healthcare

MDPI

Article

The Nijmegen Questionnaire in Non-Specific Chronic Neck Pain: Factor Structure, Construct Validity, Reliability and Responsiveness

Petros I. Tatsios [1,2,*], Eirini Grammatopoulou [1,2], Zacharias Dimitriadis [3] and George A. Koumantakis [1,2,*]

[1] Physiotherapy Department, School of Health & Care Sciences, University of West Attica (UNIWA), 12243 Athens, Greece; igrammat@uniwa.gr
[2] Laboratory of Advanced Physiotherapy, Physiotherapy Department, School of Health & Care Sciences, University of West Attica (UNIWA), 12243 Athens, Greece
[3] Health Assessment & Quality of Life Laboratory, Physiotherapy Department, University of Thessaly, 35100 Lamia, Greece; zdimitriadis@uth.gr
* Correspondence: ptatsios@uniwa.gr (P.I.T.); gkoumantakis@uniwa.gr (G.A.K.)

Highlights

What are the main findings?

- In patients with non-specific chronic neck pain, the Nijmegen Questionnaire shows preliminary support for a single-dimension interpretation: the three Van Dixhoorn subscales (shortness of breath, peripheral tetany, central tetany) could not be reproduced, and only the total score reached acceptable internal consistency ($\alpha = 0.809/\omega = 0.813$) and excellent within-day test–retest reliability (ICC = 0.92).
- The total score tracked disability, anxiety, depression and kinesiophobia (r = 0.35–0.58) yet was unrelated to every objective measure of breathing, cervical mobility and posture; measured against the minimal detectable change of 4.90 points derived here, it registered change of 2.9 × $MDC_{95\%}$ in the combined manual therapy plus breathing re-education group, whereas the change recorded under conventional physiotherapy remained within measurement error.

What are the implications of the main findings?

- Clinicians and researchers should report the Nijmegen total score alone in this population, and should not treat it as a proxy for breathing physiology—end-tidal CO_2, Hi-Lo pattern, respiratory rate, chest wall expansion, single breath count, and breath-holding time each require separate measurement.
- A raised Nijmegen score identifies a psychophysiological symptom burden that no objective respiratory, mobility or postural measure captures, supporting baseline screening with the total score and interpretation of individual change against the $MDC_{95\%}$ reported here.

Abstract

Background/Objectives: The Nijmegen Questionnaire (NQ) is the most widely used dysfunctional breathing screener that has not been psychometrically evaluated in non-specific chronic neck pain (NSCNP). This study examined the factor structure, construct validity, reliability and responsiveness of the NQ in NSCNP, using data from a randomized controlled trial. **Methods:** Ninety patients with NSCNP were randomized to three arms: MT with breathing re-education (EG1), MT with sham diaphragmatic component (EG2), or conventional physiotherapy (CG). Baseline NQ data ($n = 100$, including ten concurrently recruited, eligible but non-randomized patients) underwent principal axis factoring with parallel analysis. Pearson's correlations assessed construct validity against 20 clinical variables ($n = 90$); test–retest reliability in a subsample ($n = 30$); and responsiveness via

Check for updates

Academic Editor: Rodrigo Cappato De Araújo

Received: 23 August 2026
Revised: 17 November 2026
Accepted: 20 November 2026
Published: 23 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

the standardized response mean (SRM) and change-to-$MDC_{95\%}$ ratio. COSMIN-guided hypotheses: $r \geq 0.35$ with NDI, HADS-A/D and TSK; $r < 0.20$ with objective respiratory measures. **Results:** Parallel analysis supported a single factor explaining 24.35% of variance; the forced three-factor solution explained 36.52%. The total score showed acceptable internal consistency ($\alpha = 0.809/\omega = 0.813/\omega_h = 0.53$) and excellent test–retest reliability (ICC = 0.92 (95% CI: 0.85–0.96); SEM = 1.77; $MDC_{95\%} = 4.90$). It converged with disability, anxiety, depression and kinesiophobia ($r = 0.35$–0.58), but diverged from objective respiratory, cervical ROM and postural measures. Responsiveness was large in EG1 (SRM = 2.76; $2.9 \times MDC_{95\%}$) and EG2 (SRM = 1.31; $2.0 \times MDC_{95\%}$), whereas CG change did not exceed it. **Conclusions:** In NSCNP, the NQ measures psychophysiological dysregulation rather than breathing dysfunction. Its subscales lack structural validity in this population; only the total score should be used, with change interpreted against an $MDC_{95\%}$ of approximately five points.

Keywords: Nijmegen Questionnaire; chronic neck pain; dysfunctional breathing; factor analysis; construct validity; responsiveness; manual therapy; breathing re-education

1. Introduction

Non-specific chronic neck pain (NSCNP) is one of the most prevalent and burdensome musculoskeletal conditions globally, affecting approximately 30% of the adult population in any given year and ranking as the fourth leading cause of years lived with disability [1]. Beyond its musculoskeletal and psychological dimensions, accumulating evidence indicates that NSCNP is frequently accompanied by respiratory dysfunction, including altered breathing patterns [2], reduced respiratory muscle strength [3] and biochemical hypocapnia [4]. The anatomical rationale is well established: the accessory respiratory muscles—the scalenes, sternocleidomastoid and upper trapezius—are also primary cervical stabilizers, and their chronic overactivation through upper-chest breathing generates a self-reinforcing cycle of cervical overstrain, pain amplification and further breathing dysfunction [5].

Dysfunctional breathing (DB) is conceptualized as a multidimensional construct with three partially independent dimensions: biochemical (hyperventilation, hypocapnia), biomechanical (abnormal breathing pattern, thoracic dominance) and psychophysiological (anxiety-driven respiratory symptoms, breathing vigilance) [6]. This multidimensional framework, operationalized by Courtney [7,8] and recently extended by Vlemincx [9] into a transdiagnostic dimensional model, predicts that objective measures of each dimension may not be intercorrelated, and that subjective symptom questionnaires will primarily capture the psychophysiological dimension.

The Nijmegen Questionnaire (NQ) is the most widely used screening instrument for DB globally [10]. Developed by Van Dixhoorn and Duivenvoorden [11] and validated against the hyperventilation provocation test, the NQ comprises 16 symptom items yielding three subscales—shortness of breath (SOB), peripheral tetany (PT) and central tetany (CT)—with a clinical cut-off originally set at >22 for hyperventilation syndrome (HVS), derived against a comparison group who were themselves clinic patients. Re-analysis of the original validation datasets against normal rather than clinical subjects instead supports a threshold of ≥19 [12]. However, Van Dixhoorn and Folgering [12] subsequently clarified that an elevated NQ score does not indicate a specific syndrome, but rather reflects functional respiratory complaints mediated by stress and anxiety. A COSMIN systematic review by Li Ogilvie and Kersten [13] found the psychometric properties of the NQ to be poor to fair, with inadequate structural validity evidence. Li Ogilvie et al. [14] subsequently applied Rasch analysis and developed a revised 15-item version with superior structural validity.

Despite increasing recognition of the neck pain–breathing interface [2,3,5], the NQ has never been psychometrically evaluated in an NSCNP population. Stephen et al. [15] demonstrated higher NQ scores in neck pain patients than in controls ($p < 0.01$) and significant correlations with neck disability ($r = 0.50$) and perceived stress ($r = 0.54$), but did not examine factor structure or multivariate construct validity. Deshmukh et al. [16] reported a 68% prevalence of DB (assessed with the SEBQ) in neck pain—second only to low back pain—without psychometric analysis. The theoretical context is further enriched by Perri and Halford [17], who found in a foundational pilot study that neck pain was the only musculoskeletal condition significantly correlated with faulty breathing mechanics ($r = 0.243$), with 83.8% of neck pain participants exhibiting DB. A recent systematic review and meta-analysis by López-de-Uralde-Villanueva et al. [18] confirmed that respiratory dysfunction is prevalent across NSCNP subgroups, with patients with higher pain, disability and psychological comorbidity at greatest risk of concurrent DB.

The present study aimed to address three research gaps, firstly the factor structure and psychometric adequacy of the NQ in an NSCNP population, including whether the Van Dixhoorn three-factor model replicates; secondly, the convergent–divergent construct validity of the NQ against a comprehensive DB assessment battery comprising biochemical, biomechanical, and multidimensional measures, as well as musculoskeletal, and psychological measures; and lastly, the responsiveness of the NQ as an outcome measure in a three-group RCT comparing cervical MT combined with diaphragmatic MT and breathing re-education, cervical MT with sham diaphragmatic treatment, and conventional physiotherapy.

