, typically defined as the force generated by muscles during a maximal isometric contraction (1), is a crucial measure in clinical practice for the assessment of physical function and for guiding rehabilitation progress (2). People with neck pain, like other musculoskeletal disorders, commonly present with reduced strength (3), and resistance training programmes are often prescribed to address these deficits (4,5). Additionally, individuals in remission from neck pain, particularly following a whiplash injury, frequently display reduced neck strength, with lower flexion strength being a predictor of higher neck disability at a 6-month follow-up (6). Assessing neck muscle strength is therefore essential to establish baseline physical function, monitor rehabilitation progress, and guide prognostic decisions. However, traditional manual muscle testing (MMT) methods, where clinicians provide resistance and estimate the force generated by the individual, have been criticized for their subjectivity, variability in application based on clinician skill, and lack of standardization (7). These limitations are especially problematic when testing neck flexion and

Received: January 27, 2025
Accepted: May 12, 2025
Published online: June 10, 2025

This article includes supplementary material

Corresponding author:
Michail Arvanitidis
email: m.arvanitidis@bham.ac.uk

© 2025 The Authors. This article is published by AboutScience and licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).
Commercial use is not permitted and is subject to Publisher's permissions. Full information is available at www.aboutscience.eu

extension, where no opposite side or limb is available for comparison, and factors like individual positioning and clinician experience further affect measurement accuracy (7). A recent systematic review and meta-analysis (8) confirmed that manual neck strength measurements lack clinical reliability, raising concerns about their ability to track changes in muscle function over time.

With advancements in technology, more objective methods for measuring muscle strength have emerged, such as force transducers, fixed frame dynamometry, and isokinetic dynamometry. Isokinetic dynamometers are devices that control movement speed and provide accommodating resistance throughout a predetermined range of motion (9). Isokinetic dynamometry, provides valuable insights into muscle strength across varying speeds, allowing for detailed analysis of muscle performance under controlled conditions. It is widely used for peripheral joints (10-14) but is also applied to the spine, particularly for assessing trunk muscles (15-21). For neck strength measurements, fixed frame dynamometry is more commonly used, with the Multi-Cervical Rehabilitation Unit (MCU) being a well-known example, due to its high reliability and validity (22). The MCU is a more sophisticated device, offering controlled testing conditions and greater stability, making it more comparable to isokinetic dynamometers, the gold standard (23) for objective muscle strength assessment. However, handheld dynamometers (HHDs) have become popular among clinicians due to their convenience, affordability, and ease of use. Various studies have assessed the reliability (2,24-33) and validity (24,29,30,32) of different HHDs. These studies used different protocols, including variations in participant positioning, which can significantly affect results. For example, Krause et al., (2018) demonstrated that neck strength measurements differ significantly between sitting and lying positions (28). Studies evaluating HHDs in seated positions, the position used when evaluating neck muscle strength with the MCU, are particularly relevant. However, only five studies (24,27-29,33) have assessed neck strength in seated participants, with only two (24,29) evaluating the concurrent validity of the HHD. Although these studies demonstrated high validity and reliability, neither directly compared an HHD to the MCU (i.e., they compared between HHDs and handheld mode versus a wall-mounted configuration), leaving a gap in the understanding of the accuracy of HHDs for the measurement of neck muscle strength.

Recently, a new handheld dynamometer called "*Neuromuscular Cranio-Cervical Device*" (NOD; OT Bioelettronica, Turin, Italy) was developed, offering distinct advantages over existing HHDs. The NOD provides greater stability through its two-handed operation and allows real-time visualization of the force produced via Bluetooth, enhancing interaction, as individuals can see their force output. The NOD's portability and enhanced features make it a promising candidate, but its validity, specifically concurrent validity, and reliability need to be rigorously tested. Therefore, this study aims to evaluate the validity of the NOD handheld dynamometer for measuring neck flexor and extensor muscle strength, in comparison to the MCU, and to also assess its reliability. This comparison will provide essential insights into the NOD's potential use in both clinical and research settings.

Methods

Design and setting

This validity and reliability study was conducted in accordance with the Declaration of Helsinki, approved by the Ethics Committee of the University of Birmingham (approval number: MCR2122_05). The methods and reporting of findings were carried out in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (34).

Data was collected between 01 April 2022 and 01 July 2022 at a laboratory located within the Centre of Precision Rehabilitation for Spinal Pain, School of Sport, Exercise and Rehabilitation Sciences at the University of Birmingham, UK. All participants provided written, informed consent before participating in the study and attended the laboratory on three separate days.

The study involved two raters and three sessions. Session 1 was conducted by Rater A (HHKM), who tested the participants using the handheld dynamometer (NOD) and the MCU (randomized order; rest of 15 mins between). This session was used to assess the validity of the handheld dynamometer compared to the MCU device. After 48 hours (session 2), Rater B (MA) performed the same neck muscle strength measurements using the handheld dynamometer to evaluate inter-rater reliability. The final session (session 3), conducted seven days after session one, was performed by Rater A using the handheld dynamometer to assess intra-rater reliability, capturing potential variability over different days while ensuring no significant changes in strength.

Session 2 was carried out within 48 hours after session 1 to achieve a balance between fatigue recovery and homogeneity, as conducting both measurements within the same session might over-exert participants and introduce performance bias (35). A time interval longer than one day allows for sufficient recovery while maintaining consistency in the maximal voluntary contraction (MVC). Session 3 took place seven days after session 1, which was five days after session 2, to further ensure adequate recovery and minimize the effects of fatigue, ensuring stability in MVC measurements over time.

Participants

A total of 20 participants (8 males, 12 females) were recruited from the student and staff population of the University of Birmingham using various recruitment methods, including email, posters, and peer contacts. To ensure a homogeneous sample and minimize confounding factors, this initial validation study was conducted exclusively with individuals without neck pain.

The sample size was calculated based on a moderate expected effect size of Pearson's $r = 0.6$ (36) for the assessment of concurrent validity using G*Power software version 3.1 (37,38). The parameters included an expected power (β) of 0.8, a significance level (α) of 0.05, and a two-tailed test. An initial sample size of 17 participants was determined. To account for a potential drop rate of 15%, the final required sample size was increased to 20 participants.

