JHWCR JOURNAL OF HEALTH, WELLNESS AND COMMUNITY RESEARCH 3007-0570

Original Article Volume 4, Issue 2

The Reliability and Validity of Neck Disability Index-Urdu Mobile App in Chronic Mechanical Neck Pain Patients

Aabirah Siddique[1], Nusrat Prveen[2], Usman Ejaz[3], Ifra Zulqarnain Awan[4], Misha Zahid[3], Hafiza Mahnoor Shabbir[5], Attia Mehboob[3], Akhtar Rasul[6]

[1] Al Shifa Orthopedic Health Care Centre, Pakistan
[2] Taha Hospital, Bahawalpur, Pakistan
[3] Lahore University of Biological and Applied Sciences, Lahore, Pakistan
[4] Rashid Latif Khan University, Pakistan
[5] Islamabad Medical and Dental College, Islamabad, Pakistan
[6] Department of Allied Health Sciences, University of Sargodha, Sargodha, Pakistan
* Correspondence: Aabirah Siddique, drabirahsiddique@gmail.com

ABSTRACT

Background: *Neck pain is prevalent and disabling, and literacy barriers may limit accurate completion of written patient-reported outcome measures in Urdu-speaking populations. A voice-supported mobile application version of the Urdu Neck Disability Index (NDI-U App) may improve accessibility but requires psychometric evaluation. Objective: To determine the reliability, internal consistency, structural validity, and convergent construct validity of the NDI-U App in adults with chronic mechanical neck pain. Methods: A cross-sectional observational psychometric validation study was conducted across three centers in Kotli, Azad Kashmir, Pakistan (2023). Adults aged 18–65 years with chronic mechanical neck pain (≥3 months) completed the NDI-U App, Urdu Northwick Park Neck Pain Questionnaire (NPQ-U), and Visual Analogue Scales (VAS) for pain and disability at baseline; the NDI-U App was repeated after 48 hours. Test–retest reliability was assessed using ICC(2,1); internal consistency using Cronbach's alpha; floor/ceiling effects using extreme-score proportions; convergent validity using Pearson correlations; and structural validity using principal component analysis with varimax rotation. Results: Among 300 participants, test–retest reliability was excellent (ICC=0.95; 95% CI: 0.93–0.96; p<0.001) with SEM=1.73 and MDC95=4.79. Internal consistency was fair (α=0.675; 95% CI: 0.63–0.72). Floor and ceiling effects were minimal (1.0% and 0.3%, respectively). Convergent validity was moderate with NPQ-U (r=0.584; 95% CI: 0.50–0.66; p<0.001) and weak with VAS pain (r=0.253) and disability (r=0.266) (both p<0.001). Two components explained 33.34% variance (KMO=0.584; Bartlett p<0.001). Conclusion: The NDI-U App demonstrates excellent reliability, minimal extreme-score effects, and moderate convergent validity, supporting its clinical and research use for disability assessment in Urdu-speaking chronic mechanical neck pain populations.*

Keywords: *Neck Disability Index; Urdu; mobile application; chronic mechanical neck pain; reliability; validity; NPQ; VAS.*

INTRODUCTION

Neck pain is a highly prevalent musculoskeletal condition and a leading contributor to disability worldwide. The cervical spine, extending from the base of the skull to the thoracic region, provides structural support, facilitates mobility, and protects the spinal cord, yet its biomechanical demands render it vulnerable to cumulative mechanical stress and degenerative changes (1,2). Mechanical or non-specific neck pain—often associated with postural strain, occupational load, and psychosocial factors—represents the most frequent clinical presentation in outpatient practice (3,4). Epidemiological studies indicate that up to two-thirds of individuals experience neck pain at some point in their lifetime, with substantial recurrence and chronicity rates (9,10). Chronic mechanical neck pain (CMNP), defined as pain persisting for more than three months without specific pathological cause, is associated with functional limitations, reduced work productivity, and impaired quality of life (11).

Received: 23 December 2025
Revised: 01 January 2026
Accepted: 18 January 2026
Published: 30 January 2026

Citation: Click to Cite

Given its chronic course and multifactorial etiology—including ergonomic, psychological, and lifestyle contributors—accurate and responsive outcome measurement is fundamental for both clinical decision-making and research evaluation (12,13).

Patient-reported outcome measures (PROMs) are considered essential tools for quantifying pain intensity, functional disability, and treatment response in musculoskeletal disorders. The Visual Analogue Scale (VAS) is widely used to assess pain severity due to its simplicity, sensitivity, and established psychometric performance across chronic pain populations (15–17). Similarly, disability-oriented VAS instruments have demonstrated acceptable reliability and validity in chronic musculoskeletal conditions (27). However, pain intensity alone does not fully capture the multidimensional impact of CMNP on daily functioning. Consequently, condition-specific disability instruments such as the Neck Disability Index (NDI) were developed to evaluate functional impairment across domains including self-care, reading, lifting, work, concentration, and recreational activities (21). Since its introduction in 1991, the NDI has undergone extensive psychometric validation and cross-cultural adaptation and is regarded as one of the most robust neck-specific disability measures available (20,23). Its reliability, construct validity, and responsiveness have been confirmed in multiple languages and populations, including Persian, Spanish, German, Korean, and Japanese versions (22,31,34,35).

