

Development and content validation of the Nursing Practice Environment Scale for Long-Term Care: a methodological study in Portugal

Vera Braga

vbraga@uat.la.pt

Escola Superior de Saúde Atlântica

Ana Ramos

Escola Superior de Enfermagem da Universidade de Lisboa

Ricardo Braga

Unidade Local de Saúde de São José

Helena José

Escola Superior de Saúde Atlântica

Research Article

Keywords: nursing practice environment, long-term care, instrument development, content validity, nursing management, psychometrics

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10813959/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Nursing practice environments influence care quality, professional outcomes and service sustainability. Measures developed in acute hospitals may not adequately represent the organisational interdependence, longitudinal care and integration required in long-term care (LTC). This study aimed to develop the Nursing Practice Environment Scale for Long-Term Care (NPES-LTC) and assess its content validity, comprehensibility and preliminary item behaviour in Portugal.

Methods

This methodological instrument-development study comprised conceptual clarification and attribute mapping, deductive item generation, content validation by 13 experts and a fully electronic pre-test with a final open-ended feedback question. An initial 56-item pool was informed by the Magnet Model and a World Health Organization framework for integrated LTC. Experts independently assessed item relevance and clarity. Item- and scale-level content validity indices, Krippendorff's alpha and Fleiss' kappa were calculated. Of 45 nurses invited to pre-testing, 39 participated (86.7%). Corrected item-total correlations, Cronbach's alpha and McDonald's omega were used only as exploratory indicators of item behaviour.

Results

For the initial pool, the scale-level content validity index was 0.93 for relevance and 0.89 for clarity. Thirteen items below the predefined 0.80 criterion were removed. Among the 43 retained items, the corresponding indices were 0.96 and 0.93. Agreement was high for relevance (Krippendorff's $\alpha = 0.83$; Fleiss' $\kappa = 0.82$) and clarity ($\alpha = 0.81$; $\kappa = 0.80$). Exploratory domain estimates ranged from 0.81 to 0.90 for alpha and 0.78 to 0.95 for omega; the 43-item scale yielded $\alpha = 0.96$ and $\omega = 0.97$. One item was removed and one negatively worded item was retained with reverse scoring, producing a 42-item candidate version.

Conclusions

The expert assessment supported the relevance and clarity of the NPES-LTC item pool, while pre-testing informed comprehensibility and item behaviour. The 42-item candidate version is theoretically grounded and context-sensitive, but its comprehensiveness, four-domain structure, revised item and scoring model require further cognitive and large-scale psychometric evaluation before research, benchmarking or organisational use.

Background

Population ageing, multimorbidity, functional dependency and longer care trajectories are increasing demand for long-term care (LTC) internationally [1–3]. Unlike episodic acute care, LTC depends on continuity, coordination across health and social sectors, sustained therapeutic relationships and the capacity to respond to changing health and social needs over time [1, 4]. These features make the organisation of nursing work particularly consequential.

The nursing practice environment (NPE) comprises the organisational characteristics that facilitate or constrain professional nursing practice [5, 6]. Favourable environments are generally characterised by effective nursing leadership, nurse participation in organisational affairs, adequate staffing and resources, strong interprofessional relationships and support for quality improvement [5–8]. Evidence from hospital and residential settings links favourable NPEs with improved nurse retention and job satisfaction, lower burnout and missed care, and better safety and patient outcomes [7–11].

The Practice Environment Scale of the Nursing Work Index (PES-NWI) remains the most widely used measure of the construct [5]. Portuguese adaptations of the PES-NWI, the Nursing Work Index-Revised and the Indicators of Quality Nursing Work Environments have strengthened assessment in hospital and primary care contexts [12–14]. Nevertheless, instruments originating in acute care may not fully represent LTC-specific conditions, including prolonged temporality of care, integration of health and social responses, community and inter-institutional coordination, regulatory and financing arrangements, staff retention, career development and the participation of residents and families [1, 4, 15–18]. Using a measure that omits contextually important attributes risks reducing the NPE to what existing instruments happen to measure.

From a nursing disciplinary perspective, the practice environment is not a neutral organisational backdrop. It conditions nurses' clinical judgement, professional agency, relational continuity and capacity to deliver person-centred, ethically responsible care [19, 20]. Within Donabedian's structure-process-outcome logic, the NPE can be understood as an organisational condition that shapes care processes and, through them, outcomes for people, professionals and services [21]. A context-sensitive measure can therefore support nursing management and governance by making modifiable conditions of practice visible.

This study aimed to develop the Nursing Practice Environment Scale for Long-Term Care (NPES-LTC) and assess its content validity, response processes and preliminary item behaviour for use in Portugal.