2. Materials and Methods

2.1. Study Design and Participants

Group comparisons of NQ scores across the three arms are reported elsewhere [19] and are not re-presented as findings here; the present paper is confined to the psychometric evaluation of the instrument itself. Ethical approval was granted by the Research Ethics Committee of the University of West Attica (No. 51758/01-06-2022), in accordance with the principles of the Declaration of Helsinki. The trial was registered with ClinicalTrials.gov under the identifier NCT05229393 (https://clinicaltrials.gov/study/NCT05229393, accessed on 24 November 2023). All participants provided written informed consent. Ninety outpatients with NSCNP, aged 25 to 65 years (two participants were enrolled marginally outside this pre-specified range, at 24 and 69 years, as reported in [20]). Their data were retained: both met all other inclusion criteria (mechanical neck pain >3 months duration, Grade I or II, able to complete Greek-language questionnaires), and their clinical profiles were indistinguishable from the remaining participants. Excluding them a priori on age grounds alone would have introduced selection bias without clinical justification; a sensitivity analysis excluding them is reported in Section 3.7. They had mechanical neck pain of more than three months' duration classified as Grade I or II, were able to complete Greek-language questionnaires, and were enrolled from a private physiotherapy practice. The full inclusion and exclusion criteria, the block randomization sequence and the allocation concealment procedure are set out in previous publications [19,21] and are not repeated here. Assessor blinding was maintained: all outcome assessments were performed by a physiotherapist without knowledge of group allocation. Participant blinding was not feasible. The randomization sequence was computer-generated using Research Randomizer (https://www.randomizer.org/, accessed on 10 May 2022).

EFA sample. To reach the commonly recommended minimum of 100 participants for factor analysis [22,23], the EFA sample additionally included ten patients with NSCNP who were recruited from the same private physiotherapy practice during the trial's recruitment period (June 2022–August 2023), but who did not enter the randomized trial (unable

to commit to the ten-session treatment schedule). They were screened against the same inclusion and exclusion criteria as the trial participants, provided written informed consent under the same ethics approval (No. 51758/01-06-2022), and completed the same Greek version of the NQ, in the same setting and with the same administration procedure (self-completed on a single occasion, before any treatment) as the trial participants at baseline. Because they were not trial participants, the remainder of the trial assessment battery was not administered to them and they received no study intervention. Their data therefore contributed only to the item-level descriptive statistics and to the EFA; all construct validity, reliability and responsiveness analyses were restricted to the 90 randomized participants. The decision to include these ten patients in the EFA was taken after the initial factor analysis of the 90 trial participants, on the basis of the above sample-size recommendations and not of the factor solution obtained; all such patients with complete NQ data were included, without selection. The trial's registered enrolment target ($n = 90$) and recruitment period were unaffected; therefore, no protocol amendment or trial registry update was required.

2.2. Interventions

All three groups received 10 treatment sessions over one month. Experimental Group 1 (EG1) received cervical spine manual therapy combined with diaphragmatic manual therapy and a structured breathing re-education program; Experimental Group 2 (EG2), the active control, received the same cervical manual therapy with sham diaphragmatic treatment and postural advice; and the Control Group (CG) received conventional physiotherapy. The specific techniques, dosages, equipment and the qualifications of the treating therapist and the blinded assessor are described in full in the published protocol [21] and in the trial reports [19,24], and are not reproduced here. It is acknowledged that the sham diaphragmatic ultrasound procedure in EG2 involved tactile contact; treatment time and therapist contact were equivalent across EG1 and EG2.

2.3. Outcome Measures

All measures were assessed at baseline (pre-treatment), immediately post-treatment (1 month) and at 3-month follow-up.

Primary psychophysiological DB measure. The Nijmegen Questionnaire [11] (NQ; 16 items, 0–4 Likert scale, total range 0–64) was the primary instrument under psychometric investigation. The Greek cultural adaptation was established according to international cross-cultural adaptation guidelines [25]. A score of ≥19 was used to define a positive screen, following the derivation of Van Dixhoorn and Folgering [12] against normal subjects, and consistent with the threshold specified in the trial protocol. Because several thresholds are in circulation, prevalence is additionally reported at the conventional >23 and at the Greek asthma-validated >17 [25] as a sensitivity analysis.

Biochemical DB. End-tidal CO_2 ($ETCO_2$, mmHg) was measured via portable capnography (Nonin Medical Inc., Plymouth, MN, USA, LifeSense® II), with a cut-off of ≤35 mmHg for hypocapnia [12]. Participants breathed through a nasal cannula for 10 min.

Biomechanical DB. Hi-Lo breathing assessment (Hi = chest-dominant, Lo = diaphragmatic, or Pa = paradoxical); respiratory rate (RR, breaths/min); and chest wall expansion (CWE) of the upper and lower rib cage (cm), measured with a tape measure at the axillary and xiphoid levels as maximal inspiration minus maximal expiration.

Multidimensional DB. Breath-holding time (BHT, seconds) after exhalation to tidal volume [26]; single breath count (SBC, counts) [7].

Musculoskeletal. Numeric Rating Scale for pain intensity (PI-NRS, 0–10) [27]; Neck Disability Index (NDI, 0–100%; validated Greek version) [28,29].

Psychological. Hospital Anxiety and Depression Scale (HADS-A and HADS-D, 0–21 each; validated Greek version) [30,31]; Tampa Scale for Kinesiophobia (TSK, 17–68; validated Greek version) [32,33].

Cervical Spine range of motion (CS-ROM). Flexion, extension, bilateral side flexion and bilateral rotation were recorded in degrees using an inertial sensor-based electrogoniometer, as described and validated previously [20].

Cervical spine postural measure. Craniovertebral angle (CVA), in degrees, obtained from lateral photography, as described and validated previously [20].

2.4. Statistical Analysis

All analyses were performed in IBM SPSS Statistics 29.0.2.0 (IBM Corp., Armonk, NY, USA). Normality was assessed via the Kolmogorov–Smirnov test with the Lilliefors significance correction, together with inspection of the Q–Q plots. Descriptive statistics were calculated and presented, according to the normal distribution assessment.

Exploratory factor analysis. The EFA was performed on the baseline NQ responses of all 100 participants (Section 2.1). The principal axis factoring (PAF) with direct oblimin rotation and Kaiser normalization was applied [34]. Direct oblimin rotation was selected in preference to orthogonal rotation (e.g., varimax) because the subscales of a psychophysiological measure were expected to be intercorrelated; oblique rotation permits factors to correlate and thereby avoids underestimating loadings when true factors are not independent. The Bartlett test of sphericity and the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy were used to examine the sufficiency of the population used as a sample for this study. Factor retention used parallel analysis (10,000 Monte Carlo simulations, 95th-percentile criterion) as the primary criterion [35,36], supported by scree plot and eigenvalue evaluation. The Kaiser criterion (eigenvalue > 1) was evaluated but not used as the primary retention rule, given its documented tendency to overextract factors. Communalities $\geq$0.30 and factor loadings $\geq$0.40 were used as adequacy thresholds [37].

Internal consistency. Cronbach's α was computed for the total and subscale scores [38]. McDonald's omega (ω) was additionally computed as a model-based internal consistency estimate [39]. Given the multidimensional structure of the scale comprising a general construct and three specific group factors, the McDonald's hierarchical omega (ω_h) was additionally computed via bifactor Schmid–Leiman transformation as a model-based estimate of the proportion of total-score variance attributable to the general factor [39,40]. This analysis was performed using the psych package in R via Jamovi (version 2.7.41), with 100 bootstrap iterations utilized to generate bootstrap 95% confidence intervals. Furthermore, to isolate the shared variance structure, the Explained Common Variance (ECV) was calculated to measure the proportion of common variance accounted for by the general factor. Additionally, the mean percentage of item variance (mean p^2) was utilized to assess the item-level saturation and ensure the general factor uniformly influenced individual items. Finally, to evaluate the structural validity and dimensionality of the scale, the global model fit was assessed using the Root Mean Square Error of Approximation (RMSEA), with values below 0.06 indicating good fit and values up to 0.08 indicating acceptable fit. To evaluate model parsimony and penalize over-parameterization, the Bayesian Information Criterion (BIC) was compared across models, where a lower (more negative) value indicates a preferred model.

Within-day test–retest reliability. Reliability was estimated in a third of the sample ($n = 30$), with a one-hour interval between the two administrations, using the two-way random effects, absolute agreement, single-measures model ($ICC_{2,1}$), single rater intraclass correlation coefficient ($ICC_{2,1}$) [41], together with the standard error of measurement (SEM) and the minimum detectable change at the 95% confidence level ($MDC_{95\%}$) [38]. ICC values

below 0.5 were taken to indicate poor, 0.5 to 0.75 moderate, 0.75 to 0.90 good and above 0.90 excellent reliability [37]. The SEM expresses the typical error attached to a single score, and the $MDC_{95\%}$ is the smallest change in an individual score that can be distinguished from that error [37].

Floor/ceiling effects. These would be considered present if the total score of the NQ scale corresponded to either of the extreme scale values (0 or 64) for more than 13 participants (15% of the sample) [42].

Construct validity. Pearson's correlations (r) examined bivariate associations between the NQ total score and 20 clinical variables. Convergent validity was inferred when significant correlations were found with theoretically related constructs, and divergent validity when correlations with theoretically unrelated constructs were small and non-significant [37].