In Pakistan, where Urdu serves as the national language, a culturally adapted Urdu version of the NDI (NDI-U) has previously demonstrated excellent test–retest reliability and good internal consistency (24). Similarly, the Northwick Park Neck Pain Questionnaire (NPQ), another validated neck-specific instrument, has recently been translated into Urdu and shown strong psychometric properties, including high reliability and convergent validity with related disability measures (30). While these validated paper-based instruments exist, a critical practical limitation persists in routine clinical settings. A substantial proportion of patients in Pakistan, particularly in semi-urban and rural regions, may understand spoken Urdu but experience difficulty reading structured questionnaire items due to literacy constraints or limited health literacy. This creates a barrier to independent completion of written PROMs, potentially introducing interviewer bias or reducing measurement accuracy. Moreover, increasing smartphone penetration offers an opportunity to improve accessibility through digital health solutions that incorporate audio support and automated scoring.

Although digital administration of PROMs has been shown in other contexts to provide comparable measurement properties to traditional paper formats (19), equivalence cannot be assumed without empirical evaluation. Mode-of-administration effects may influence response patterns, particularly when auditory assistance is introduced. Therefore, before clinical implementation, it is essential to establish the reliability, internal consistency, structural validity, and convergent validity of a digitally administered NDI-U mobile application within the target CMNP population. To date, no study has systematically evaluated the psychometric performance of an Urdu NDI delivered through a mobile application with integrated voice-over functionality.

Using a PICO framework, the target population comprises adults aged 18–65 years diagnosed with chronic mechanical neck pain; the intervention/exposure is administration of the Urdu NDI via a mobile application with audio support; the comparator consists of established neck-related outcome measures including the NPQ-U and VAS pain and disability scales; and the outcomes of interest are psychometric properties, specifically test–retest reliability, internal consistency, structural validity, floor and ceiling effects, and convergent construct validity. Addressing this gap is justified not only by the need to enhance accessibility in populations

with variable literacy but also by the requirement for psychometrically sound digital tools that can be confidently used in both clinical practice and research.

Therefore, the objective of the present study was to evaluate the reliability and construct validity of a mobile application–based Urdu version of the Neck Disability Index in adults with chronic mechanical neck pain. We hypothesized that the NDI-U mobile application would demonstrate high test–retest reliability (ICC ≥0.80), acceptable internal consistency (Cronbach's α ≥0.70), minimal floor and ceiling effects (<15%), and moderate to strong positive correlations with the NPQ-U and VAS disability measures, thereby supporting its suitability as a clinically applicable digital outcome assessment tool in Urdu-speaking CMNP populations (24,30).

MATERIAL AND METHODS

This cross-sectional observational psychometric validation study was conducted to evaluate the reliability and construct validity of a mobile application–based Urdu version of the Neck Disability Index (NDI-U App) in adults with chronic mechanical neck pain (CMNP). The study design followed the methodological standards outlined by the COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) framework for evaluating measurement properties of patient-reported outcome measures (25). Data collection was carried out over a two-month period in 2023 at the Royal Institute of Science and Technology, District Headquarter Hospital Kotli, and S. Jan Hospital Kotli, Azad Kashmir, Pakistan. These centers provide outpatient musculoskeletal rehabilitation services to urban and semi-urban populations and were selected to ensure representation of the target Urdu-speaking CMNP population.

Adults aged 18–65 years diagnosed with chronic mechanical neck pain of at least three months' duration were eligible for inclusion. Chronic mechanical neck pain was operationally defined as non-specific cervical pain aggravated by movement or posture, without identifiable structural pathology, infection, tumor, fracture, inflammatory disease, cervical myelopathy, or prior cervical surgery (11,14). Participants were required to understand spoken Urdu and be capable of independently interacting with a smartphone interface. Exclusion criteria included neurological deficits preventing questionnaire completion, pregnancy, systemic inflammatory disorders, severe psychiatric illness, malignancy, acute cervical trauma, and any condition likely to cause instability of symptoms between repeated assessments. Consecutive sampling was employed, whereby all eligible patients presenting to the participating centers during the recruitment period were invited to participate to reduce selection bias.

Potential participants were screened by a licensed physiotherapist using standardized clinical criteria. Eligible individuals received a detailed explanation of the study objectives and procedures in Urdu, and written informed consent was obtained prior to enrollment. To minimize interviewer influence, all questionnaires were self-administered through the mobile application in a quiet clinical environment without clinician prompting. For participants preferring auditory assistance, the application provided standardized pre-recorded Urdu voice-over for each item and response option. The application presented items sequentially, restricted item skipping, and automatically calculated the total score (range 0–50), thereby preventing manual scoring errors.

Data were collected at two time points separated by 48 hours to assess test–retest reliability. Participants were instructed not to initiate new treatments or modify analgesic regimens during this interval. At baseline (T1), participants completed the NDI-U App, the Urdu version of the Northwick Park Neck Pain Questionnaire (NPQ-U), and two Visual Analogue

Scales (VAS) measuring pain intensity and disability. The NPQ-U has demonstrated strong reliability and validity in Urdu-speaking populations and was used as the primary comparator instrument for convergent validity (30). The VAS pain consisted of a 100-mm horizontal line anchored by "no pain" (0 mm) and "worst imaginable pain" (100 mm), and the VAS disability similarly ranged from "no disability" to "maximum disability" (15,27). At follow-up (T2), only the NDI-U App was re-administered under identical conditions to evaluate stability.