Methods

Study design

A methodological instrument-development study was conducted in four sequential stages: (1) conceptual clarification of the construct and identification of context-specific attributes; (2) deductive item generation; (3) content validation by an expert panel; and (4) pre-testing with nurses working in LTC. The NPES-LTC was developed within the academic research programme underlying the present

methodological study. The initial instrument-development and content-validation work was undertaken as part of a Master's dissertation completed in 2024 and deposited in a Portuguese institutional repository [33]. Although the repository record is publicly available, the full text remains under restricted access, and the questionnaire has not previously been published as a publicly accessible, psychometrically validated instrument. The process followed established recommendations for health measurement instrument development, including the sequential construction and content-validation method described by Coluci et al. [22–25], and was reported with reference to the COSMIN Reporting Guideline 2.0, adapted to a professional-reported organisational measure rather than a patient-reported outcome measure [31, 32]. This study addressed only the initial development phase of the NPES-LTC: construct clarification, item generation, expert-based content validation and pre-testing with the target population. It was not designed as a complete psychometric validation study. Structural validity, internal consistency based on an empirically supported dimensional structure, test-retest reliability, measurement error, construct validity and measurement invariance were outside the scope of this phase and will be evaluated in a subsequent multicentre study. The study was completed in 2024.

Conceptual clarification and attribute mapping

The construct was initially examined through scientific and normative literature concerning NPEs, quality nursing work environments and LTC. The 14 Forces of Magnetism and the later Magnet Model components—transformational leadership, structural empowerment, exemplary professional practice, new knowledge and innovation, and empirical outcomes—provided an organisational nursing reference [26]. These components were compared with established NPE instruments and with instruments previously adapted to Portuguese nursing contexts [12–14].

The comparison identified attributes that were incompletely represented in commonly used measures, particularly community engagement, regulatory conditions, financing, career progression, professional recognition, workforce retention, integration of care, organisational learning and the relationship between nursing governance and system-level conditions.

To ensure contextual relevance beyond the organisational level, the identified attributes were aligned with the World Health Organization Regional Office for Europe country assessment framework for the integrated delivery of long-term care [1]. This framework organises assessment around health and social care needs, performance, service delivery and system enablers. These components informed four a priori theoretical categories for item coverage: system facilitators, service delivery, performance, and health and social needs. They were not considered empirically established factors.

Item generation and response format

Items were generated deductively from the conceptual mapping. Each proposed item was traced to a relevant Magnet component, an LTC attribute and one of the four theoretical domains. The initial pool

contained 56 items spanning individual nursing practice, team and service processes, organisational governance and wider regulatory conditions.

Items were written as declarative statements and organised from the professional and service levels to the organisational and system levels. The intended response format was a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The initial wording was reviewed for redundancy, specificity and applicability to Portuguese LTC. Respondents were instructed to rate their current nursing practice environment; no fixed retrospective recall period was specified.

The 42-item candidate version is scored from 1 (strongly disagree) to 5 (strongly agree). Items are assigned theoretically to system facilitators (11 items; possible score 11–55), service delivery (11 items; 11–55), performance (12 items; 12–60), and health and social needs (8 items; 8–40). A provisional total score ranges from 42 to 210, with higher scores intended to indicate a more favourable nursing practice environment. One negatively worded item concerning whether part-time versus full-time contractual arrangements constrain care quality requires reverse scoring. Domain and total scores remain provisional because the dimensional structure has not been empirically established, and no interpretative cut-off points have been defined.

The complete 42-item candidate NPES-LTC is provided in Supplementary File 1, including the Portuguese study wording, an English-language reporting translation, five-point response anchors, theoretical domain assignment, reverse-scoring instructions and provisional scoring guidance. This supplementary version is provided to ensure accessibility of the questionnaire described in the present study. The English-language translation has not undergone a separate cross-cultural adaptation or validation process.

Expert panel and content-validation procedure

A purposive panel of experts was recruited according to recognised indicators of nursing expertise, including relevant specialist knowledge, demonstrated experience and professional authority [27]. Eligibility required experience in LTC, nursing management, specialist nursing, education or research relevant to the construct. Thirteen experts participated.

Experts were contacted by email and received a link to a fully electronic, self-administered questionnaire created in Microsoft Forms using the institutional account. Data collection occurred during the first 15 working days of January 2024. One courteous reminder was emailed seven days after the initial invitation to those who had not yet responded. Experts independently evaluated each item for relevance and clarity using four-point ordinal scales. Scores of 3 or 4 were considered endorsement. Space was provided for written comments and suggestions. Responses were collected remotely and anonymously. A second round had been planned if the predefined consensus criteria were not achieved; because the retained pool exceeded the criteria after the first assessment and comments could be incorporated

without substantive disagreement, no further round was undertaken. Accordingly, the procedure is reported as an expert content-validation panel rather than as a completed multi-round Delphi study.