Correlations were interpreted as negligible (0.0–0.10), weak (0.1–0.39), moderate (0.4–0.69), strong (0.70–0.89), or very strong (0.90–1.00) [38]. The required sample size to achieve 90% power with a relatively weak correlation level of r = 0.35 would be $n = 82$ participants, as calculated with an online program for computing sample size for correlational designs (https://sample-size.net/correlation-sample-size/ (accessed on 21 May 2022)) [43]. Due to the several correlations performed, the significance level was adjusted according to the Bonferroni method [44]. The adjusted a-level was calculated at a = 0.0025 based on the 20 correlations of interest between the NQ and all other outcomes. It is noted that this procedure assumes independence and may be overly conservative when variables are intercorrelated; uncorrected *p*-values are also reported for transparency.

Age-group and sex differences in NQ total score. Age-group differences (24–40 vs. 41–69 y) were assessed with the independent samples *t*-test (normality confirmed by Shapiro–Wilk per group). A pre-specified sensitivity analysis excluded the two participants enrolled outside the protocol age range. In addition, sex differences in NQ total score were examined with the independent samples *t*-test (normality confirmed by Shapiro–Wilk per group).

Responsiveness. The group-level change in NQ scores across the three assessment points, and the between-group comparison of that change, form part of the clinical effectiveness analysis of the parent trial and are reported in [19]. Responsiveness is therefore addressed here in psychometric rather than clinical terms, by means of two distribution-based indices that depend on the reliability parameters derived above and that have not previously been reported for this instrument in an NSCNP population. First, the standardized response mean (SRM) was calculated for each group and interval as the mean change divided by the standard deviation of that change, and interpreted as small at 0.20, moderate at 0.50 and large at 0.80 [38]. Second, the observed mean change was expressed as a ratio to the $MDC_{95\%}$ obtained in the reliability analysis, so that change exceeding measurement error could be distinguished from change that does not. The published change data on which these indices are computed are reproduced solely as their input, with the source acknowledged [19].

3. Results

3.1. Participants and Baseline Characteristics

Distributional assumptions were examined for all variables entering the psychometric analyses using the Kolmogorov–Smirnov test together with visual inspection of the corresponding Q-Q plots. The Kolmogorov–Smirnov test was applied with the Lilliefors significance correction, which is required whenever the distributional parameters are estimated from the sample rather than specified a priori. The parent trial reports [19,24] used the Shapiro–Wilk test; both procedures address the same distributional assumption, and

none of the analytical decisions taken here depends on which of the two is applied. Twelve of the 21 variables satisfied the formal test; for the remaining nine, the plotted standardized values departed only marginally from the reference line, and parametric procedures were therefore retained. Descriptive statistics for all variables are presented in Table 1. Ninety participants were enrolled and randomized (n = 30 per group), and the groups were equivalent at baseline; their demographic and clinical baseline characteristics are reported in [19,24]. The mean NQ total score for the sample of 90 participants was 20.36 (SD 6.95). Data from ten more patients were collected to increase the sample size to 100 participants for the factor analysis. The mean NQ total score for the sample of 100 participants was 21.30 (SD 7.54). Using the ≥19 threshold, 60.0% of participants (n = 60/100) screened positive at baseline. Prevalence was sensitive to the threshold applied: 68.0% (n = 68/100) at the Greek asthma-validated cut-off of >17 [25], 56.0% (n = 56/100) at >19, 45.0% (n = 45/100) at ≥23 and 41.0% (n = 41/100) at the conventional >23. This range in the same cohort is itself informative about the instrument. Biochemical hypocapnia ($ETCO_2$ of 35 mmHg or below) was present in 81.1% of participants, and 81 of 90 participants (90%) displayed a Hi breathing pattern, with the remaining 9 a Lo pattern. The response percent frequencies per item and the total score of the NQ are displayed in Table 2 with both mean (SD) and median (IQR) scores, as individual item scores were not normally distributed.

Table 1. Descriptive statistics of all outcome assessments (n = 90).

	Mean (SD)	**Min–Max**
NQ	20.36 (6.95)	4.0–38.0
NDI	31.71 (8.77)	18.0–66.0
PI-NRS	4.49 (1.21)	3.0–7.0
Pain duration	6.39 (1.72)	4.0–11.0
HADS-A	7.81 (2.62)	1.0–14.0
HADS-D	5.33 (3.03)	0.0–16.0
TSK	46.39 (5.62)	33.0–60.0
$ETCO_2$	31.29 (3.82)	20.8–40.3
BHT	28.15 (9.62)	13.5–54.5
SBC	37.63 (11.43)	13.0–64.0
RR	15.13 (3.09)	8.59–24.12
CWE upper	3.93 (0.76)	2.0–6.5
CWE lower	5.17 (1.06)	3.0–7.5
Flexion	46.83 (13.31)	21.2–76.7
Extension	50.45 (10.09)	29.4–73.4
Side flexion right	37.25 (8.14)	17.9–57.0
Side flexion left	37.23 (7.44)	20.9–55.3
Rotation right	67.15 (9.66)	42.7–99.83
Rotation left	67.20 (10.05)	38.6–89.4
CVA	47.91 (8.05)	23.5–66.4

NQ: Nijmegen Questionnaire, NDI: Neck Disability Index, PI-NRS: pain intensity numeric rating scale, HADS-A/HADS-D: Hospital Anxiety and Depression Scale, anxiety/depression subscales, TSK: Tampa Scale of Kinesiophobia, $ETCO_2$: end-tidal carbon dioxide, BHT: breath holding time, SBC: single breath count, RR: respiratory rate, CWE: chest wall expansion, CVA: craniovertebral angle. Baseline values for all variables other than the NQ have been published, disaggregated by treatment group, in [19,24]; they are presented here pooled across the whole sample as the reference distribution for the psychometric analyses reported below.

Table 2. Response percent frequencies per item and mean (SD)/median (IQR) scores per item and total of the NQ ($n = 100$).

Item	Never (%)	Seldom (%)	Sometimes (%)	Often (%)	Very Often (%)	Mean (SD)	Median (IQR)
1. Chest pain	31	30	32	7	0	1.15 (0.95)	1.0 (2.0)
2. Feeling tense	2	12	57	27	2	2.15 (0.73)	2.0 (1.0)
3. Blurred vision	52	30	16	2	0	0.68 (0.81)	0.0 (1.0)
4. Dizzy spells	41	25	28	5	1	1.00 (0.99)	1.0 (2.0)
5. Confused/losing touch	53	32	13	2	0	0.64 (0.78)	0.0 (1.0)
6. Rapid or deep breathing	27	36	29	8	0	1.18 (0.92)	1.0 (2.0)
7. Shortness of breath	21	25	44	10	0	1.43 (0.93)	2.0 (1.0)
8. Chest tightness	20	26	45	9	0	1.43 (0.91)	2.0 (1.0)
9. Bloated sensation	9	19	53	19	0	1.82 (0.84)	2.0 (1.0)
10. Tingling fingers	51	24	9	15	1	0.91 (1.14)	0.0 (1.2)
11. Unable to breathe deeply	27	41	25	7	0	1.12 (0.89)	1.0 (2.0)
12. Stiff fingers or arms	65	26	6	3	0	0.47 (0.74)	0.0 (1.0)
13. Tight around the mouth	70	21	7	2	0	0.41 (0.71)	0.0 (1.0)
14. Cold hands or feet	24	16	16	25	19	1.99 (1.47)	2.0 (2.0)
15. Palpitations	8	21	51	17	3	1.86 (0.90)	2.0 (1.0)
16. Anxiety	0	3	20	45	32	3.06 (0.80)	3.0 (1.0)
Total Score						21.30 (7.54)	20.5 (10.7)

3.2. Factor Analysis: One-Factor PAF Solution (Optimal Solution)

The Kaiser–Meyer–Olkin measure (KMO = 0.746, middling [45]) and Bartlett's test of sphericity ($\chi^2(120) = 441.906$, $p < 0.001$) confirmed factorability of the $n = 100$ sample. Parallel analysis (10,000 Monte Carlo simulations, 95th-percentile criterion [35,36]) identified a single factor: the F1 eigenvalue (4.587) substantially exceeded its 95th-percentile random threshold (1.934), whereas F2 (1.682) marginally fell below its threshold (1.700), and F3 (1.380) through F5 (1.028) fell clearly below their respective thresholds despite exceeding the Kaiser criterion of >1. Using the random-mean criterion (the more liberal alternative), two factors would be retained (F2 actual = 1.682 vs. random mean = 1.584); however, the 95th-percentile standard is more conservative and is recommended [35,36]. The initial eigenvalues for the five Kaiser-criterion factors were F1 = 4.587 (28.67%), F2 = 1.682 (10.51%), F3 = 1.380 (8.62%), F4 = 1.169 (7.31%), F5 = 1.028 (6.42%); cumulative 61.54%. Therefore, forcing three factors constitutes overextraction by the primary PA criterion. The one-factor PAF solution (extraction sums of squared loadings) explained 24.35% of variance. Additionally, 10 of 16 items had communalities below 0.30 in the one-factor solution, and iterative item reduction converged on only three items (NQ1 chest pain, NQ4 dizzy spells and NQ5 confused or losing touch) that met all psychometric thresholds (communality ≥0.30, loading ≥0.40; final $\alpha = 0.671$, variance explained 40.8%). Internal consistency (Cronbach's α, McDonald's ω, and ωh) is reported in Section 3.5.