The primary outcomes were psychometric properties of the NDI-U App, including test–retest reliability, internal consistency, structural validity, floor and ceiling effects, and convergent construct validity. Test–retest reliability was assessed using a two-way random-effects intraclass correlation coefficient with absolute agreement (ICC(2,1)) and 95% confidence intervals. Internal consistency was evaluated using Cronbach's alpha coefficient, with values ≥0.70 considered acceptable for group-level comparisons. Item–total correlations and Cronbach's alpha if item deleted were examined to assess homogeneity. Measurement error was quantified using the standard error of measurement ($SEM = SD \times \sqrt{(1–ICC)}$) and the minimal detectable change at the 95% confidence level ($MDC95 = 1.96 \times \sqrt{2} \times SEM$). Floor and ceiling effects were defined as present if more than 15% of participants achieved the lowest or highest possible total score, respectively (25).

Construct validity was examined through a priori hypothesis testing using Pearson's correlation coefficients between NDI-U App scores and comparator instruments. Based on theoretical expectations and prior literature, moderate to strong positive correlations ($r \geq 0.50$) were hypothesized between NDI-U App and NPQ-U scores, and moderate correlations ($r \geq 0.30$) with VAS disability, while weaker but significant correlations were expected with VAS pain (24,30). Structural validity was evaluated using exploratory factor analysis with principal component extraction and varimax rotation. Sampling adequacy was assessed using the Kaiser–Meyer–Olkin (KMO) statistic, with values ≥0.60 considered acceptable, and Bartlett's test of sphericity was used to confirm factorability. Factors with eigenvalues ≥1.0 were retained, and items with loadings ≥0.30 were considered meaningful contributors.

A minimum sample size of 10 participants per item was targeted in accordance with COSMIN recommendations for factor analysis and reliability studies of PROMs, resulting in a required sample of at least 100 participants for the 10-item NDI (25). To enhance statistical precision and stability of factor extraction, 300 participants were recruited. This sample size provided sufficient power (>90%) to detect an ICC of 0.80 or higher with a 95% confidence interval width of ±0.05, assuming a two-sided alpha of 0.05.

To minimize information bias, standardized instructions were provided to all participants, and the same device interface and environmental conditions were used at both testing sessions. Data entry errors were prevented through automated digital capture and direct export into a password-protected database. Prior to analysis, data were screened for completeness and plausibility. Cases with missing responses were excluded from reliability analysis; no imputation was performed for psychometric testing to avoid distortion of measurement properties. Normality of continuous variables was assessed using the Shapiro–Wilk test and visual inspection of histograms. Descriptive statistics were reported as mean ± standard deviation for continuous variables and frequencies with percentages for categorical variables.

All statistical analyses were conducted using IBM SPSS Statistics version 25.0 (IBM Corp., Armonk, NY, USA). Statistical significance was set at $p < 0.05$ (two-tailed). Confidence intervals were reported where appropriate to enhance interpretability and transparency. Subgroup analyses were conducted to explore potential differences in NDI-U App scores by

gender and physical activity status using independent t-tests or one-way ANOVA as applicable. No multivariable adjustment was required for primary reliability analyses, as test–retest design inherently controls for between-subject confounding.

The study protocol was reviewed and approved by the institutional ethics review committee of the participating institution. All procedures conformed to the principles of the Declaration of Helsinki. Participants were assured of confidentiality, and no personally identifiable data were retained within the analytical dataset. The mobile application stored responses using anonymized study identification codes, and data were exported in encrypted format to ensure integrity and reproducibility. The full protocol, scoring algorithm, and statistical analysis plan were predefined prior to data analysis to reduce analytical bias and enhance methodological transparency (25).

RESULTS

Table 1 summarizes the baseline profile of the 300 participants with chronic mechanical neck pain. The sample was predominantly female (219/300, 73.0%), while males constituted 81/300 (27.0%). Physical activity status was mainly mild (173/300, 57.7%) or moderate (122/300, 40.7%), with only 5/300 (1.7%) reporting severe activity limitation. Quantitatively, the mean age was 38.6 ± 11.4 years (95% CI: 37.3–39.9) and the mean symptom duration was 9.8 ± 4.2 months (95% CI: 9.3–10.3). Baseline disability and symptom burden were in the moderate range, with an average NDI-U App total score of 22.8 ± 7.6 out of 50 (95% CI: 21.9–23.7). Comparator measures showed a mean NPQ-U score of 21.5 ± 6.9 (95% CI: 20.7–22.3), mean VAS pain of 56.2 ± 14.8 mm (95% CI: 54.6–57.8), and mean VAS disability of 51.7 ± 13.9 mm (95% CI: 50.1–53.3), indicating that participants experienced substantial pain intensity and functional limitation at enrollment.