The item-level content validity index (I-CVI) was calculated as the proportion of experts assigning a score of 3 or 4. An I-CVI of at least 0.80 for both relevance and clarity was the predefined retention criterion. The scale-level content validity index based on the average method (S-CVI/Ave) was calculated as the mean of the I-CVIs. Items below the criterion for either property were removed or considered for reformulation. Agreement beyond raw endorsement was examined after dichotomisation using Krippendorff's alpha and Fleiss' kappa.

Comprehensiveness was not assessed as a separate quantitative criterion; experts could propose omitted content in their written comments. This aspect was therefore treated as exploratory rather than as established evidence of content validity.

No hypothesis-based sample-size calculation was performed. The purposive panel size was determined by the developmental purpose of the study and the need to include complementary expertise in LTC, nursing management, specialist nursing, education and research, in accordance with the adopted instrument-development method [22, 27].

Pre-test and comprehensibility appraisal

The 43-item version resulting from content validation was administered to a convenience sample of registered nurses working in Portuguese integrated long-term care services. No minimum duration of LTC experience was required. Forty-five nurses were contacted by email and invited to complete a fully electronic, self-administered questionnaire created in Microsoft Forms using the institutional account. Data collection occurred during the first 15 working days of April 2024. One courteous reminder was emailed seven days after the initial invitation to those who had not yet responded. Of the 45 nurses invited, 39 completed the pre-test, corresponding to a participation rate of 86.7%. Participants represented districts across mainland Portugal. The pre-test was intended to explore comprehensibility and preliminary item behaviour, not to establish structural validity or definitive reliability.

At the end of the Microsoft Forms questionnaire, participants answered an open-ended question inviting them to report difficulties, ambiguities and suggestions concerning the instructions, items, response options and overall completion experience. Written feedback was considered descriptively alongside corrected item-total correlations; no formal qualitative coding was undertaken. Because the sample was small relative to the number of items, no exploratory or confirmatory factor analysis was performed.

No hypothesis-based sample-size calculation was undertaken for the pre-test. The sample was pragmatic and intended to explore comprehensibility, response processes and preliminary item behaviour. It was not considered sufficient for factor analysis or definitive psychometric evaluation.

Statistical analysis

Descriptive statistics were used to characterise experts and pre-test participants. Content validity was summarised using I-CVI and S-CVI/Ave. Inter-expert agreement was assessed using Krippendorff's alpha and Fleiss' kappa, complementary coefficients commonly used to evaluate agreement among multiple raters beyond that expected by chance [28–29]. Preliminary internal consistency was examined for each a priori domain and the total scale using corrected item-total correlations, Cronbach's alpha and McDonald's omega. A corrected item-total correlation of at least 0.30 and reliability coefficients of at least 0.70 were used as interpretive reference values [23–25]. Alpha and omega values reported for individual items represent the coefficient if that item were deleted.

All 39 nurses included in the pre-test completed all 43 items. No item-level data were missing and no imputation was required. Alpha and omega were calculated solely as exploratory indicators of the behaviour of the theoretically grouped items. Because dimensionality had not been empirically established, these coefficients were not interpreted as evidence of internal consistency reliability according to COSMIN standards.

Analyses were performed using IBM SPSS Statistics, version 29.0 (IBM Corp., Armonk, NY, USA), and R statistical software, version 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria).

Ethical considerations

The study was conducted in accordance with the Declaration of Helsinki and received a favourable opinion from the Ethics Council of the Nursing School of Lisbon (Escola Superior de Enfermagem de Lisboa), process no. 4147/2023. Before participation, experts and nurses received written information explaining the study objectives, procedures, voluntary nature of participation, confidentiality safeguards and right to withdraw at any time without penalty. All participants had capacity to consent and provided written informed consent before data collection. Data were pseudonymised and stored electronically in a password-protected database accessible only to the research team. In accordance with the approved protocol, data will be retained for five years after study completion and then securely destroyed.

Results

Expert characteristics

The 13 experts were predominantly women ($n = 12$; 92.3%). Their mean age was 54.2 years (SD 11.2; range 35–66), and mean professional experience was 32.0 years (SD 11.5; range 12–46). They had a mean of 10.9 years of LTC experience (SD 6.0; range 2–20). Specialist nurses and nurse managers were the largest professional groups (each $n = 4$; 30.8%). Detailed characteristics are presented in Table 1.