3.3. Forced Three-Factor PAF with Oblimin Rotation Solution

The forced three-factor PAF with direct oblimin rotation produced a pattern matrix with positive primary loadings and factor intercorrelations of r = 0.266 (F1–F2), −0.319 (F1–F3) and −0.255 (F2–F3). The solution explained 36.52% of the total variance (extraction sums: F1 = 24.99%, F2 = 6.77%, F3 = 4.75%; initial eigenvalues: 47.81%), which remains well below the conventionally accepted 50% benchmark. It should be noted that application of PAF to ordinal item data may suppress communalities and inflate uniqueness, potentially contributing to the low variance explained independently of any structural inadequacy of the scale.

The factor structure corresponded only partially to the Van Dixhoorn & Duivenvoorden (1985) [11] taxonomy. Specifically, F1 (SOB) partially loaded (>0.40) on 4 out of 7 SOB items of the original publication, specifically NQ1 (chest pain, 0.596), NQ2 (feeling tense, 0.468), NQ8 (chest tightness, 0.678), and NQ15 (palpitations, 0.422); however, it also loaded on NQ9 (bloated sensation, 0.669), originally corresponding to CT. Additionally, F2 loaded on NQ10 (tingling fingers, 0.677) and NQ12 (stiff fingers or arms, 0.657), corresponding to the originally identified PT subscale; however, it was missing the two other items (NQ13 and NQ14). Furthermore, F3 (CT) did not load >0.40 on the four originally identified items (NQ3, NQ4, NQ5, and NQ9) but primarily loaded on NQ6 (rapid or deep breathing, −0.726), NQ7 (shortness of breath, −0.505), and NQ11 (unable to breathe deeply, 0.466); however, these latter three items were under the SOB subscale in the original publication. Importantly, 5 of 16 items had communalities below 0.30. Also, the F1 (SOB), F2 (PT) and F3 (CT) subscales yielded an internal consistency Cronbach's α index of 0.744, 0.631 and 0.699, the first only considered acceptable and the latter two questionable. Detailed loadings and communalities for both solutions are presented in Table 3.

Table 3. Principal axis factoring results: one-factor and forced three-factor direct oblimin solutions ($n = 100$).

			One-Factor		Forced Three-Factor			
Item	**VD85**	**Item**	**1F**	**h^2 (1F)**	**F1 SOB**	**F2 PT**	**F3 CT**	**h^2 (3F)**
NQ1	SOB	Chest pain	0.657	0.431	0.596	0.046	−0.177	0.475
NQ2	SOB	Feeling tense	0.544	0.296 †	0.468	0.203	−0.038	0.327
NQ3	CT	Blurred vision	0.214	0.046 †	−0.274	0.372	−0.352	0.289 †
NQ4	CT	Dizzy spells	0.701	0.491	0.322	0.351	−0.316	0.509
NQ5	CT	Confused/losing touch	0.476	0.226 †	0.170	0.264	−0.255	0.249 †
NQ6	SOB	Rapid or deep breathing	0.613	0.376	0.222	−0.075	−0.726	0.648
NQ7	SOB	Shortness of breath	0.453	0.205 †	0.106	0.036	−0.505	0.312
NQ8	SOB	Chest tightness	0.562	0.316	0.678	0.022	0.020	0.460
NQ9	CT	Bloated sensation	0.546	0.298 †	0.669	−0.060	−0.036	0.445
NQ10	PT	Tingling fingers	0.292	0.085 †	−0.006	0.677	0.099	0.432
NQ11	SOB	Unable to breathe deeply	0.564	0.318	0.37	−0.112	0.466	0.427
NQ12	PT	Stiff fingers or arms	0.402	0.161 †	0.008	0.657	−0.064	0.460
NQ13	PT	Tight around the mouth	0.280	0.079 †	0.279	0.321	0.197	0.200 †
NQ14	PT	Cold hands or feet	0.261	0.068 †	0.309	−0.032	−0.023	0.096 †
NQ15	SOB	Palpitations	0.549	0.301	0.422	0.078	−0.202	0.305
NQ16	_	Anxiety	0.447	0.200 †	0.228	0.255	−0.146	0.209 †

1F = one-factor (unrotated) loading; F1/F2/F3 = three-factor direct oblimin pattern matrix; h^2 = communality. † marks communality <0.30. VD85 = Van Dixhoorn and Duivenvoorden (1985) [11] subscale assignment (SOB = shortness of breath; PT = peripheral tetany; CT = central tetany), also highlighted in gray within the forced three-factor solution respective squares.

3.4. Convergent–Divergent Construct Validity

Convergent validity. The NQ total score demonstrated moderate, yet highly significant correlations with NDI ($r = 0.58$, $p < 0.001$), HADS-A ($r = 0.53$, $p < 0.001$), HADS-D ($r = 0.42$, $p < 0.001$), and TSK ($r = 0.35$, $p < 0.001$). Anxiety retained a highly significant association with the NQ beyond disability—with HADS-A and HADS-D retaining significance in partial correlation analysis when controlling for NDI ($r = 0.35$, $p < 0.001$), while the TSK was not significantly associated if the Bonferroni criteria for significance correction were applied ($r = 0.22$, $p = 0.034$) when controlling for NDI.

Divergent validity. None of the biochemical or biomechanical DB measures reached significance: $ETCO_2$ ($r = -0.16$, $p = 0.121$), BHT ($r = -0.13$, $p = 0.209$), respiratory rate ($r = -0.003$, $p = 0.975$), Hi-Lo ($r = -0.15$, $p = 0.155$), CWE upper ($r = -0.14$, $p = 0.191$),

CWE lower (r = −0.17, $p = 0.111$) and SBC (r = −0.26, $p = 0.014$, higher than the corrected p value level of 0.0025). Also, the correlation with PI-NRS was weak and non-significant according to the Bonferroni correction (r = 0.22, $p = 0.038$, higher than the corrected p-value level of 0.0025), and pain duration was non-significant (r = −0.08, $p = 0.447$). All CS-ROM variables and the CVA did not correlate significantly with the NQ. The NQ total score was therefore independent of all objective respiratory physiological measures, as well as pain and CS-ROM. All correlations are presented in Table 4.

Table 4. Pearson's correlations between the Nijmegen questionnaire scores and clinical domain variables ($n = 90$).

Variable	Domain	NQ Total	Uncorrected p
NDI	Disability (musculoskeletal)	0.58 ***	<0.001
PI-NRS	Pain intensity (musculoskeletal)	0.22 †	0.038
Pain duration	Clinical history	−0.08	0.447
HADS-A	Anxiety (psychological)	0.53 ***	<0.001
HADS-D	Depression (psychological)	0.42 ***	<0.001
TSK	Kinesiophobia (psychological)	0.35 ***	<0.001
$ETCO_2$	Biochemical DB	−0.16	0.121
BHT	Multidimensional DB	−0.13	0.209
SBC	Multidimensional DB	−0.26 †	0.014
RR	Biomechanical DB	−0.003	0.975
Hi-Lo	Biomechanical DB	−0.15	0.155
CWE upper	Biomechanical DB	−0.14	0.191
CWE lower	Biomechanical DB	−0.17	0.111
Flexion	CS-ROM	−0.16	0.132
Extension	CS-ROM	−0.03	0.759
Side flexion right	CS-ROM	−0.16	0.138
Side flexion left	CS-ROM	−0.07	0.491
Rotation right	CS-ROM	−0.06	0.591
Rotation left	CS-ROM	−0.05	0.618
CVA	Posture	−0.006	0.954

*** $p < 0.001$, † $p < 0.05$ uncorrected, non-significant after Bonferroni, Note that the two participants aged outside the pre-specified range (24 and 69 years) were included in the full analysis but excluded from the sensitivity analysis reported in Section 3.7, CS-ROM: cervical spine range of motion, NDI: Neck Disability Index, PI-NRS: pain intensity numeric rating scale, HADS-A/HADS-D: Hospital Anxiety and Depression Scale, anxiety/depression subscales, TSK: Tampa Scale of Kinesiophobia, $ETCO_2$: end-tidal carbon dioxide, BHT: breath holding time, SBC: single breath count, RR: respiratory rate, CWE: chest wall expansion, CVA: craniovertebral angle.

3.5. Internal Consistency and Test–Retest Reliability

The total-score Cronbach's α was 0.809 (95% CI: 0.764–0.855) and McDonald's $\omega = 0.813$ (95% CI: 0.759–0.866), confirming good (>0.80) internal consistency by both estimators.

McDonald's hierarchical omega ($\omega_h = 0.53$, bootstrapped 95% CI: 0.13–0.94) was additionally computed as a model-based estimate of the general factor's contribution to internal consistency. This indicates that while traditional internal consistency indices are at a good level, just over half of the total score variance (53%) is driven by the overarching general construct. At the structural and item levels, the general factor captured less than half of the common variance (ECV = 0.44), and individual items exhibited a parallel trend, with the general factor accounting for a moderate portion of item-level variance (mean $p^2 = 0.43$). The single general-factor model demonstrated poor absolute fit (RMSEA = 0.095, 90% CI: 0.076–0.116). Conversely, the bifactor model provided an improved and acceptable absolute fit to the data (RMSEA = 0.06, 90% CI: 0.027–0.089). However, when accounting for model

complexity, the information criteria favored the simpler, strict unidimensional structure (BIC = −279.9) over the heavily parameterized bifactor model (BIC = −242.2).