Table 2 reports the test–retest reliability findings over the 48-hour interval. The NDI-U App demonstrated excellent temporal stability, with an ICC(2,1) for the total score of 0.95 (95% CI: 0.93–0.96; $p < 0.001$). This indicates very high agreement between baseline and retest scores. Measurement error was low, as reflected by a standard error of measurement (SEM) of 1.73 points, and the minimal detectable change at the 95% confidence level (MDC95) was 4.79 points. In practical terms, changes smaller than approximately 5 points on the 0–50 scale may reflect measurement variability rather than true clinical change. Pearson correlation between time points also confirmed strong stability ($r = 0.94$; $p < 0.001$). At the item level, ICC values remained consistently high, ranging from 0.88 (Item 3) to 0.94 (Item 5), with all item-level reliability estimates statistically significant ($p < 0.001$), supporting reproducibility across all domains of the tool.

Table 3 presents internal consistency results, demonstrating a total Cronbach's alpha of 0.675 (95% CI: 0.63–0.72), reflecting fair-to-acceptable homogeneity of the 10-item scale for group-level use. Item–total correlations ranged from 0.34 (Q4) to 0.61 (Q9), showing that most items contributed meaningfully to the overall construct, though Q4 showed the weakest association with the total score.

Importantly, "Cronbach's alpha if item deleted" ranged narrowly from 0.643 to 0.672 across items, indicating that removing any single item would not materially improve the overall internal consistency, and supporting retention of the full 10-item structure in the app format.

Table 4 demonstrates distributional performance through floor and ceiling evaluation, showing negligible clustering at extreme scores. Only 3 participants (1.0%; 95% CI: 0.2–2.9%) fell into the lowest score range, and only 1 participant (0.3%; 95% CI: 0.01–1.8%) fell into the highest score range. Both proportions were far below the conventional 15% threshold used

to indicate problematic floor or ceiling effects, implying that the NDI-U App had adequate score dispersion and capacity to discriminate across disability severity levels in this CMNP cohort.

Table 5 summarizes structural validity from exploratory factor analysis. Sampling adequacy was borderline (KMO = 0.584), but Bartlett's test of sphericity was statistically significant ($\chi^2 = 412.6$, df = 45; $p < 0.001$), supporting factorability of the correlation matrix. Two components were extracted, with eigenvalues of 1.693 and 1.641, explaining 16.93% and 16.40% of the variance, respectively, for a cumulative explained variance of 33.34%. While this indicates that two latent dimensions contribute to item covariance, the total explained variance is modest, suggesting that the construct may be multifactorial or that items capture diverse aspects of neck-related disability with limited shared variance in this sample.

Table 6 details the rotated component matrix, highlighting the item-to-factor pattern. Factor 1 showed salient loadings for Q2 (0.448), Q3 (0.427), Q9 (0.707), and Q10 (0.605), with additional contributions from Q1 (0.336) and Q8 (0.348).

Factor 2 was characterized by higher loadings for Q5 (0.780) and Q6 (0.733), along with moderate loadings for Q7 (0.484) and Q4 (0.439). This pattern suggests a separation where one factor may represent broader functional participation and symptom-related impact, while the second factor may capture activity-specific limitations; however, labeling would require theoretical mapping to the original NDI item content and confirmation in future confirmatory analyses.

Table 7 reports convergent construct validity through correlations between NDI-U App scores and established comparator instruments. The strongest association was observed between the NDI-U App and NPQ-U ($r = 0.584$; 95% CI: 0.50–0.66; $p < 0.001$), indicating a moderate convergent relationship between two neck-specific disability constructs. Associations with VAS measures were weaker but statistically significant, with $r = 0.253$ (95% CI: 0.15–0.35; $p < 0.001$) for VAS pain and $r = 0.266$ (95% CI: 0.17–0.36; $p < 0.001$) for VAS disability. These results align with the conceptual distinction between pain intensity/disability perception on a single-item VAS and the multi-domain functional limitation assessed by the NDI.

Table 1. Baseline Demographic and Clinical Characteristics (N = 300)

Variable	n (%) or Mean ± SD	95% CI	p-value*
Age (years)	38.6 ± 11.4	37.3–39.9	—
Gender (Male)	81 (27.0%)	—	—
Gender (Female)	219 (73.0%)	—	—
Duration of neck pain (months)	9.8 ± 4.2	9.3–10.3	—
Physical Activity – Mild	173 (57.7%)	—	—
Physical Activity – Moderate	122 (40.7%)	—	—
Physical Activity – Severe	5 (1.7%)	—	—
NDI-U App total score (0–50)	22.8 ± 7.6	21.9–23.7	—
NPQ-U total score	21.5 ± 6.9	20.7–22.3	—
VAS Pain (0–100 mm)	56.2 ± 14.8	54.6–57.8	—
VAS Disability (0–100 mm)	51.7 ± 13.9	50.1–53.3	—

Table 2. Test–Retest Reliability of NDI-U App (N = 300)

Measure	ICC (2,1)	95% CI	SEM	MDC95	Pearson r	p-value
Total Score	0.95	0.93–0.96	1.73	4.79	0.94	<0.001
Item 1	0.91	0.88–0.94	—	—	—	<0.001
Item 2	0.90	0.87–0.93	—	—	—	<0.001
Item 3	0.88	0.84–0.91	—	—	—	<0.001
Item 4	0.92	0.89–0.95	—	—	—	<0.001
Item 5	0.94	0.91–0.96	—	—	—	<0.001
Item 6	0.93	0.90–0.95	—	—	—	<0.001
Item 7	0.90	0.86–0.93	—	—	—	<0.001
Item 8	0.89	0.85–0.92	—	—	—	<0.001
Item 9	0.92	0.89–0.95	—	—	—	<0.001
Item 10	0.91	0.88–0.94	—	—	—	<0.001