© 2025 The Authors. Published by AboutScience - www.aboutscience.eu

Inclusion and exclusion criteria

The study included individuals between the ages of 18 and 60 years old who reported pain-free neck movements. Individuals were excluded from the study if they had a history of neck pain that required treatment from a healthcare professional, cervical radiculopathy or myelopathy, spinal fractures or dislocation, spinal or thoracic surgery, osteoporosis, spinal infection, previous whiplash injury or spinal trauma, or any condition that required hospital medical care in the past 12 months. These criteria aimed to recruit an asymptomatic population with no significant prior history of neck conditions or injuries, ensuring participant safety and minimizing potential confounding factors.

Instrumentation and raters

The NOD dynamometer (OT Bioelettronica, Turin, Italy) is a 2-in-1 handheld dynamometer and biofeedback device consisting of a force cell with magnetic attachments. To record neck muscle strength, the NOD was connected to an Android tablet via Bluetooth, and its associated app was used. A silicon pad was magnetically attached to the force cell to enhance grip and participant comfort. The MCU consists of an armchair with an adjustable seat, back support and armrest, and a head fixation/assembly system where a force cell is positioned to measure the force of participants (22).

The neck muscle strength measurements from the NOD were recorded directly using the Android app in Newtons (NOD app, OT Bioelettronica, Turin, Italy). For the MCU device, the force signal was extracted via an Ethernet cable connected to a Forza device (OT Bioelettronica, Turin, Italy), which amplified the force signal (sampling rate: 100 Hz). The Forza was then connected to a data acquisition board (National Instruments/Emerson, Austin, Texas, USA), which was connected to a laptop, allowing real-time visualization, recording, and calculation of the peak force through a custom MATLAB script (version 9.9.0.1718557, R2020b, The MathWorks Inc., Natick, Massachusetts, USA). The force signal from the MCU was initially recorded in volts and subsequently converted into Newtons. Both dynamometers were calibrated once at the start of the study.

Both raters were male and experienced in using each device. Rater A had one year of experience using dynamometers, while Rater B had five years of clinical and research experience using dynamometers.

NOD measurements

The NOD handheld dynamometer (OT Bioelettronica, Turin, Italy) was used to assess MVCs of neck flexion and neck extension (randomized a priori using a randomization app). For these measurements, participants were secured with a strap around their waist while seated in a specific chair to minimize upper body compensation during force exertion (Fig. 1). For the measurement of neck flexion, the NOD was placed just superior to the participants' eyebrows (Fig. 1A), while for neck extension, the NOD was positioned against the occipital region of the skull (Fig. 1B). For each predetermined direction, participants performed one practice trial (one submaximal and one MVC contraction) to familiarize themselves with the procedure and to ensure the raters exerted a static force equal to the participants' force. After this familiarization, participants performed three MVCs in the specified direction. Verbal prompts were also provided by the raters to encourage a true maximal contraction. Participants were instructed to rest for two minutes between each MVC. The same procedure was then repeated for the opposite direction. These measurements were performed by Rater A (Session 1 and Session 3) and by Rater B (Session 2, 48 hours after Session 1).

Multi-Cervical Unit measurements

Participants were seated and stabilized using straps across their chest to prevent excessive upper body compensation during force exertion. Cervical flexion was measured by positioning the force cell of the MCU machine just above the participants' eyebrows, with a stabilization band placed behind the occipital protuberance and a stabilization clamp securing the back of the head (Fig. 1C, D). For neck extension, the force cell was positioned perpendicular to the occipital protuberance, and a stabilization band was secured just superior to the eyebrows (Fig. 1D). To familiarize participants with the testing procedure, they performed one submaximal and one MVC in the pre-determined direction (e.g., flexion). Participants were then instructed to perform three MVCs in that direction, with verbal encouragement provided by the rater, and a two-minute rest between each MVC. This protocol was repeated in the opposite direction (e.g., extension). These measurements were performed only during session 1, by Rater A.

Outcome measures and statistical analyses

Appropriate outcome measures and statistical analyses were selected in line with the GRRAS guidelines (34). Descriptive analysis was conducted to summarize the participants' demographics, with data presented as mean ± standard deviation. For statistical analysis, the mean peak isometric force from three measurements in both flexion and extension MVCs was used. Data normality was assessed using the Shapiro-Wilk test. Once normality was confirmed, parametric tests were used.

Concurrent validity was assessed by calculating Pearson's correlation coefficients (r) and the corresponding 95% CIs between the NOD and MCU measurements taken by rater 1 in session 1. Validity thresholds were classified as poor ($r < 0.50$), moderate ($0.50 \leq r < 0.70$), good ($0.70 \leq r < 0.90$), or excellent ($r \geq 0.90$), similarly to others (24). Pearson's correlation (*r*) coefficients were computed using SPSS Statistics, version 29 (IBM, USA).

The reliability of the NOD handheld dynamometer was evaluated using ICCs (3, k) and their 95% Confidence Intervals (CIs), calculated using a single-rating, absolute-agreement, two-way mixed-effects model, with two raters across 20 subjects. ICC values and their 95% CIs were calculated for inter-rater reliability (rater 1 vs rater 2 in session 2) and intra-rater reliability (rater 1 between sessions 1 and 3). Based on the criteria from (39), reliability was classified as poor ($ICC < 0.50$), moderate ($0.50 \leq ICC < 0.75$), good ($0.75 \leq ICC < 0.90$), or excellent ($ICC \geq 0.90$).

To assess measurement precision and interpretability, the Standard Error of Measurement (SEM) and Minimal

© 2025 The Authors. Arch Physioter - ISSN 2057-0082 - www.archivesofphysiotherapy.com