The test–retest within-day reliability measured with the $ICC_{2,1}$ (95% CI) was excellent (>0.90), with the error reliability indices (SEM and $MDC_{95\%}$) sufficiently low (Table 5). Specifically, in terms of absolute reliability, the SEM and $MDC_{95\%}$ values were at acceptable levels relative to the range of values of the NQ (0–64); if interpreted as percent error for the NQ in relation to the grand mean of 18.32, these amounted, respectively, to SEM% = (1.77/18.32) × 100 = 9.66% and $MDC_{95\%}$% = (4.90/18.32) × 100 = 26.75%.

Table 5. Descriptive statistics and test–retest intra-rater reliability coefficients of the NQ (n = 30).

Mean (SD) Measurement 1	Mean (SD) Measurement 2	$ICC_{2,1}$ (95% CI)	SEM	$MDC_{95\%}$
18.53 (6.97)	18.10 (5.91)	0.92 (0.85–0.96)	1.77	4.90

ICC: intraclass correlation coefficient, CI: confidence interval, MDC: minimum detectable change, SEM: standard error of the measurement, SD: standard deviation.

3.6. Floor and Ceiling Effects

No floor or ceiling effects were noted, as participants' range of scores for the NQ was between 4 and 38, with most of them scoring between 12 and 30 (Figure 1).

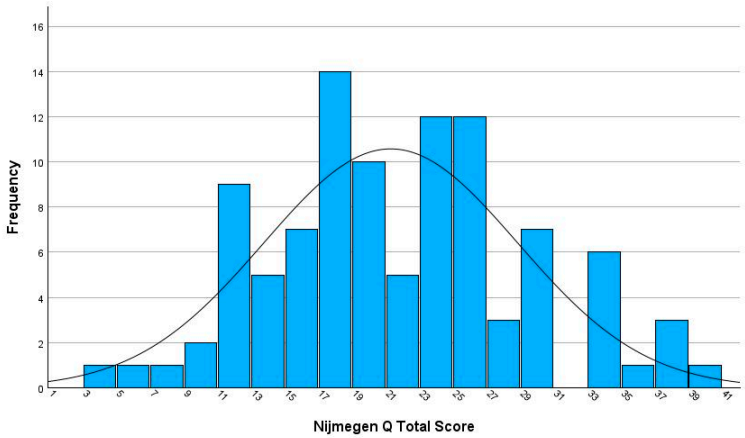

Figure 1. Frequency distribution of the responses from participants (n = 100) in the Nijmegen questionnaire.

3.7. Sensitivity Analysis

A post hoc sensitivity analysis excluding the two participants enrolled outside the protocol age range (aged 24 and 69; n = 98) yielded almost identical conclusions for internal consistency indices: α = 0.816 (95% CI: 0.762–0.869), ω = 0.812 (95% CI: 0.766–0.858), ω_h = 0.55 (bootstrapped 95% CI: 0.10–0.91); for test–retest reliability $ICC_{2,1}$ = 0.91 (95% CI: 0.82–0.95) in n = 28 participants. Also, the convergent–divergent correlation pattern was unchanged (all $\Delta r \leq 0.02$; n = 88).

3.8. Age-Group and Sex Differences in NQ Total Score

An independent samples t-test demonstrated non-significant between-group differences in the NQ total score for age (24–40 y vs. 41–69 y), assessed in all (n = 100) participants (24–40 y group (n = 50): M (SD) = 20.94 (7.45) y vs. 41-69 y group (n = 50): M (SD) = 21.66 (7.69) y; t(98) = 0.475, p = 0.636, mean difference (95% CI) = 0.72 (−2.29, 3.73) y, Cohen's d = 0.095).

Also, an independent samples *t*-test revealed that female participants ($n = 60$) scored significantly higher on the NQ total score than male participants ($n = 40$) (females: M = 22.95, SD = 7.26; males: M = 18.82, SD = 7.36; t(98) = 2.77, $p = 0.007$, mean difference (95% CI) = 4.12 (1.17, 7.08), Cohen's d = 0.565).

3.9. Responsiveness

The change in NQ scores across the three assessment points (Table 6), and the between-group comparisons of that change (Table 7), are reported as part of the clinical effectiveness analysis of the parent trial in [19] and are not restated as findings here. Applying the reliability parameters obtained above to those published change data yields the responsiveness indices set out below. Between baseline and the 3-month follow-up, the standardized response mean was 2.76 in EG1 and 1.31 in EG2, both large by conventional criteria, corresponding to a change of 2.9 and 2.0 times the $MDC_{95\%}$ of 4.90 points respectively. Over the same interval, the mean change recorded in the CG (SRM = 1.04) amounted to 0.84 times the MDC95% and therefore falls within measurement error at the group-mean level. Change between post-treatment and follow-up did not exceed the $MDC_{95\%}$ in any group. The NQ therefore registers a change well above its own measurement error in patients receiving breathing-directed treatment, which is the property required of a responsive outcome measure.

Table 6. Within-group changes in the Nijmegen questionnaire, reproduced from [19] as the input to the responsiveness indices reported in the text.

Group	Comparison	ΔM	*p*
EG1	Baseline—post-treatment	−10.80	<0.001
EG1	Baseline—3 months	−14.33	<0.001
EG1	Post-treatment—3 months	−3.53	<0.001
EG2	Baseline—post-treatment	−7.29	<0.001
EG2	Baseline—3 months	−9.73	<0.001
EG2	Post-treatment—3 months	−2.44	0.011
CG	Baseline—post-treatment	+0.97	0.167
CG	Baseline—3 months	+4.10	<0.001
CG	Post-treatment—3 months	+3.13	<0.001

EG = experimental group, CG = control group, ΔM = mean change in NQ points (negative values denote a reduction in NQ score, i.e., improvement).

Table 7. Between-group comparisons of the Nijmegen questionnaire mean (SD) change scores, reproduced from [19] as the input to the responsiveness indices reported in the text.

Change Score (Δ)	EG1 M (SD)	EG2 M (SD)	CG M (SD)	EG1 vs. EG2	EG1 vs. CG	EG2 vs. CG
Baseline—post-treatment	−10.80 (4.31)	−7.29 (5.55)	+0.97 (3.48)	0.003	<0.001	<0.001
Baseline—3 months	−14.33 (5.19)	−9.73 (7.42)	+4.10 (3.96)	0.007	<0.001	<0.001
Post-treatment—3 months	−3.53 (2.53)	−2.44 (4.70)	+3.13 (2.50)	0.528 *ns*	<0.001	<0.001

EG = experimental group, CG = control group; negative values denote improvement (a reduction in NQ score).

4. Discussion

4.1. Factor Structure: Preliminary Support for a Single Psychophysiological Burden Dimension

This is the first study to systematically examine the factor structure of the NQ in an NSCNP population. Parallel analysis—the methodologically superior criterion to the commonly used Kaiser eigenvalue rule [35]—identified a single factor as the optimal solution. The forced three-factor oblimin solution partially replicated the Van Dixhoorn SOB/PT/CT

taxonomy in terms of item direction, but failed all psychometric quality thresholds: the variance explained (36.52% extraction sums) was below the 50% benchmark—a threshold that is not universal and commonly unmet when items span heterogeneous domains; part of the low communalities may also reflect a PAF-on-ordinal-items artifact, all subscales yielded lower Cronbach's α, the two out of three below the accepted minimum of 0.70, 5 of 16 items had communalities below 0.30, and the factor intercorrelations (range approximately −0.319 to 0.266) indicated that the three factors were related rather than independent constructs. However, a single-factor solution explained only 24.35% of total variance and 10 of 16 items had communalities below 0.30, indicating limited shared variance among the items. These results support the use of the total score but should not be interpreted as proof of unidimensionality.

These findings extend and contextualize prior psychometric work on the NQ. Li Ogilvie and Kersten [13] demonstrated that the original three-factor structure of the NQ was established using non-metric principal components analysis on an inadequate sample of 75 patients. Li Ogilvie et al. [14] subsequently found all 16 items to have disordered response thresholds in Rasch analysis and identified NQ14 as a misfitting item with gender differential item functioning. Grammatopoulou et al. [25] found a single-factor 11-item solution in Greek asthma patients ($\alpha = 0.92$, variance explained 58.6%)—a more coherent structure than the one found here. The discrepancy is clinically expected, as neck pain patients may also endorse peripheral tetany items (tingling fingers, stiff fingers, cold hands) for radiculopathic rather than hypocapnic reasons, contaminating the intended PT construct of the NQ. The current findings converge with a growing consensus [12–14,25] that the NQ subscales lack robustness across populations and that only the total score warrants clinical use.