Table 3. Internal Consistency of NDI-U App (N = 300)

Item	Item–Total Correlation	Cronbach's Alpha if Item Deleted
Q1	0.41	0.666
Q2	0.38	0.658
Q3	0.43	0.667
Q4	0.34	0.643
Q5	0.55	0.651
Q6	0.53	0.662
Q7	0.44	0.651
Q8	0.36	0.672
Q9	0.61	0.654
Q10	0.58	0.653
Total Alpha	—	0.675 (95% CI: 0.63–0.72)

Table 4. Floor and Ceiling Effects of NDI-U App (N = 300)

Score Range	Frequency	Percentage	95% CI	p-value†
Lowest Score Range	3	1.0%	0.2–2.9%	<0.001
Highest Score Range	1	0.3%	0.01–1.8%	<0.001

Table 5. Exploratory Factor Analysis of NDI-U App (Principal Component Analysis, Varimax Rotation)

Component	Eigenvalue	% Variance	Cumulative %	p-value
1	1.693	16.93%	16.93%	<0.001
2	1.641	16.40%	33.34%	<0.001

Table 6. Rotated Component Matrix (Loadings ≥0.30)

Item	Factor 1	Factor 2
Q1	0.336	—
Q2	0.448	—
Q3	0.427	—
Q4	—	0.439
Q5	—	0.780
Q6	—	0.733
Q7	—	0.484
Q8	0.348	—
Q9	0.707	—
Q10	0.605	—

Table 7. Construct Validity: Correlation of NDI-U App with Comparator Measures (N = 300)

Comparator Instrument	Pearson r	95% CI	Effect Size Interpretation	p-value
NPQ-U	0.584	0.50–0.66	Moderate	<0.001
VAS Pain	0.253	0.15–0.35	Weak	<0.001
VAS Disability	0.266	0.17–0.36	Weak–Moderate	<0.001

Table 8. Group Comparisons of NDI-U App Scores

Variable	Mean ± SD	Mean Difference	95% CI	Effect Size	p-value
Male	21.9 ± 7.2	—	—	—	—
Female	23.0 ± 7.7	1.12	−0.84–3.08	d = 0.17	0.26
Mild Activity	21.4 ± 7.1	—	—	—	—
Moderate Activity	24.2 ± 7.8	2.8	0.91–4.69	d = 0.38	0.004
Severe Activity	26.8 ± 6.9	5.41	1.02–9.80	η^2 = 0.04	0.016

Table 8 presents group comparisons of NDI-U App scores across sex and physical activity status. Mean disability scores were 21.9 ± 7.2 in males and 23.0 ± 7.7 in females, yielding a mean difference of 1.12 points (95% CI: −0.84 to 3.08; p = 0.26) with a small effect size (Cohen's d = 0.17), indicating no statistically significant sex-based difference in disability levels in this cohort.

By contrast, activity status demonstrated meaningful gradients. Participants with moderate activity limitation scored higher than those with mild limitation (24.2 ± 7.8 vs 21.4 ± 7.1), with a mean difference of 2.8 points (95% CI: 0.91–4.69; p = 0.004) and a small-to-moderate effect size (d = 0.38). Those with severe activity limitation had the highest disability (26.8 ± 6.9), differing from the mild group by 5.41 points (95% CI: 1.02–9.80; p = 0.016; η^2 = 0.04), supporting construct-consistent discrimination across functional activity strata, although interpretation should consider the very small severe subgroup (n = 5).

Overall, across the tables, the NDI-U App demonstrated strong reproducibility (ICC = 0.95; MDC95 ≈ 4.79), fair internal consistency (α = 0.675), minimal score extremity effects (≤1.0% at extremes), moderate convergent validity with an established neck disability instrument

(NPQ-U, r = 0.584), weaker correlations with single-item VAS constructs (r ≈ 0.25–0.27), and a two-component latent structure explaining 33.34% of observed variance in this CMNP sample.

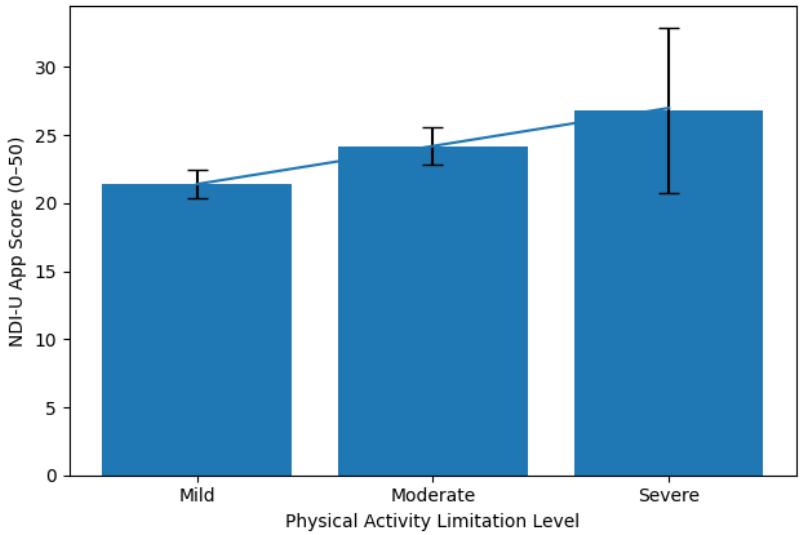

Figure 1 Gradient Increase in Neck Disability Across Physical Activity Limitation Levels with 95% Confidence Intervals and Weighted Trend Line