Table 1
Characteristics of the expert panel (n = 13)

Characteristic	Result
Age, mean (SD), range	54.2 (11.2), 35–66
Women, n (%)	12 (92.3)
Total professional experience, mean years (SD), range	32.0 (11.5), 12–46
LTC experience, mean years (SD), range	10.9 (6.0), 2–20
Registered nurse, n (%)	1 (7.7)
Specialist nurse, n (%)	4 (30.8)
Head nurse, n (%)	1 (7.7)
Nurse coordinator, n (%)	2 (15.4)
Nurse manager, n (%)	4 (30.8)
Nurse educator, n (%)	1 (7.7)
Bachelor's degree, n (%)	6 (46.2)
Postgraduate qualification, n (%)	4 (30.8)
Master's degree, n (%)	2 (15.4)
Doctoral degree, n (%)	1 (7.7)

Content validity and expert agreement

Across the initial 56-item pool, S-CVI/Ave was 0.93 for relevance and 0.89 for clarity. Thirteen items did not reach 0.80 for relevance and/or clarity and were removed. They concerned interdisciplinary interface wording, monitoring of client complexity, safe staffing, management of turbulence during organisational change, an innovative professional environment, institutional partnerships, implementation of strategic plans, organisational recognition and reward, professional differentiation, public financing and the role of professional and union bodies.

The retained 43-item pool achieved an S-CVI/Ave of 0.96 for relevance and 0.93 for clarity. Agreement indices were at or above 0.80 for both properties: Krippendorff's α was 0.83 for relevance and 0.81 for clarity; Fleiss' κ was 0.82 and 0.80, respectively (Table 2).

No separate quantitative judgement of comprehensiveness was obtained.

Table 2
Content validity and inter-expert agreement

Measure	Relevance	Clarity
S-CVI/Ave, initial 56-item pool	0.93	0.89
S-CVI/Ave, retained 43-item pool	0.96	0.93
Krippendorff's α , retained pool	0.83	0.81
Fleiss' κ , retained pool	0.82	0.80

Pre-test participant characteristics

Of the 45 nurses invited, 39 participated (86.7%). They represented 12 districts across mainland Portugal; Lisbon contributed the largest subgroup ($n = 14$; 35.9%). Most participants were women ($n = 30$; 76.9%). Mean age was 37.2 years (SD 10.2; range 22–64), mean professional experience was 13.8 years (SD 10.5; range 1–40), and mean LTC experience was 6.4 years (SD 5.0; range 0.3–18). Most were registered nurses without a specialist title ($n = 25$; 64.1%), followed by specialist nurses ($n = 7$; 17.9%); 30 (76.9%) held a bachelor's degree. Completion time ranged from 3 to 7 minutes. All 39 participants completed every item, and no item-level data were missing.

Pre-test feedback and preliminary item behaviour

Participants' written responses to the final open-ended question were reviewed descriptively to identify apparent instructional, wording or response-option difficulties and suggestions for improvement. Because these responses were not formally coded, no qualitative frequencies or thematic findings are claimed. For the four a priori domains, Cronbach's alpha ranged from 0.81 to 0.90 and McDonald's omega from 0.78 to 0.95. For the complete 43-item version, α was 0.96 and ω was 0.97 (Table 3). These estimates were calculated solely to explore preliminary item behaviour and were not interpreted as evidence of reliability because the dimensional structure had not been empirically tested and the sample was small.

Table 3
Preliminary internal consistency of the theoretically assigned domains
($n = 39$)

Theoretical domain	Items, n	Cronbach's α	McDonald's ω
System facilitators	11	0.83	0.95
Service delivery	11	0.90	0.90
Performance	12	0.81	0.78
Health and social needs	9	0.84	0.80
Total preliminary scale	43	0.96	0.97

Most corrected item-total correlations exceeded 0.30. The item stating that part-time versus full-time contractual arrangements constrained care quality showed a negative corrected item-total correlation within its assigned domain (-0.18) and in the total scale (-0.07). Because the underlying attribute was considered theoretically relevant, the negatively worded item was provisionally retained and designated for reverse scoring in the candidate version. This decision requires confirmation through renewed cognitive assessment and independent psychometric evaluation. The item concerning persistent notification by the Portuguese nursing regulator to employers and regulatory authorities had low corrected item-total correlations (0.22 within its domain and 0.19 for the total scale) and was removed. The process produced a 42-item candidate NPES-LTC assigned theoretically to system facilitators, service delivery, performance, and health and social needs.