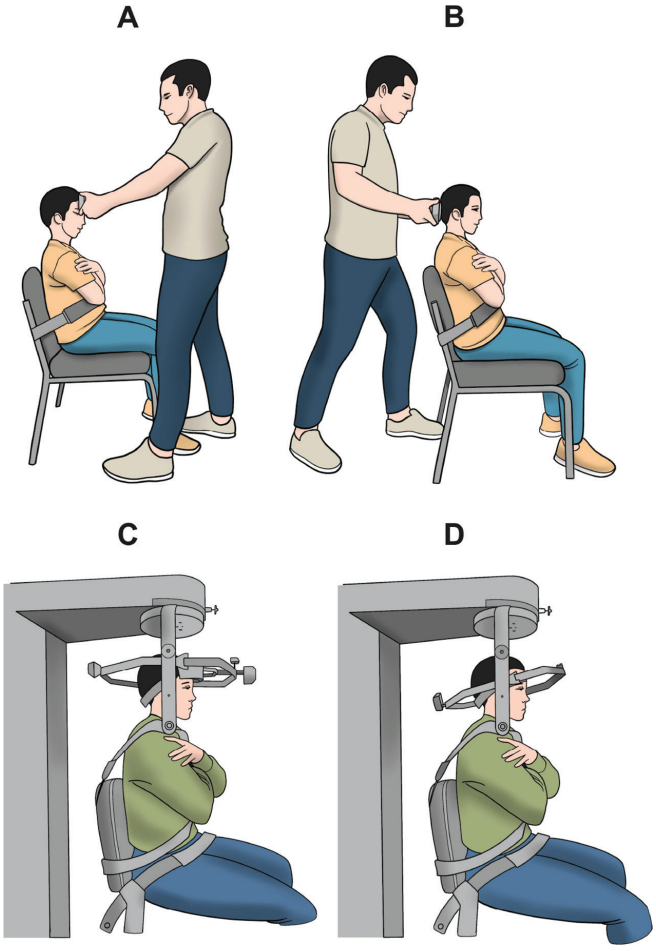


FIGURE 1 - Neck muscle strength testing for flexion and extension using two different methods: a handheld dynamometer (NOD) (A, B) and the Multi-Cervical Unit (MCU) (C, D). In panels A and B, the individual was seated with their arms crossed and a belt around the waist to stabilize the trunk, minimizing compensatory movements. The experimenter held the handheld dynamometer (NOD) on the forehead (A) for flexion and on the occiput (B) for extension, applying force while ensuring consistent contact with the head. In panels C and D, the individual is seated with their arms crossed and secured at the waist and chest with belts to stabilize the trunk and prevent extraneous movement. The head is positioned within the adjustable frame of the MCU, with the device set to the appropriate attachment for each movement. For extension (D), the attachment is tilted 15 degrees so that the point of force application is correctly aligned on the occiput.

Detectable Change (MDC) were calculated. SEM provides an estimate of the expected random variation in scores when no real change has occurred, reflecting the standard deviation of measurement error associated with each specific movement (flexion or extension) due to variability either between raters (inter-rater reliability) or across different sessions with the same rater (intra-rater reliability) (40,41).

For each movement, the pooled standard deviation (SD_{pooled}) was calculated by combining the standard deviations of the measurements across raters (for inter-rater reliability) or across sessions with the same rater (for intra-rater reliability). This represents the overall variability due to differences between or within raters and was calculated as:

$$SD_{pooled} = \sqrt{\frac{\sum(n_i - 1) \times SD_i^2}{\sum(n_i - 1)}}$$

where: SD_i is the standard deviation of measurements for each set, and n_i is the sample size for that set. The SEM was calculated as:

$$SEM = SD_{pooled} \times \sqrt{1 - ICC}$$

where the *ICC* reflects the test–retest reliability of the measurement. The MDC, representing the smallest real change beyond measurement error with 95% confidence (40,41), was then calculated as:

$$MDC = SEM \times \sqrt{2} \times 1.96$$

These metrics are crucial for identifying meaningful changes in clinical assessments and research, ensuring that observed differences are not due to random variability.

Values for SEM and MDC are reported in Newtons (N) for both flexion and extension, with separate calculations for inter-rater and intra-rater reliability. These measures provide insights into the consistency and accuracy of the NOD for assessing neck muscle strength across different testing conditions and serve as benchmarks for interpreting significant changes in muscle performance.

Scatter plots with their respective regression lines were created to visually assess the concurrent validity between MCU and NOD measurements, illustrating the strength and direction of their relationship. Additionally, Bland-Altman plots were created using GraphPad Prism version 10.3.1 (GraphPad Software, San Diego, California, USA) to visually assess the agreements and biases for the validity between the NOD and MCU measurements, as well as between the inter- and intra-rater NOD measurements for reliability. For all analyses, statistical significance was set at $\alpha = 0.05$.

Results

Participants and neck muscle strength

All 20 recruited participants (8M, 12F) attended all three sessions of the study. The characteristics of the participants are documented in Table 1. Additionally, the mean isometric neck muscle force of participants measured in each session are presented in Table 2.

TABLE 1 - Characteristics of participants

Participants' characteristics (N = 20)	Mean ± SD
Age (years)	20.7 ± 2.23
Height (cm)	165.2 ± 8.88
Weight (kg)	56.6 ± 8.07
BMI (kg/m²)	20.61 ± 1.77

© 2025 The Authors. Published by AboutScience - www.aboutscience.eu

TABLE 2 - Mean isometric neck muscle forces of all participants across sessions measures by both instruments. Data is reported as mean ± SD

Session	Equipment	Movement	Force
Session 1	*NOD*	*Flexion*	22.0 ± 5.97 N
		Extension	30.5 ± 6.63 N
	MCU	*Flexion*	19.2 ± 5.31 N
		Extension	30.7 ± 7.16 N
Session 2	*NOD*	*Flexion*	22.8 ± 5.82 N
		Extension	31.0 ± 5.99 N
Session 3	*NOD*	*Flexion*	23.4 ± 5.16 N
		Extension	30.3 ± 6.32 N

Concurrent validity

Poor to good agreement was found between the measurements of NOD and MCU for neck flexion strength. There was a significant positive correlation between the two variables, *r(18) = 0.705 [0.382, 0.875]*, *p<0.001.* (Fig. 2A). Good to excellent agreement was found for neck extension strength. There was a significant positive correlation between the two variables, *r(18) = 0.954 [0.885, 0.982], p<0.001. (*Fig. 2B)

Bland-Altman plots are presented in Figure 2C for neck flexion strength and 2D for neck extension strength. For flexion, the mean bias was −2.87 N, with limits of agreement ranging from −11.44 N to 5.70 N. For extension, the mean bias was 0.22 N, with limits of agreement ranging from −4.44 N to 4.88 N. For neck flexion, the bias suggests that NOD tends to slightly underestimate the values compared to MCU. However, for neck extension, the bias is very small, indicating a close agreement between the two methods. Most of the data points for both flexion and extension fall within the limits of agreement, showing good agreement overall, but the wider range for flexion indicates more variability in the measurements.