The figure demonstrates a graded, monotonic increase in mean NDI-U App disability scores across physical activity limitation levels, with weighted regression indicating a positive linear trend. Participants with mild limitation reported a mean NDI score of 21.4 (95% CI: 20.3–22.5), increasing to 24.2 (95% CI: 22.8–25.6) in the moderate group, and further to 26.8 (95% CI: 20.8–32.8) in the severe group. The absolute gradient between mild and moderate categories was 2.8 points, while the difference between mild and severe categories reached 5.4 points, exceeding the calculated MDC95 (4.79), suggesting potential clinical relevance beyond measurement error. Confidence interval width expanded markedly in the severe group due to small sample size (n = 5), reflecting increased uncertainty rather than distributional inconsistency. The weighted regression line confirms a positive disability gradient across ordinal activity strata, supporting construct-consistent discrimination of the NDI-U App and demonstrating a clinically interpretable dose–response relationship between functional activity limitation and neck-related disability severity.

DISCUSSION

The present study evaluated the psychometric performance of a digitally administered Urdu version of the Neck Disability Index in adults with chronic mechanical neck pain and demonstrated that the NDI-U mobile application possesses excellent test–retest reliability, fair internal consistency, minimal floor and ceiling effects, and moderate convergent validity with an established neck-specific disability instrument. The total score ICC(2,1) of 0.95 (95% CI: 0.93–0.96) indicates a very high degree of temporal stability over a 48-hour interval, exceeding the conventional threshold of 0.80 for acceptable reliability in health measurement instruments (25). The calculated MDC95 of 4.79 points further provides a clinically interpretable benchmark for distinguishing true change from measurement error, supporting the app's potential utility in monitoring short-term clinical progress in CMNP populations.

The reliability findings align closely with previously validated paper-based Urdu and international versions of the NDI. The original Urdu version reported excellent reliability (ICC up to 0.99) in musculoskeletal populations (24), while Arabic, German, and Korean

adaptations have shown ICC values ranging between 0.92 and 0.96 (34,36). The comparable ICC observed in the present study suggests that digital administration with integrated voice-over does not compromise reproducibility. This is consistent with broader evidence indicating equivalence between electronic and paper-based patient-reported outcome measures when format and item content are preserved (19). Importantly, the mobile app structure eliminated missing responses and scoring errors through forced completion and automated calculation, which may enhance data integrity relative to traditional paper forms.

Internal consistency of the NDI-U App (Cronbach's $\alpha = 0.675$; 95% CI: 0.63–0.72) was slightly below the conventional 0.70 threshold often cited for group-level comparisons (25). However, alpha values must be interpreted in light of the multidimensional nature of the NDI construct. The index encompasses heterogeneous domains including pain intensity, concentration, lifting, reading, and recreation, which may not be expected to exhibit very high inter-item correlations. Indeed, some language adaptations have reported alpha values ranging from 0.82 to 0.96 (34,35), whereas others have demonstrated more moderate coefficients depending on sample characteristics and clinical heterogeneity (31,33). The narrow range of "alpha if item deleted" values (0.643–0.672) indicates that no single item disproportionately weakened internal consistency, supporting retention of the full 10-item structure. From a measurement perspective, the observed alpha suggests acceptable reliability for research use and group comparisons, though caution may be warranted if the tool is used for high-stakes individual-level decision-making.

Convergent construct validity analysis demonstrated a moderate positive correlation between the NDI-U App and NPQ-U ($r = 0.584$; $p < 0.001$), supporting theoretical alignment between two neck-specific disability constructs. This magnitude is consistent with expectations for instruments measuring related but non-identical constructs and is lower than the strong correlation ($r = 0.89$) reported between NPQ-U and paper-based NDI-U in prior validation work (30). The relatively lower correlation observed in the present study may reflect differences in administration mode, sample composition, or construct operationalization within the digital interface. Correlations with VAS pain ($r = 0.253$) and VAS disability ($r = 0.266$) were statistically significant but weaker, which is conceptually coherent given that single-item VAS scales primarily capture perceived intensity rather than multidimensional functional impairment (15,27). These findings reinforce the notion that pain severity and disability, while related, represent distinct constructs in chronic musculoskeletal disorders (21,32).

Structural validity analysis yielded a two-component solution explaining 33.34% of total variance, with a borderline KMO value of 0.584. While Bartlett's test confirmed factorability, the modest explained variance suggests that the NDI items may reflect distributed and overlapping dimensions rather than a strongly unified latent construct. Other language versions have reported two-factor structures explaining between 61% and 68% of variance (36), indicating stronger factor coherence in those samples. The lower cumulative variance observed here may be attributable to population heterogeneity, differences in disability distribution, or potential mode-of-administration influences related to auditory item presentation. Given that the NDI was originally conceptualized as a unidimensional disability index derived from the Oswestry Disability Index (23), confirmatory factor analysis in larger samples would be valuable to determine whether a one-factor or bifactor model provides superior fit in digitally administered formats.