4.2. Construct Validity: The NQ as a Transdiagnostic Psychophysiological Measure

The convergent–divergent validity profile is striking in its clarity. The NQ total score showed moderate convergent validity with disability, anxiety, depression and kinesiophobia, and consistent divergent validity from every objective respiratory measure (all correlations non-significant with 95% CIs spanning zero), consistent with dimensional independence. This pattern is precisely what the transdiagnostic dimensional model of Vlemincx [9] predicts: the biochemical, biomechanical and psychophysiological dimensions of dysfunctional breathing are partially independent mechanisms, and the NQ—measuring subjective respiratory symptom burden—operates on the psychophysiological dimension while remaining dissociated from the biochemical ($ETCO_2$) and biomechanical (Hi-Lo, CWE, respiratory rate) dimensions. Therefore, measures of other dimensions of DB (CWE, Hi-Lo, BHT, ETCO2, and RR) require independent assessment and are not substitutable by the NQ.

The complete independence of the NQ from CWE, a direct measure of rib cage mobility in both the upper and lower regions, further underscores that chest wall mechanical restriction, despite its near-universal presence in this sample (90% chest-dominant breathing), does not drive subjective respiratory symptom burden. This finding is consistent with Courtney's [7,8] observation that the three dimensions of dysfunctional breathing are operationally independent in populations with musculoskeletal concerns. Also, the non-significant associations between NQ total and all CS-ROM directions, as well as the CVA, suggest that these variables do not drive subjective respiratory symptom burden.

There were moderate but highly significant correlations of the NQ with the NDI, HADS-A, HADS-D, and TSK documented in the current sample. These findings are broadly consistent with those of Stephen et al. [15] in South African patients with neck pain, who reported a closely comparable NQ–perceived stress association ($r = 0.54$, against $r = 0.53$ for the NQ and HADS-A in the current study) and a somewhat weaker NQ–NDI asso-

ciation (r = 0.50, against r = 0.58 in the current study). Perceived stress and anxiety are related but not interchangeable constructs, so the latter comparison is indicative rather than direct. The present results also align with the clarification by Van Dixhoorn and Folgering [12] that an elevated NQ score signals that factors other than the patient's index medical condition—principally stress, anxiety and heightened perception of bodily and respiratory sensations—may be contributing to the reported complaints, rather than identifying a specific HVS. The independent contribution of anxiety beyond disability—with HADS-A retaining significance in partial correlation analysis (r = 0.35 controlling for NDI)—implicates the breathing vigilance construct described by Steinmann et al. [46] as a potential explanatory mechanism warranting future investigation.

The significantly higher NQ total scores in female participants compared with male participants are consistent with well-established sex differences in anxiety, somatic symptom burden and hyperventilation-related complaints [47–49]. Rather than representing a confound, this finding corroborates the construct validity of the NQ as a psychophysiological burden measure: the instrument is sensitive to a dimension—anxious somatic arousal—that is known to be more prevalent in women. Clinicians should be aware that female NSCNP patients are more likely to screen positive for dysfunctional breathing on the NQ, and future normative work should evaluate whether sex-stratified cut-off thresholds improve clinical utility.

In contrast, age group (24–40 vs. 41–69 y) did not significantly differentiate NQ total scores, suggesting that the degree of functional respiratory complaints measured by the NQ is not strongly age-dependent within the working-age adult range studied here. This finding is broadly consistent with the observation that dysfunctional breathing and hyperventilation-related symptoms are not exclusively age-linked phenomena but may reflect individual psychophysiological reactivity patterns that cut across age strata. Future studies with wider age ranges or larger sub-group samples would be required to detect more subtle age effects.

4.3. Reliability and Internal Consistency of the NQ

The NQ total score demonstrated excellent test-retest reliability in a sub-sample of the population. A limitation, however, is that only within-day data were collected. The short one-hour time interval increases the likelihood of a memory effect, which may artificially inflate the reliability estimates. To build upon these findings, future studies should evaluate the long-term test–retest reliability of the NQ total score. Future research should implement longer, more methodologically robust time intervals between assessments, such as a two-week period. This extension is essential to mitigate the short-term memory effects and recall bias associated with within-day testing. Additionally, testing the instrument across diverse and larger population samples will help establish its broader generalizability, structural stability, and temporal reliability over time.

However, our findings present a nuanced case regarding the dimensionality and scoring composition of the NQ. At first glance, traditional internal consistency reliability metrics demonstrated a robust and cohesive scale, with both Cronbach's alpha (0.809) and McDonald's total ω (0.813) comfortably exceeding standard thresholds [39,40,42]. This finding may suggest that the items of the questionnaire are adequately interrelated and measure a cohesive underlying construct—specifically, subjective 'functional respiratory complaints'. However, relying entirely on α or total ω creates a misleading illusion of unidimensionality. When we disattenuate the variance using a bifactor framework, the model-based indices strongly caution against treating this scale as a cleanly unidimensional instrument. The ω_h of 0.53 fell well below the widely accepted threshold (typically 0.70 or 0.80) required to justify the interpretation or clinical use of a single total score. Because

nearly half of the total score variance (47%) remains unaccounted for by the general factor, a global total sum score would conflate the general construct with substantial unique subscale variance.

This is strongly reinforced by the structural and item-level indices. The ECV (0.44) and mean p^2 (0.43) consistently demonstrate that the general factor possesses only a moderate degree of saturation. Because the specific subscales retain more than half of the common and item-level variance, they are contributing substantial unique information rather than acting as mere random error. Furthermore, the bifactor model achieved a superior absolute structural fit (RMSEA = 0.06) compared to the poorly fitting single-factor model (RMSEA = 0.095). This implies that ignoring the multidimensional subscale architecture of the instrument severely distorts how well the model maps onto the empirical data. Finally, the BIC values favor the single-factor model (−279.9 vs. −242.2), signaling that the statistical complexity and loss of parsimony required to estimate the bifactor model may not be entirely justified if the ultimate goal is to isolate a strong, dominant general factor.

A notable limitation of this structural evaluation is the sample size (N = 100). This lower sample size is directly reflected in the wide bootstrapped 95% confidence interval for ω_h (0.13 to 0.94). While the point estimate (0.53) suggests poor unidimensionality, the true population value remains highly uncertain, ranging from severely weak to highly acceptable.

These findings warrant caution when interpreting the total score as a predominantly unidimensional measure. Nevertheless, the failure to reproduce the originally proposed three-subscale structure does not necessarily establish that the questionnaire is unidimensional. The total score may still provide a clinically useful summary of NQ, but its interpretation should be supported by the combined evidence from factor structure, reliability and construct validity. Further studies with larger samples and independent confirmation of the factor structure are needed to establish the dimensionality and general-factor reliability of the questionnaire in this population.

4.4. Responsiveness: Change in Relation to Measurement Error

Responsiveness was evaluated here against the instrument's own measurement error rather than as a clinical result. In the two arms that received manual therapy, the standardized response mean was large, and the observed change exceeded the $MDC_{95\%}$ by a factor of two to three, indicating that the NQ registers treatment-related change well beyond what measurement error alone would produce. This property is consistent with evidence from breathing retraining trials in asthma [50–52], in which normalization of the NQ is a primary therapeutic target. The clinical interpretation of the between-group differences themselves belongs to the trial report [19] and is not repeated here. It is important to distinguish responsiveness (sensitivity to detecting change) from treatment efficacy. The $MDC_{95\%}$ is a measurement-error index and should not be equated with a patient-important minimal important change (MIC); an anchor-based MIC for the NQ in NSCNP has not been established.

One observation nevertheless bears directly on the use of the instrument. In the arm receiving conventional physiotherapy, the NQ did not improve, and the group means at follow-up remained above the ≥19 screening threshold, although the magnitude of that change fell below the $MDC_{95\%}$ established here and should not be read as reliable deterioration at the individual level. Taken together with the absence of any correlation between the NQ and the objective respiratory, mobility and postural measures, this indicates that psychophysiological respiratory symptom burden constitutes a distinct clinical domain that will go undetected unless it is screened for directly, and supports administering the NQ at baseline in this population.

4.5. Clinical Cut-Off Sensitivity

As reported in Section 3.1, the proportion of participants classified as NQ-positive varied substantially depending on the threshold applied: 68.0% at >17 (Greek asthma-validated cut-off [25]), 56.0% at >19, 45.0% at ≥23, and 41.0% at the conventional >23. This range in NQ–positive proportion of participants within the same cohort has direct implications for clinical and research interpretation of NQ scores in NSCNP populations. The >19 cut-off, re-derived from the original validation datasets using a normative rather than clinical comparison group [12], produced intermediate prevalence estimates and may represent a more conservative screening threshold than the >17 value validated in a Greek asthma sample. Until a cut-off is formally validated in NSCNP populations, researchers and clinicians should report and justify their chosen threshold explicitly, as prevalence figures are not directly comparable across studies applying different cut-offs. The present study's findings at multiple thresholds are reported to enable such comparisons.