Reliability

Inter-rater reliability

Good to excellent reliability was found for neck flexion; the mean ICC was 0.931, with a 95% CI ranging from 0.833 to 0.972 (*F(19,19) = 30.0, p<0.001*). Similarly, good to excellent reliability was observed for neck extension; the mean ICC was 0.896, with a 95% CI ranging from 0.760 to 0.957 (*F(19,19) = 17.9, p<0.001*).

Bland-Altman plots are presented in Figure 3A and 3B. For both flexion and extension, most of the data points fall within the limits of agreement, suggesting no significant mean biases in the inter-rater measurements of NOD for either flexion (−0.75 N) or extension (−0.48 N).

Intra-rater reliability

Moderate to excellent intra-rater reliability was observed for neck flexion measurements; the mean ICC was 0.844, with a 95% CI ranging from 0.626 to 0.937 (*F(19,19) = 13.8, p<0.001*). Good to excellent intra-rater reliability was found for neck extension; the mean ICC was 0.893, with a 95% CI ranging from 0.751 to 0.956 (*F(19,19) = 17.0, p<0.001*).

Bland-Altman plots are shown in Figure 3C and 3D. For flexion, most of the data points fell within the limits of agreement, but a higher mean bias of −1.38 N was found when comparing the NOD measurements of neck flexion in session 1 against session 3. For extension, most of the data points also fell within the limits of agreement, with no significant bias observed (mean bias = 0.22 N).

The ICC (95% CI), SEM, and MDC values for both inter-rater and intra-rater reliability for neck flexion and extension strength measurements are summarized in Table 3.

TABLE 3 - Reliability measures

Characteristic	Movement	ICC (95% CI)	SEM (N)	MDC (N)
Inter-rater reliability	Flexion	0.931 [0.833, 0.972]	1.549	4.292
	Extension	0.896 [0.760, 0.957]	2.037	5.646
Intra-rater reliability	Flexion	0.844 [0.626, 0.937]	2.204	6.110
	Extension	0.893 [0.751, 0.956]	2.118	5.869

Discussion

The NOD showed good to excellent agreement with the MCU for neck extension, but the agreement for flexion was lower, ranging from poor to good. Overall, the NOD proved to be a reliable instrument, especially for neck extension, though some methodological refinements may be needed to improve its accuracy for neck flexion strength measurements.

Concurrent validity

The agreement between the NOD and MCU for neck flexion strength was lower than expected, ranging from only poor to good. One explanation for this lies in the different measurement setups for each device. The MCU employs a more rigid setup with a metal head brace and multiple chest straps, providing greater stabilization of the head and trunk. In contrast, the NOD relies on the rater to provide static resistance, which is inherently more variable. Moreover, the lack of precise head positioning markers in the NOD setup may introduce small discrepancies in initial positioning of the dynamometer against the person's forehead. It is also possible that the MCU setup, by restricting movement, encourages participants to perform a combined movement of neck flexion and protraction, whereas the NOD, with its greater freedom of movement, allows participants to perform a more natural neck flexion. This difference in movement direction could potentially affect the direction of force applied and

© 2025 The Authors. Arch Physioter - ISSN 2057-0082 - www.archivesofphysiotherapy.com