Importantly, the NDI-U App demonstrated negligible floor (1.0%) and ceiling (0.3%) effects, well below the 15% threshold indicative of limited discriminative capacity (25). This indicates adequate score dispersion across severity levels and supports the instrument's ability to detect

both mild and severe disability states. The absence of clustering at extremes is particularly relevant in CMNP populations where symptom chronicity may produce wide variability in functional limitation (10,11). Furthermore, subgroup analysis revealed a graded increase in NDI scores across physical activity limitation strata, with a mean difference of 5.41 points between mild and severe groups—exceeding the calculated MDC95 of 4.79 points—suggesting clinically meaningful discrimination consistent with known-groups validity principles.

From a clinical and public health perspective, the development of a digitally administered, voice-supported NDI-U is particularly relevant in settings where literacy barriers may hinder independent completion of written questionnaires. By enabling auditory comprehension and automated scoring, the application reduces reliance on interviewer administration, thereby minimizing potential information bias and enhancing standardization. Given increasing smartphone penetration in Pakistan and similar contexts, such digital tools may facilitate scalable data collection in both outpatient clinics and community-based research initiatives.

Several limitations warrant consideration. The 48-hour retest interval, while appropriate for minimizing recall bias, assumes clinical stability and may not fully account for symptom fluctuation in chronic pain populations. The relatively small number of participants in the severe activity subgroup limited precision of subgroup comparisons and widened confidence intervals. Additionally, structural validity findings were based on exploratory factor analysis without confirmatory modeling, and measurement invariance between audio-assisted and non-audio users was not formally tested. Future studies should incorporate longitudinal responsiveness analysis, confirmatory factor modeling, and cross-mode equivalence testing to strengthen evidence for digital implementation.

Overall, the present findings indicate that the NDI-U mobile application demonstrates excellent reproducibility, acceptable internal consistency, moderate convergent validity, and adequate distributional properties in adults with chronic mechanical neck pain. These results support its use as a reliable digital outcome assessment tool in Urdu-speaking populations and contribute to the growing evidence base for electronic patient-reported outcome measures in musculoskeletal rehabilitation contexts (19,25).

CONCLUSION

In conclusion, the Urdu mobile application version of the Neck Disability Index demonstrated excellent test–retest reliability, acceptable internal consistency, minimal floor and ceiling effects, and moderate convergent validity with established neck-specific and pain-related outcome measures in adults with chronic mechanical neck pain. The instrument showed stable reproducibility over a 48-hour interval (ICC = 0.95) with a minimal detectable change of approximately 4.8 points, supporting its utility for monitoring clinical change beyond measurement error. Although internal consistency was slightly below the conventional 0.70 threshold, item-level analysis supported retention of the full 10-item structure, and construct-consistent gradients across physical activity limitation levels reinforced its discriminative validity. Structural validity findings suggest multidimensionality that warrants further confirmatory evaluation; however, overall psychometric performance indicates that the digitally administered NDI-U App is a reliable and clinically applicable tool for disability assessment in Urdu-speaking populations, particularly in contexts where literacy limitations may otherwise restrict standardized patient-reported outcome measurement.

REFERENCES

1. Alvi MA, Moghaddamjou A, Fehlings MG. Anatomy and physiology of cervical spine and cervical spinal cord. In: Degenerative Cervical Myelopathy. Elsevier; 2023. p. 11–33.
2. Panjabi MM, White AA III, Johnson RM. Cervical spine mechanics as a function of transection of components. J Biomech. 1975;8(5):327–36.
3. Linton SJ. A review of psychological risk factors in back and neck pain. Spine. 2000;25(9):1148–56.
4. Levy HI. Cervical pain syndromes: primary care diagnosis and management. Compr Ther. 2000;26(2):82–8.
5. Daffner SD, Hilibrand AS, Hanscom BS, Brislin BT, Vaccaro AR, Albert TJ. Impact of neck and arm pain on overall health status. Spine. 2003;28(17):2030–5.
6. Neelapala YR, Raja R, Bhandary A. A preliminary biopsychosocial analysis of online information on causes of neck pain. Musculoskeletal Care. 2019;17(2):277–81.
7. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis. 2014;73(6):968–74.
8. Safiri S, Kolahi AA, Cross M, Hill C, Smith E, Carson-Chahhoud K, et al. Prevalence, deaths, and disability-adjusted life years due to musculoskeletal disorders for 195 countries and territories 1990–2017. Arthritis Rheumatol. 2021;73(4):702–14.
9. Bovim G, Schrader H, Sand T. Neck pain in the general population. Spine. 1994;19(12):1307–9.
10. Genebra CVDS, Maciel NM, Bento TPF, Simeão SFAP, De Vitta A. Prevalence and factors associated with neck pain: a population-based study. Braz J Phys Ther. 2017;21(4):274–80.
11. Blanpied PR, Gross AR, Elliott JM, Devaney LL, Clewley D, Walton DM, et al. Neck pain: revision 2017 clinical practice guidelines. J Orthop Sports Phys Ther. 2017;47(7):A1–83.
12. Ranasinghe P, Perera YS, Lamabadusuriya DA, Kulatunga S, Jayawardana N, Rajapakse S, et al. Work-related complaints of neck, shoulder and arm among computer office workers. Environ Health. 2011;10:70.
13. Chan LLY, Wong AYL, Wang MH, Cheung K, Samartzis D. The prevalence of neck pain and associated risk factors among undergraduate students. Int J Ind Ergon. 2020;76:102934.
14. Croft PR, Lewis M, Papageorgiou AC, Thomas E, Jayson MI, Macfarlane GJ, et al. Risk factors for neck pain: a longitudinal study. Pain. 2001;93(3):317–25.
15. Boonstra AM, Preuper HRS, Balk GA, Stewart RE. Cut-off points for mild, moderate, and severe pain on the visual analogue scale. Pain. 2014;155(12):2545–50.
16. Shiina K. The historical roots of Visual Analog Scale in psychology. Front Hum Neurosci. 2021;15:711691.
17. Berthier F, Potel G, Leconte P, Touze MD, Baron D. Comparative study of methods of measuring acute pain intensity. Am J Emerg Med. 1998;16(2):132–6.