4.6. Limitations

Several limitations should be acknowledged. First, the sample was recruited from a single outpatient physiotherapy setting and was not randomly selected, so generalizability to primary care or community-dwelling populations requires caution. Second, the exploratory factor analysis was conducted on 100 participants; while adequate by Mundfrom et al.'s [53] recommendations for a one-factor solution (the optimal solution here) with a variables-to-factors ratio of 16 and low-to-moderate communalities ($n \geq 20$–50 required; $n = 100$ substantially exceeds this). For a three-factor solution with a variables-to-factors ratio of approximately 5 and low communalities, the same table indicates $n \approx 40$–150, placing $n = 100$ within the adequate range but not at the upper end, as a larger sample would yield more precise communality estimates. Furthermore, some methodologists recommend a minimum of 100 participants for any factor analysis regardless of the variables-to-factors ratio [22,23]; the present sample meets this threshold exactly. This threshold was reached by including ten concurrently recruited, eligible patients who were not part of the randomized trial, a decision taken after the initial analysis of the trial sample; although they met identical eligibility criteria and completed the NQ under the same procedure, this post hoc augmentation, together with the absence of a separately pre-registered psychometric protocol, should be considered when interpreting the factor-analytic results. Third, factor analysis of ordinal NQ items technically violates the continuous-variable assumption of principal axis factoring; polychoric correlation-based factoring would be methodologically preferable and is recommended for future studies. Fourth, blinding of participants was not possible given the nature of the interventions, although assessors were blinded to group allocation. Fifth, the absence of an isolated breathing-specific intervention arm limits conclusions about which active components of breathing re-education drive the observed NQ change. Sixth, missing outcome values were handled by single mean imputation within the allocated group under the intention-to-treat principle, consistent with the parent trial [19], where the limitations of that approach are discussed in full. Seventh, the Bonferroni correction across 20 non-independent correlations is conservative; uncorrected p-values were also reported. Eighth, screening prevalence varies markedly with the choice of cut-off: 41.0% at >23, rising to 68.0% at >17—a substantial variation with direct clinical implications for patient identification and resource allocation in this population. No ROC-based validation has been performed in NSCNP, and the existing cut-off evidence derives primarily from unselected physiotherapy referrals and hyperventilation clinic samples [12]. Until a validated threshold specific to NSCNP is established, the ≥19 threshold [12] is recommended as the most defensible current option; future work should perform ROC analysis against a clinician-rated criterion standard to determine optimal sensitivity and specificity in this

population. Finally, single mean imputation may attenuate variance, though the missing data were confined to 3 of 30 participants (10%) in one arm, with complete data in the other two, limiting the likely impact.

5. Conclusions

The NQ shows preliminary psychometric support for use as a single-score measure of psychophysiological functional respiratory complaints in patients with non-specific chronic neck pain, with acceptable total-score internal consistency reliability but inadequate subscale validity. Therefore, only the NQ total score should be used in NSCNP populations, since the subscales are not psychometrically viable. The NQ total score demonstrates moderate convergent validity with disability-related and anxiety-related constructs, and consistent non-significant associations with all objective respiratory measures (divergent validity pattern), including biochemical, biomechanical, cervical spine range of motion and static posture parameters, supporting its use as a measure of the psychophysiological dimension of dysfunctional breathing within Courtney's multidimensional framework. Measures of other dimensions of DB require independent assessment.

As a treatment outcome measure, the NQ is responsive, registering change in the two manual therapy arms that exceeded its own minimal detectable change by a factor of two to three, while the change recorded under conventional physiotherapy remained within measurement error. Users should therefore interpret individual change against an $MDC_{95\%}$ of approximately five points, and should not infer change from scores differing by less than that. The comparative clinical effectiveness of the interventions studied is reported separately [19].

Author Contributions: Conceptualization, P.I.T., E.G., Z.D. and G.A.K.; methodology, P.I.T., E.G., Z.D. and G.A.K.; validation, P.I.T., E.G., Z.D. and G.A.K.; formal analysis, P.I.T., E.G., Z.D. and G.A.K.; investigation, P.I.T., E.G., Z.D. and G.A.K.; resources, P.I.T., E.G. and G.A.K.; data curation, P.I.T., E.G., Z.D. and G.A.K.; writing—original draft preparation, P.I.T., E.G., Z.D. and G.A.K.; writing—review and editing, P.I.T., E.G., Z.D. and G.A.K.; visualization, P.I.T. and G.A.K.; supervision, E.G., Z.D. and G.A.K.; project administration, P.I.T., E.G., Z.D. and G.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the University of West Attica, Athens, Greece (approval no: 51758; date of approval: 1 June 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to the applicable data protection law in Greece (Law 4624/2019).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BHT	Breath-holding time
CG	Control Group
CS-ROM	Cervical Spine Range of Motion
CVA	Craniovertebral Angle
CWE	Chest wall expansion
EG	Experimental Group

$ETCO_2$	End-tidal carbon dioxide
HADS-A	Hospital Anxiety and Depression Scale-Anxiety subscale
HADS-D	Hospital Anxiety and Depression Scale-Depression subscale
NDI	Neck Disability Index
NQ	Nijmegen Questionnaire
NSCNP	Non-Specific Chronic Neck Pain
PI-NRS	Pain intensity numeric rating scale
RR	Respiratory rate
SBC	Single breath count
TSK	Tampa Scale of Kinesiophobia

References

1. GBD 2021 Neck Pain Collaborators. Global, regional, and national burden of neck pain, 1990–2020, and projections to 2050: A systematic analysis of the Global Burden of Disease Study 2021. *Lancet Rheumatol.* **2024**, *6*, e142–e155. [CrossRef] [PubMed]
2. Kapreli, E.; Vourazanis, E.; Billis, E.; Oldham, J.A.; Strimpakos, N. Respiratory dysfunction in chronic neck pain patients. A pilot study. *Cephalalgia* **2009**, *29*, 701–710. [CrossRef] [PubMed]
3. Dimitriadis, Z.; Kapreli, E.; Strimpakos, N.; Oldham, J. Respiratory weakness in patients with chronic neck pain. *Man. Ther.* **2013**, *18*, 248–253. [CrossRef] [PubMed]
4. McLaughlin, L.; Goldsmith, C.H.; Coleman, K. Breathing evaluation and retraining as an adjunct to manual therapy. *Man. Ther.* **2011**, *16*, 51–52. [CrossRef] [PubMed]
5. McLaughlin, L. Breathing evaluation and retraining in manual therapy. *J. Bodyw. Mov. Ther.* **2009**, *13*, 276–282. [CrossRef] [PubMed]
6. Boulding, R.; Stacey, R.; Niven, R.; Fowler, S.J. Dysfunctional breathing: A review of the literature and proposal for classification. *Eur. Respir. Rev.* **2016**, *25*, 287–294. [CrossRef] [PubMed]
7. Courtney, R.; Greenwood, K.M.; Cohen, M. Relationships between measures of dysfunctional breathing in a population with concerns about their breathing. *J. Bodyw. Mov. Ther.* **2011**, *15*, 24–34. [CrossRef] [PubMed]
8. Courtney, R. A multi-dimensional model of dysfunctional breathing and integrative breathing therapy—Commentary on the functions of breathing and its dysfunctions and their relationship to breathing therapy. *J. Yoga Phys. Ther.* **2016**, *6*, 257.
9. Vlemincx, E. Dysfunctional breathing: A dimensional, transdiagnostic perspective. *Eur. Respir. J.* **2023**, *61*, 2300629. [CrossRef] [PubMed]
10. Courtney, R. Breathing training for dysfunctional breathing in asthma: Taking a multidimensional approach. *ERJ Open Res.* **2017**, *3*, 00065-2017. [CrossRef] [PubMed]
11. Van Dixhoorn, J.; Duivenvoorden, H.J. Efficacy of Nijmegen Questionnaire in recognition of the hyperventilation syndrome. *J. Psychosom. Res.* **1985**, *29*, 199–206. [CrossRef] [PubMed]
12. Van Dixhoorn, J.; Folgering, H. The Nijmegen Questionnaire and dysfunctional breathing. *ERJ Open Res.* **2015**, *1*, 00001-2015. [CrossRef] [PubMed]
13. Li Ogilvie, V.; Kersten, P. A critical review of the psychometric properties of the Nijmegen Questionnaire for dysfunctional breathing. *N. Z. J. Physiother.* **2015**, *43*, 3–10. [CrossRef]
14. Li Ogilvie, V.; Kayes, N.M.; Kersten, P. The Nijmegen Questionnaire: A valid measure for hyperventilation syndrome. *N. Z. J. Physiother.* **2019**, *47*, 160–171. [CrossRef]
15. Stephen, S.; Brandt, C.; Olivier, B. Neck pain and disability: Are they related to dysfunctional breathing and stress? *Physiother. Can.* **2022**, *74*, 158–164. [CrossRef] [PubMed]
16. Deshmukh, M.P.; Palekar, T.J.; Manvikar, N.; Makker, K.S.; Mali, C. Prevalence of breathing dysfunction in musculoskeletal complaints: A cross-sectional study. *Int. J. Health Sci. Res.* **2022**, *12*, 146–152. [CrossRef]
17. Perri, M.A.; Halford, E. Pain and faulty breathing: A pilot study. *J. Bodyw. Mov. Ther.* **2004**, *8*, 297–306. [CrossRef]
18. López-de-Uralde-Villanueva, I.; Del Corral, T.; Salvador-Sánchez, R.; Angulo-Díaz-Parreño, S.; López-Marcos, J.J.; Plaza-Manzano, G. Respiratory dysfunction in patients with chronic neck pain: Systematic review and meta-analysis. *Disabil. Rehabil.* **2023**, *45*, 2422–2433. [CrossRef] [PubMed]
19. Tatsios, P.I.; Grammatopoulou, E.; Dimitriadis, Z.; Koumantakis, G.A. Effectiveness of cervical spine and diaphragm manual therapy combined with breathing re-education in non-specific chronic neck pain: A randomized controlled trial. *J. Clin. Med.* **2026**, *15*, 4266. [CrossRef] [PubMed]
20. Tatsios, P.I.; Grammatopoulou, E.; Dimitriadis, Z.; Patsaki, I.; Gioftsos, G.; Koumantakis, G.A. Cervical spine range of motion reliability with two methods and associations with demographics, forward head posture, and respiratory mechanics in patients with non-specific chronic neck pain. *J. Funct. Morphol. Kinesiol.* **2025**, *10*, 269. [CrossRef] [PubMed]