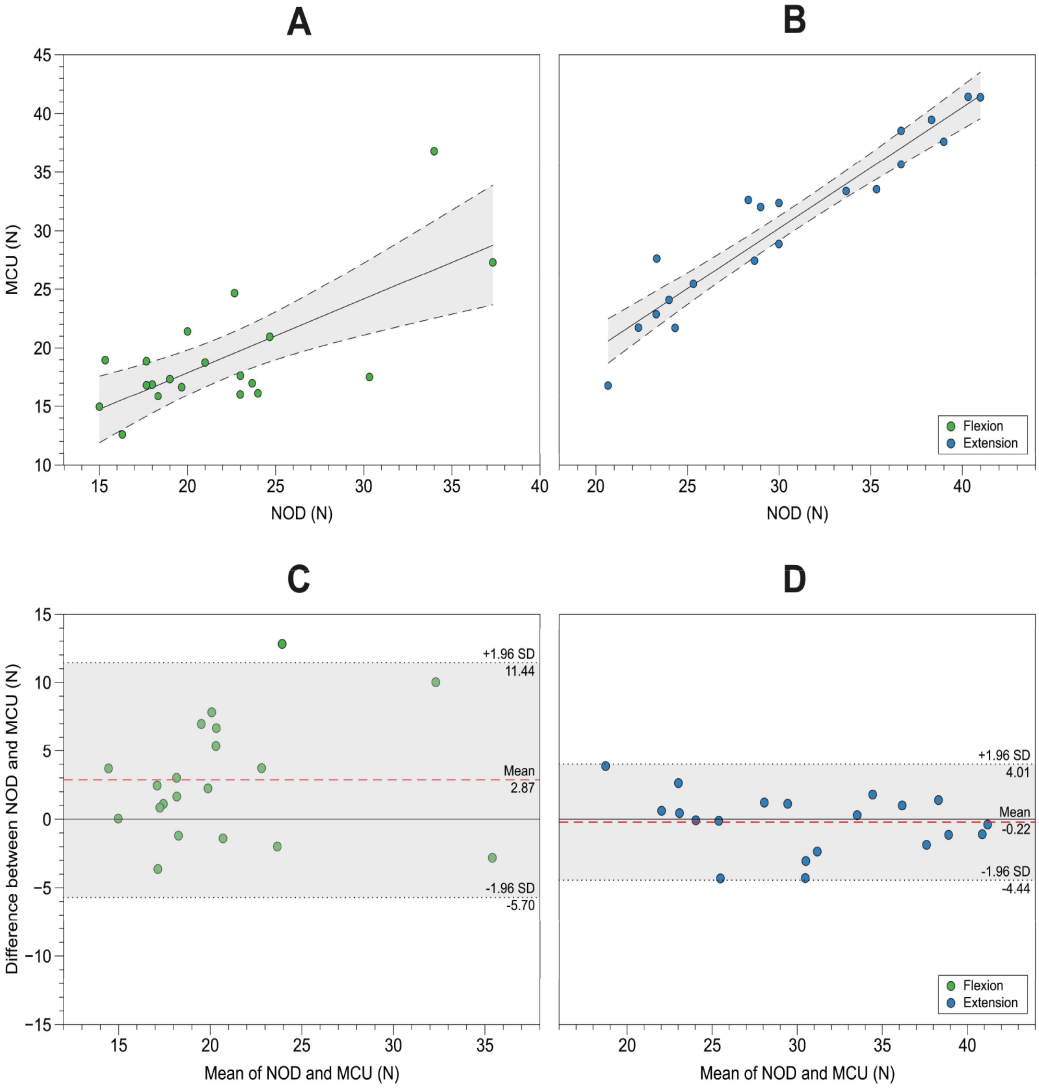


FIGURE 2 - Scatter plots with regression lines (A, B) and Bland-Altman plots (C, D) comparing NOD measurements with MCU for both neck flexion (A, C) and neck extension (B, D). In panels A and B, the x-axis represents the NOD measurements, and the y-axis represents the MCU measurements. The solid black lines indicate the regression lines, while the shaded areas represent the 95% CIs. These plots illustrate the relationship between the two methods, highlighting the strength and direction of their correlation for both flexion (A) and extension (B). In panels C and D, the x-axis represents the average of the measurements from the NOD and MCU methods, while the y-axis shows the difference between the NOD and MCU measurements. The red dashed line indicates the mean bias, reflecting the average difference between the two methods, and the black dotted lines represent the 95% limits of agreement, showing the range within which most differences are expected to fall. These plots assess the validity of the NOD method by comparing its measurements to the MCU for both flexion (C) and extension (D), highlighting any systematic bias and the level of agreement between the two methods.

contribute to the variability in force measurements. These findings are consistent with Ashall's et al., (2021) (24), who observed similar challenges when comparing measurements obtained with a HHD versus a wall-mounted dynamometer for neck flexion strength. In their study, the Pearson correlation for flexion strength was comparable to our results, with reduced agreement in flexion compared to extension. This suggests that the nature of handheld dynamometry introduces greater variability in flexion measurements due to the increased challenge in maintaining a stable, neutral head position. In addition, flexion movements tend to involve more compensatory muscle activity, especially if the head position deviates from neutral (42). These compensatory movements may lead to inconsistent force generation, resulting in lower agreement between the NOD and MCU.

In contrast, the NOD demonstrated good to excellent agreement with the MCU for neck extension strength. The rigidity provided by the seated posture and back support likely contributed to the higher consistency in extension measurements. Additionally, extension movements are generally less prone to compensatory trunk and head movements compared to flexion, resulting in more stable muscle recruitment patterns (43,44). Interestingly, these findings align with those of González-Rosalén et al., (2021) (45), who demonstrated that handheld dynamometry performs well for extension across different joints when forces remain below 200 N. Given that the neck extensors are less prone to fatigue due to their higher proportion of slow-twitch fibres (46), this could also explain the better agreement observed in extension compared to flexion, where fatigue may play a more significant role.

Reliability

Inter-rater reliability

The inter-rater reliability of the NOD was found to be good to excellent for both flexion and extension, consistent with the available literature on using HHDs to assess neck