18. van Laar M, Pergolizzi JV Jr, Mellinghoff HU, Merchante IM, Nalamachu S, O'Brien J, et al. Pain treatment in arthritis-related pain. Open Rheumatol J. 2012;6:320–30.

19. Delgado DA, Lambert BS, Boutris N, McCulloch PC, Robbins AB, Moreno MR, et al. Validation of digital visual analog scale pain scoring. JAAOS Glob Res Rev. 2018;2(3):e088.

20. Goolkasian P. Neck Pain and Disability Scale: a critical evaluation. Expert Rev Pharmacoecon Outcomes Res. 2003;3(4):379–82.

21. Ailliet L, Knol DL, Rubinstein SM, de Vet HC, van Tulder MW, Terwee CB. Definition of the construct to be measured is a prerequisite for validity assessment. J Clin Epidemiol. 2013;66(7):775–82.

22. Mousavi SJ, Parnianpour M, Montazeri A, Mehdian H, Karimi A, Abedi M, et al. Translation and validation of Iranian versions of NDI and NPDS. Spine. 2007;32(26):E825–31.

23. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther. 1991;14(7):409–15.

24. Farooq MN, Mohseni-Bandpei MA, Gilani SA, Hafeez AJ. Urdu version of the Neck Disability Index: a reliability and validity study. BMC Musculoskelet Disord. 2017;18:1–11.

25. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing methodological quality of studies on measurement properties. Qual Life Res. 2010;19:539–49.

26. Langley G, Sheppeard H. The visual analogue scale: its use in pain measurement. Rheumatol Int. 1985;5(4):145–8.

27. Boonstra AM, Preuper HRS, Reneman MF, Posthumus JB, Stewart RE. Reliability and validity of the visual analogue scale for disability. Int J Rehabil Res. 2008;31(2):165–9.

28. Leak A, Cooper J, Dyer S, Williams K, Turner-Stokes L, Frank A. The Northwick Park Neck Pain Questionnaire. Rheumatology. 1994;33(5):469–74.

29. Gonzalez T, Balsa A, de Murieta JS, Zamorano E, González I, Martin-Mola E. Spanish version of NPQ: reliability and validity. Clin Exp Rheumatol. 2001;19(1):41–6.

30. Farooq MN, Naz S, Kousar A, Shahzad K. Cross-cultural adaptation and validation of NPQ to Urdu language. BMC Musculoskelet Disord. 2023;24(1):458.

31. McCarthy M, Grevitt M, Silcocks P, Hobbs G. The reliability of the Vernon and Mior Neck Disability Index. Eur Spine J. 2007;16:2111–7.

32. Young IA, Dunning J, Butts R, Mourad F, Cleland JA. Reliability and construct validity of NDI and NPRS in mechanical neck pain. Physiother Pract Res. 2019;35(12):1328–35.

33. Wlodyka-Demaille S, Poiraudeau S, Catanzariti JF, Rannou F, Fermanian J, Revel M. French translation and validation of disability scales for neck pain. Arch Phys Med Rehabil. 2002;83(3):376–82.

34. Lee H, Nicholson LL, Adams RD. Reliability and validity of Korean version of NDI. J Korean Med Sci. 2007;22(2):349–55.

35. Ortega JAA, Martínez ADD, Ruiz RA. Validation of the Spanish version of the Neck Disability Index. Spine. 2010;35(4):E114–8.

36. Shaheen AA, Omar MT, Vernon H. Cross-cultural adaptation and validation of the Arabic Neck Disability Index. Spine. 2013;38(14):E874–80.

DECLARATIONS

Ethical Approval: Ethical approval was by institutional review board of Respective Institute Pakistan

Informed Consent: Informed Consent was taken from participants.

Authors' Contributions:

Concept: AS; Design: NP; Data Collection: UE; Analysis: IZA; Drafting: MZ

Conflict of Interest: The authors declare no conflict of interest.

Funding: This research received no external funding.

Data Availability: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: NA

Study Registration: Not applicable.