21. Tatsios, P.I.; Grammatopoulou, E.; Dimitriadis, Z.; Koumantakis, G.A. Manual therapy in the cervical spine and diaphragm, combined with breathing re-education in patients with non-specific chronic neck pain: Protocol for a randomized controlled trial. *Diagnostics* **2022**, *12*, 2690. [CrossRef] [PubMed]
22. Gorsuch, R.L. *Factor Analysis*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1983.
23. MacCallum, R.C.; Widaman, K.F.; Zhang, S.; Hong, S. Sample size in factor analysis. *Psychol. Methods* **1999**, *4*, 84–99. [CrossRef]
24. Tatsios, P.I.; Grammatopoulou, E.; Dimitriadis, Z.; Koumantakis, G.A. Effectiveness of manual therapy in the cervical spine and diaphragm with breathing re-education on range of motion and forward head posture in non-specific chronic neck pain: A randomized controlled trial. *Healthcare* **2025**, *13*, 1765. [CrossRef] [PubMed]
25. Grammatopoulou, E.P.; Skordilis, E.K.; Georgoudis, G.; Haniotou, A.; Evangelodimou, A.; Fildissis, G.; Katsoulas, T.; Kalagiakos, P. Hyperventilation in asthma: A validation study of the Nijmegen Questionnaire—NQ. *J. Asthma* **2014**, *51*, 839–846. [CrossRef] [PubMed]
26. Courtney, R.; Cohen, M. Investigating the claims of Konstantin Buteyko, M.D., Ph.D.: The relationship of breath holding time to end-tidal CO_2 and other proposed measures of dysfunctional breathing. *J. Altern. Complement. Med.* **2008**, *14*, 115–123. [CrossRef] [PubMed]
27. Euasobhon, P.; Atisook, R.; Bumrungchatudom, K.; Zinboonyahgoon, N.; Saisavoey, N.; Jensen, M.P. Reliability and responsivity of pain intensity scales in individuals with chronic pain. *Pain* **2022**, *163*, e1184–e1191. [CrossRef] [PubMed]
28. Vernon, H.; Mior, S. The Neck Disability Index: A study of reliability and validity. *J. Manip. Physiol. Ther.* **1991**, *14*, 409–415.
29. Trouli, M.N.; Vernon, H.T.; Kakavelakis, K.N.; Antonopoulou, M.D.; Paganas, A.N.; Lionis, C.D. Translation of the Neck Disability Index and validation of the Greek version in a sample of neck pain patients. *BMC Musculoskelet. Disord.* **2008**, *9*, 106. [CrossRef] [PubMed]
30. Zigmond, A.S.; Snaith, R.P. The Hospital Anxiety and Depression Scale. *Acta Psychiatr. Scand.* **1983**, *67*, 361–370. [CrossRef] [PubMed]
31. Michopoulos, I.; Douzenis, A.; Kalkavoura, C.; Christodoulou, C.; Michalopoulou, P.; Kalemi, G.; Fineti, K.; Patapis, P.; Protopapas, K.; Lykouras, L. Hospital Anxiety and Depression Scale (HADS): Validation in a Greek general hospital sample. *Ann. Gen. Psychiatry* **2008**, *7*, 4. [CrossRef] [PubMed]
32. Dupuis, F.; Cherif, A.; Batcho, C.; Massé-Alarie, H.; Roy, J.S. The Tampa Scale of Kinesiophobia: A systematic review of its psychometric properties in people with musculoskeletal pain. *Clin. J. Pain* **2023**, *39*, 236–247. [CrossRef] [PubMed]
33. Georgoudis, G.; Raptis, K.; Koutserimpas, C. Cognitive assessment of musculoskeletal pain: Validity and reliability of the Greek version of the Tampa Scale of Kinesiophobia in patients suffering from chronic low back pain. *Maedica* **2022**, *17*, 826–832. [CrossRef] [PubMed]
34. Field, A. Exploratory factor analysis. In *Discovering Statistics Using IBM SPSS Statistics*, 5th ed.; Sage: London, UK, 2015; pp. 778–834.
35. Horn, J.L. A rationale and test for the number of factors in factor analysis. *Psychometrika* **1965**, *30*, 179–185. [CrossRef] [PubMed]
36. O'Connor, B.P. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods Instrum. Comput.* **2000**, *32*, 396–402. [CrossRef] [PubMed]
37. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, 6th ed.; Pearson: Boston, MA, USA, 2013.
38. Portney, L.G.; Watkins, M.P. *Foundations of Clinical Research: Applications to Practice*, 3rd ed.; Pearson Education Ltd.: Harlow, UK, 2014.
39. McDonald, R.P. *Test Theory: A Unified Treatment*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1999.
40. Zinbarg, R.E.; Revelle, W.; Yovel, I.; Li, W. Cronbach's α, Revelle's β, and McDonald's ω_h: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* **2005**, *70*, 123–133. [CrossRef]
41. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef] [PubMed]
42. Terwee, C.B.; Bot, S.D.; de Boer, M.R.; van der Windt, D.A.; Knol, D.L.; Dekker, J.; Bouter, L.M.; de Vet, H.C. Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* **2007**, *60*, 34–42. [CrossRef] [PubMed]
43. Hulley, S.B.; Cummings, S.R.; Browner, W.S.; Grady, D.; Newman, T.B. *Designing Clinical Research: An Epidemiologic Approach*, 4th ed.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2013; p. 79.
44. Bland, J.M.; Altman, D.G. Multiple significance tests: The Bonferroni method. *BMJ* **1995**, *310*, 170. [CrossRef] [PubMed]
45. Kaiser, H.F. An index of factorial simplicity. *Psychometrika* **1974**, *39*, 31–36. [CrossRef]
46. Steinmann, J.; Lewis, A.; Ellmers, T.J.; Jones, M.; MacBean, V.; Kal, E. Validating the Breathing Vigilance Questionnaire for use in dysfunctional breathing. *Eur. Respir. J.* **2023**, *61*, 2300031. [CrossRef] [PubMed]
47. Milne, K.M.; Zhang, J.; Harris, O.D.; Guenette, J.A. Sex differences in the multidimensional evaluation of dyspnea in respiratory outpatients. *Front. Med.* **2025**, *12*, 1627496. [CrossRef] [PubMed]
48. LoMauro, A.; Aliverti, A. Sex differences in respiratory function. *Breathe* **2018**, *14*, 131–140. [CrossRef] [PubMed]

49. Asselmann, E.; Venz, J.; Pieper, L.; Wittchen, H.U.; Pittrow, D.; Beesdo-Baum, K. The role of gender and anxiety in the association between somatic diseases and depression. *Epidemiol. Psychiatr. Sci.* **2019**, *28*, 321–332. [CrossRef] [PubMed]
50. Bruton, A.; Lee, A.; Yardley, L.; Raftery, J.; Arden-Close, E.; Kirby, S.; Zhu, S.; Thiruvothiyur, M.; Webley, F.; Taylor, L.; et al. Physiotherapy breathing retraining for asthma: A randomised controlled trial. *Lancet Respir. Med.* **2018**, *6*, 19–28. [CrossRef] [PubMed]
51. Holloway, E.A.; West, R.J. Integrated breathing and relaxation training (the Papworth method) for adults with asthma in primary care: A randomised controlled trial. *Thorax* **2007**, *62*, 1039–1042. [CrossRef] [PubMed]
52. Grammatopoulou, E.P.; Skordilis, E.K.; Stavrou, N.; Myrianthefs, P.; Karteroliotis, K.; Baltopoulos, G.; Koutsouki, D. The effect of physiotherapy-based breathing retraining on asthma control. *J. Asthma* **2011**, *48*, 593–601. [CrossRef] [PubMed]
53. Mundfrom, D.J.; Shaw, D.G.; Ke, T.L. Minimum sample size recommendations for conducting factor analyses. *Int. J. Test.* **2005**, *5*, 159–168. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.