© 2025 The Authors. Published by AboutScience - www.aboutscience.eu

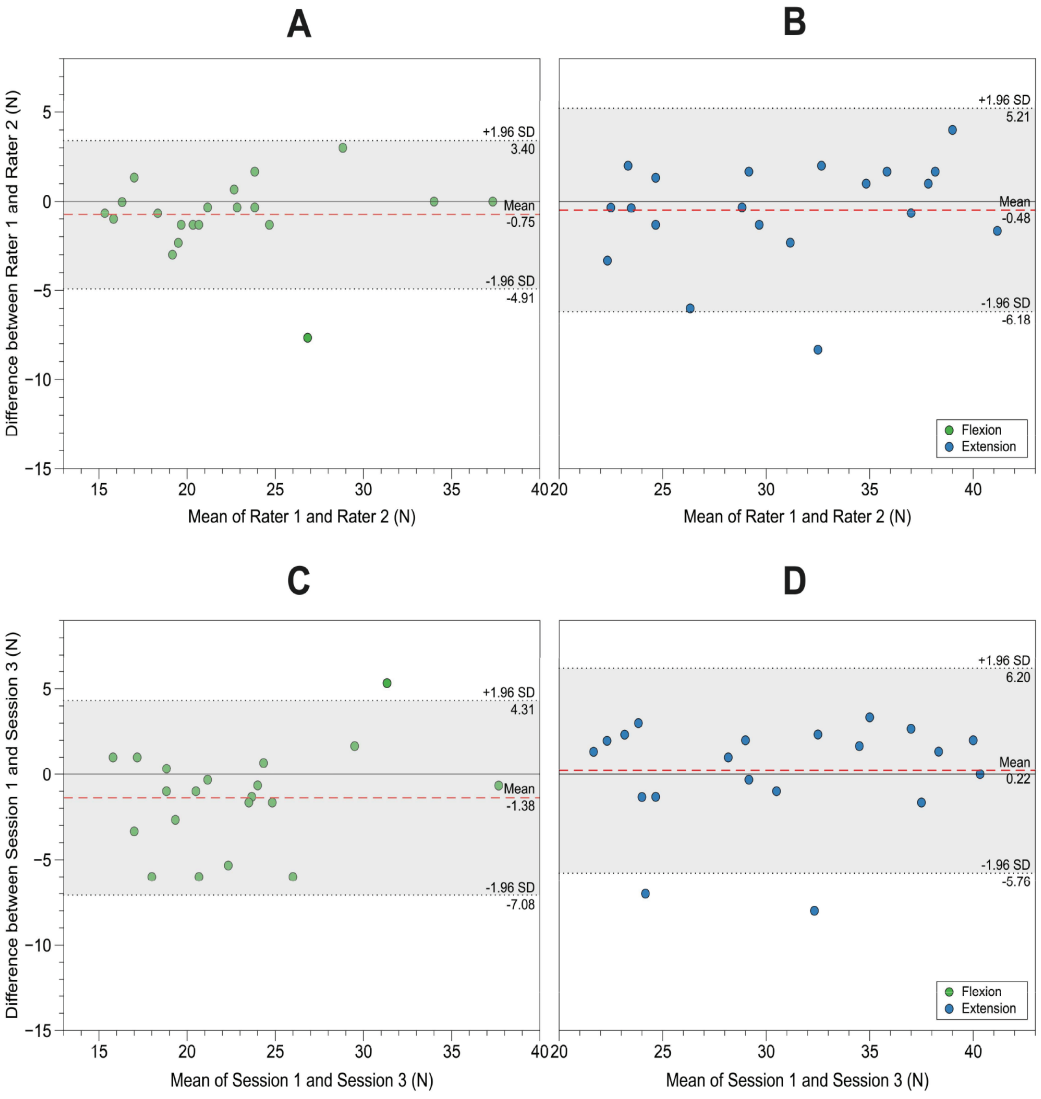


FIGURE 3 - Bland-Altman plots for inter-rater (A, B) and intra-rater (C, D) measurements of neck flexion and extension. In panels A and B, the x-axis represents the average of the two raters' measurements, while the y-axis indicates the difference between their measurements. The red dashed line represents the mean bias, showing the average difference between raters, and the black dotted lines indicate the upper and lower 95% limits of agreement, outlining the range within which most of the differences between the raters are expected to fall. In panels C and D, the x-axis represents the average of the measurements taken by the same rater across sessions 1 and 3, while the y-axis represents the difference between those repeated measurements. The red dashed line represents the mean bias, indicating the average difference between the repeated measurements by the same rater, and the black dotted lines mark the upper and lower 95% limits of agreement, showing the range within which most differences are expected to fall. Panels A and C correspond to neck flexion measurements, while panels B and D correspond to neck extension measurements.

muscle strength in a seated position. For example, Kubas et al., (2017) (29) reported similar reliability for the measurements of neck flexion and extension using an HHD. However, our study had a smaller 95% CI (0.760-0.957 vs 0.53-0.97), likely due to the larger sample size, which reduces the variability of the estimations. In contrast, the study by Kubas et al., (2021) (29) had a smaller sample (n = 10), which may have contributed to greater variability. Additionally, their study tested multiple directions (flexion, extension, side flexion, and rotation), possibly increasing muscle fatigue, which is known to affect force output (47). Our protocol limited the number of MVCs, likely reducing fatigue and contributing to more consistent reliability.

Intra-Rater Reliability

Moderate to excellent intra-rater reliability was observed for flexion using the NOD. This aligns with five previous studies examining the intra-rater reliability of HHD measurements for neck flexion in a seated position (24,27-29,33), all of which reported good reliability (ICC > 0.75). However, the 95%CIs in these studies varied substantially. For example, the study by Kubas et al., (2021) (29) reported a wide CI (0.08 to 0.96), while Vannebo et al., (2018) (33) reported a much narrower CI (0.92 to 0.97), indicating more consistent measurements. The differences in study design, including participant characteristics, time intervals between sessions, and the types of HHDs used, likely contributed to the variability in CIs. The study by Vannebo et al., (2018) (33) stands out for its higher intra-rater reliability, potentially due to participants being seated against a wall, which provided better proprioceptive input and trunk control, reducing compensatory movements and leading to more consistent measurements.

The intra-rater reliability of the NOD for neck extension strength was demonstrated to be good to excellent, aligning with four previous studies on neck extension strength assessed using HHDs with the participant in a seated position (24,27-29). Most of these studies reported similar

results, except for the study by Krause et al., (2019) (28), who reported a much wider CI range (0.21 to 0.82). This discrepancy could be due to over-exertion and muscular fatigue, as their study involved a higher number of submaximal holds and MVCs across multiple directions. In contrast, the current study required fewer MVCs, likely reducing fatigue and contributing to more consistent intra-rater measurements, as also mentioned for inter-rater reliability.

Systematic bias and learning Effect

A key finding in this study was the presence of systematic bias, particularly for the neck flexion measurements. Bland-Altman plots revealed a high bias in intra-rater flexion measurements, where force output increased by 6.27% from session 1 to session 3. This likely reflects a learning effect, where participants become more familiar with the testing procedure (33,48). The learning effect has been well documented in isometric dynamometry, where participants exhibit greater force production in later sessions due to improved technique and familiarity with the equipment (49). The absence of a similar bias in extension measurements may be due to the greater stability and consistency of extension movements as mentioned above. The back support provided during testing limits compensatory movements, ensuring that the neck extensors are more consistently recruited in the same manner across sessions.

Clinical implications

The NOD's portability and real-time feedback make it a valuable tool for clinical and research settings. However, the neck flexion results should be interpreted with some caution, as the lower agreement with the MCU suggests that variability in the testing setup may impact the accuracy of neck flexion strength measurements. Further refinement of the NOD's design or the development of standardized protocols for flexion testing could help mitigate this issue. Despite the challenges with neck flexion strength measurement, the NOD's excellent reliability and validity in extension suggest it is a reliable tool for tracking changes in neck strength over time, particularly in rehabilitation settings or in environments where portability is critical.

Methodological considerations

While this study demonstrated the validity and reliability of the NOD for the measurement of neck strength in a seated position, the sample was limited to young, asymptomatic individuals, which may restrict the generalisability of the findings to clinical populations with neck pain or older adults. The modest overall sample size, given that larger cohorts are typically advised to strengthen reliability estimates, and the relatively narrow range of demographics, particularly in terms of age and BMI, further limit the external validity of the results, as these factors are known to influence both muscle strength and measurement reliability (50). Inter- and intra-rater reliability were assessed with only one pair of raters, so variability introduced by different examiner combinations remains to be explored. Although the use of the NOD in the testing procedure adopted in this study could be considered to be reliant on the experience and physical strength of the examiner, we mitigated this by stabilizing the trunk with a strap, visually monitoring participants to ensure that they were primarily using their neck muscles with minimal trunk movement, and having both an experienced and a less experienced examiner perform the measurements, which increased our confidence in the reliability of the results. Lastly, the observed learning effect in flexion measurements suggests that further refinement of testing protocols may be needed to minimize systematic bias over multiple sessions.

Conclusions

This study provides preliminary evidence that the NOD is a reliable and valid tool for measuring neck muscle strength, particularly for neck extension. The good to excellent reliability across both inter-rater and intra-rater measurements confirms its consistency and potential suitability for clinical and research applications. Although its accuracy for neck flexion strength is influenced by some systematic bias and variability, the NOD remains a promising and reliable instrument. It may serve as a practical tool for frequent, reliable assessments in both clinical and research settings.

Acknowledgments

We would like to thank Dr. David Jiménez-Grande for his assistance with the experimental setup to extract force signals from the MCU and for developing the custom MATLAB script. We also extend our gratitude to all the participants of this study for their time and effort. This research was supported by the UK Space Agency [grant number ST/W003058/1].

Disclosures

Conflict of Interest: The authors declare no conflict of interest

Financial Support: This research was supported by the UK Space Agency [grant number ST/W003058/1].

Author Contributions: M.A., H.H.K.M., E.M.-V., M.B. and D.F. conceived and designed research; M.A. and H.H.K.M. performed experiments; M.A. and H.H.K.M. analysed data; M.A., H.H.K.M., E.M.-V., M.B. and D.F. interpreted results of experiments; M.A. and M.B. prepared figures: M.A. and H.H.K.M. drafted manuscript; M.A., H.H.K.M., E.M.-V., M.B. and D.F. edited and revised manuscript; M.A., H.H.K.M., E.M.-V., M.B. and D.F. approved final version of manuscript.

Data Availability Statement: The data supporting the findings of this study are provided in the supplementary material as an Excel file.

References

1. Enoka RM. Neuromechanics of human movement. Fifth edition ed. Champaign, IL: Human Kinetics; 2015. CrossRef
2. Versteegh T, Beaudet D, Greenbaum M, Hellyer L, Tritton A, Walton D. Evaluating the reliability of a novel neck-strength assessment protocol for healthy adults using self-generated resistance with a hand-held dynamometer. Physiother Can. 2015;67(1):58-64. CrossRef PubMed
3. Miranda IF, Wagner Neto ES, Dhein W, Brodt GA, Loss JF. Individuals With Chronic Neck Pain Have Lower Neck Strength Than Healthy Controls: A Systematic Review With

© 2025 The Authors. Published by AboutScience - www.aboutscience.eu

Meta-Analysis. J Manipulative Physiol Ther. 2019;42(8):608-622. CrossRef PubMed
4. Falla D, Lindstrøm R, Rechter L, Boudreau S, Petzke F. Effectiveness of an 8-week exercise programme on pain and specificity of neck muscle activity in patients with chronic neck pain: a randomized controlled study. Eur J Pain. 2013;17(10):1517-1528. CrossRef PubMed
5. Ylinen JJ, Häkkinen AH, Takala EP, et al. Effects of neck muscle training in women with chronic neck pain: one-year follow-up study. J Strength Cond Res. 2006;20(1):6-13. CrossRef PubMed
6. Alalawi A, Devecchi V, Gallina A, Luque-Suarez A, Falla D. Assessment of Neuromuscular and Psychological Function in People with Recurrent Neck Pain during a Period of Remission: Cross-Sectional and Longitudinal Analyses. J Clin Med. 2022;11(7):2042. CrossRef PubMed
7. Cuthbert SC, Goodheart GJ Jr. On the reliability and validity of manual muscle testing: a literature review. Chiropr Osteopat. 2007;15(1):4. CrossRef PubMed
8. Selistre LFA, Melo CS, Noronha MA. Reliability and Validity of Clinical Tests for Measuring Strength or Endurance of Cervical Muscles: A Systematic Review and Meta-analysis. Arch Phys Med Rehabil. 2021;102(6):1210-1227. CrossRef PubMed
9. Osternig LR. Isokinetic dynamometry: implications for muscle testing and rehabilitation. Exerc Sport Sci Rev. 1986;14:45-80. CrossRef PubMed
10. Arvanitidis M, Falla D, Sanderson A, Martinez-Valdes E. Does pain influence control of muscle force? A systematic review and meta-analysis. Eur J Pain. 29(2):e4716 CrossRef PubMed
11. Chamorro C, Armijo-Olivo S, De la Fuente C, Fuentes J, Javier Chirosa L. Absolute Reliability and Concurrent Validity of Hand Held Dynamometry and Isokinetic Dynamometry in the Hip, Knee and Ankle Joint: Systematic Review and Meta-analysis. Open Med (Wars). 2017;12(1):359-375. CrossRef PubMed
12. Contreras-Hernandez I, Arvanitidis M, Falla D, Negro F, Martinez-Valdes E. Achilles tendon morpho-mechanical parameters are related to triceps surae motor unit firing properties. J Neurophysiol. 2024;132(4):1198-1210. CrossRef PubMed
13. Lesnak JB, Anderson DT, Farmer BE, Katsavelis D, Grindstaff TL. Ability of Isokinetic Dynamometer to Predict Isotonic Knee Extension 1-Repetition Maximum. J Sport Rehabil. 2019;29(5):616-620. CrossRef PubMed
14. Sørensen L, Oestergaard LG, van Tulder M, Petersen AK. Measurement Properties of Isokinetic Dynamometry for Assessment of Shoulder Muscle Strength: A Systematic Review. Arch Phys Med Rehabil. 2021;102(3):510-520. CrossRef PubMed
15. Arvanitidis M, Bikinis N, Petrakis S, et al. Spatial distribution of lumbar erector spinae muscle activity in individuals with and without chronic low back pain during a dynamic isokinetic fatiguing task. Clin Biomech (Bristol). 2021;81:105214. CrossRef PubMed
16. Arvanitidis M, Jiménez-Grande D, Haouidji-Javaux N, Falla D, Martinez-Valdes E. People with chronic low back pain display spatial alterations in high-density surface EMG-torque oscillations. Sci Rep. 2022;12(1):15178. CrossRef PubMed
17. Arvanitidis M, Jiménez-Grande D, Haouidji-Javaux N, Falla D, Martinez-Valdes E. Eccentric exercise-induced delayed onset trunk muscle soreness alters high-density surface EMG-torque relationships and lumbar kinematics. Sci Rep. 2024;14(1):18589. CrossRef PubMed
18. Arvanitidis M, Jiménez-Grande D, Haouidji-Javaux N, Falla D, Martinez-Valdes E. Low Back Pain-Induced Dynamic Trunk Muscle Control Impairments Are Associated with Altered Spatial EMG-Torque Relationships. Med Sci Sports Exerc. 2024;56(2):193-208. CrossRef PubMed
19. Estrázulas JA, Estrázulas JA, de Jesus K, de Jesus K, da Silva RA, Libardoni Dos Santos JO. Evaluation isometric and isokinetic of trunk flexor and extensor muscles with isokinetic dynamometer: A systematic review. Phys Ther Sport. 2020;45:93-102. CrossRef PubMed
20. García-Vaquero MP, Barbado D, Juan-Recio C, López-Valenciano A, Vera-Garcia FJ. Isokinetic trunk flexion-extension protocol to assess trunk muscle strength and endurance: Reliability, learning effect, and sex differences. J Sport Health Sci. 2020;9(6):692-701. CrossRef PubMed
21. Reyes-Ferrada W, Chirosa-Rios L, Martinez-Garcia D, Rodríguez-Perea Á, Jerez-Mayorga D. Reliability of trunk strength measurements with an isokinetic dynamometer in non-specific low back pain patients: A systematic review. J Back Musculoskelet Rehabil. 2022;35(5):937-948. CrossRef PubMed
22. Chiu TT, Sing KL. Evaluation of cervical range of motion and isometric neck muscle strength: reliability and validity. Clin Rehabil. 2002;16(8):851-858. CrossRef PubMed
23. Althobaiti S, Falla D. Reliability and criterion validity of hand-held dynamometry for measuring trunk muscle strength in people with and without chronic non-specific low back pain. Musculoskelet Sci Pract. 2023;66:102799. CrossRef PubMed
24. Ashall A, Dobbin N, Thorpe C. The concurrent validity and intrarater reliability of a hand-held dynamometer for the assessment of neck strength in semi-professional rugby union players. Phys Ther Sport. 2021;49:229-235. CrossRef PubMed
25. Carnevalli APO, Bevilaqua-Grossi D, Oliveira AIS, Carvalho GF, Fernández-De-Las-Peñas C, Florencio LL. Intrarater and Interrater Reliability of Maximal Voluntary Neck Muscle Strength Assessment Using a Handheld Dynamometer in Women With Headache and Healthy Women. J Manipulative Physiol Ther. 2018;41(7):621-627. CrossRef PubMed
26. Cibulka MT, Herren J, Kilian A, Smith S, Mahmutovic F, Dolles C. The reliability of assessing sternocleidomastoid muscle length and strength in adults with and without mild neck pain. Physiother Theory Pract. 2017;33(4):323-330. CrossRef PubMed
27. Geary K, Green BS, Delahunt E. Intrarater reliability of neck strength measurement of rugby union players using a handheld dynamometer. J Manipulative Physiol Ther. 2013;36(7):444-449. CrossRef PubMed
28. Krause DA, Hansen KA, Hastreiter MJ, Kuhn TN, Peichel ML, Hollman JH. A Comparison of Various Cervical Muscle Strength Testing Methods Using a Handheld Dynamometer. Sports Health. 2019;11(1):59-63. CrossRef PubMed
29. Kubas C, Chen YW, Echeverri S, et al. Reliability and Validity of Cervical Range of Motion and Muscle Strength Testing. J Strength Cond Res. 2017;31(4):1087-1096. CrossRef PubMed
30. Martins F, Bento A, Silva AG. Within-Session and Between-Session Reliability, Construct Validity, and Comparison Between Individuals With and Without Neck Pain of Four Neck Muscle Tests. PM R. 2018;10(2):183-193. CrossRef PubMed
31. Shahidi B, Johnson CL, Curran-Everett D, Maluf KS. Reliability and group differences in quantitative cervicothoracic measures among individuals with and without chronic neck pain. BMC Musculoskelet Disord. 2012;13(1):215. CrossRef PubMed
32. Tudini F, Myers B, Bohannon R. Reliability and validity of measurements of cervical retraction strength obtained with a hand-held dynamometer. J Man Manip Ther. 2019;27(4):222-228. CrossRef PubMed
33. Vannebo KT, Iversen VM, Fimland MS, Mork PJ. Test-retest reliability of a handheld dynamometer for measurement of isometric cervical muscle strength. J Back Musculoskelet Rehabil. 2018;31(3):557-565. CrossRef PubMed

34. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96-106. CrossRef PubMed
35. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26(4):217-238. CrossRef PubMed
36. Chinn S. The assessment of methods of measurement. Stat Med. 1990;9(4):351-362. CrossRef PubMed
37. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods. 2009;41(4):1149-1160. CrossRef PubMed
38. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175-191. CrossRef PubMed
39. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-163. CrossRef PubMed
40. Furlan L, Sterr A. The Applicability of Standard Error of Measurement and Minimal Detectable Change to Motor Learning Research-A Behavioral Study. Front Hum Neurosci. 2018;12:95. CrossRef PubMed
41. Portney LG, Watkins MP. Foundations of Clinical Research: Applications To Practice. Pearson/Prentice Hall; 2015.
42. Lee DK, Moon DC, Hong KH. Effect of neck flexion restriction on sternocleidomastoid and abdominal muscle activity during curl-up exercises. J Phys Ther Sci. 2016;28(1):90-92. CrossRef PubMed
43. Bogduk N, Mercer S. Biomechanics of the cervical spine. I: normal kinematics. Clin Biomech (Bristol). 2000;15(9):633-648. CrossRef PubMed
44. Feipel V, Rondelet B, Le Pallec J, Rooze M. Normal global motion of the cervical spine: an electrogoniometric study. Clin Biomech (Bristol). 1999;14(7):462-470. CrossRef PubMed
45. González-Rosalén J, Benítez-Martínez JC, Medina-Mirapeix F, Cuerda-Del Pino A, Cervelló A, Martín-San Agustín R. Intra- and Inter-Rater Reliability of Strength Measurements Using a Pull Hand-Held Dynamometer Fixed to the Examiner's Body and Comparison with Push Dynamometry. Diagnostics (Basel). 2021;11(7):1230. CrossRef PubMed
46. Mannion AF, Dumas GA, Cooper RG, Espinosa FJ, Faris MW, Stevenson JM. Muscle fibre size and type distribution in thoracic and lumbar regions of erector spinae in healthy subjects without low back pain: normal values and sex differences. J Anat. 1997;190 (Pt 4)(Pt 4):505-13. CrossRef
47. Cvetko E, Karen P, Eržen I. Myosin heavy chain composition of the human sternocleidomastoid muscle. Ann Anat. 2012;194(5):467-472. CrossRef PubMed
48. Ylinen J, Ruuska J. Clinical use of neck isometric strength measurement in rehabilitation. Arch Phys Med Rehabil. 1994;75(4):465-469. CrossRef PubMed
49. Blazevich, A. J., Gill, N. D., Deans, N., & Zhou, S. (2007). Lack of human muscle architectural adaptation after short-term strength training. Muscle & nerve, 35(1), 78–86. CrossRef
50. Liu J, Liu P, Ma Z, et al. The effects of aging on the profile of the cervical spine. Medicine (Baltimore). 2019;98(7):e14425. CrossRef PubMed

© 2025 The Authors. Published by AboutScience - www.aboutscience.eu