Table 4
Provisional composition and scoring of the 42-item
candidate NPES-LTC

Theoretical domain	Items, n	Possible score
System facilitators	11	11–55
Service delivery	11	11–55
Performance	12	12–60
Health and social needs	8	8–40
Provisional total score	42	42–210

Items are scored from 1 (strongly disagree) to 5 (strongly agree). One negatively worded item concerning contractual arrangements requires reverse scoring.

Higher scores are intended to represent a more favourable nursing practice environment. For descriptive interpretation, the mean item score may be calculated after reverse-scoring the negatively worded contractual-arrangements item. A mean score > 3.0 is provisionally interpreted as indicating a positive or favourable nursing practice environment; a score of 3.0 corresponds to the neutral midpoint of the response scale, whereas a mean score < 3.0 indicates a non-positive or unfavourable environment. For the 42-item total scale, a mean score > 3.0 corresponds to a raw total score > 126. These interpretative thresholds are a priori descriptive criteria based on the midpoint of the response scale and should not be regarded as empirically validated diagnostic or clinical cut-off values. The domain composition, scoring model and interpretative criteria remain provisional pending structural and construct validation.

Discussion

This study developed a 42-item candidate measure intended to represent organisational and system conditions of nursing practice in LTC. Expert assessment supported the relevance and clarity of the retained item pool, while the electronic pre-test and exploratory quantitative analysis provided

preliminary information on comprehensibility and item behaviour. The study therefore establishes an initial content-valid candidate version rather than a fully validated psychometric instrument.

The contribution of the NPES-LTC lies first in its contextual specificity. Established NPE instruments have generated extensive evidence and remain valuable for comparison across settings [5–8]. However, their acute-care origins may leave LTC attributes less visible. The present scale extends assessment to continuity and integration of care, health and social needs, workforce development and retention, organisational monitoring, professional regulation and wider system facilitators.

These additions are consistent with international descriptions of LTC as a continuum requiring coordination across professional, organisational and policy boundaries [1, 4].

Second, the scale was built through theoretical traceability rather than by generating an undifferentiated list of workplace characteristics. Linking Magnet principles with an integrated LTC framework made it possible to connect micro-level nursing practice with meso-level organisational governance and macro-level regulation. This is relevant because nurses' capacity to provide person-centred, evidence-informed and ethically responsible care depends not only on individual competence but also on the conditions under which professional action occurs [19–21].

The initial content-validity results also illustrate the importance of assessing clarity separately from relevance. Several items were judged relevant but insufficiently clear, including statements concerning client complexity, organisational innovation, strategic partnerships, professional recognition and financing. Their removal reduced conceptual breadth in some areas but protected interpretability at this developmental stage. Future qualitative work may permit clearer replacement items where the attribute remains theoretically important.

Agreement estimates complemented the CVI by examining consistency beyond simple endorsement. Values of 0.80 or above across both Krippendorff's alpha and Fleiss' kappa support the homogeneity of expert judgements. Nevertheless, these coefficients do not establish construct validity. Similarly, the high total-scale alpha and omega should not be interpreted as proof of unidimensionality or validity. With 43 items, very high coefficients may partly reflect test length and overlapping content [23–25]. The domain-level coefficients are encouraging but remain conditional on theoretical item assignment.

The negative item-total behaviour of the contractual-regime item may have resulted from its direction, ambiguity or variation in participants' employment contexts. Retaining it as a reverse-scored item preserves the attribute but requires formal cognitive and empirical testing in an independent sample; reverse wording can itself introduce method effects. The removal of the regulator-notification item suggests that respondents may have had limited knowledge of macro-level regulatory activity or that the item was too remote from their observable practice environment. This finding reinforces the need to distinguish organisational conditions nurses can reliably report from wider system conditions that may require other data sources.

Implications for nursing management and research

After further validation, the NPES-LTC may help nurse managers identify modifiable conditions affecting professional practice, compare patterns across services and guide improvement priorities. Its multilevel coverage may also support governance discussions connecting staffing, leadership, learning, safety, continuity, regulation and organisational outcomes. It should not yet be used for high-stakes benchmarking or performance assessment.

The next validation phase should recruit a substantially larger, multicentre sample representing different LTC models, regions and organisational sectors. Structural validity should be examined using factor analysis appropriate to ordinal data. Competing models should be tested rather than assuming the four theoretical domains. Subsequent analyses should evaluate internal consistency conditional on the supported factor structure, convergent and discriminant validity, known-groups validity, test-retest reliability, measurement error and invariance across relevant groups. Formal cognitive interviewing should precede large-scale testing of the reverse-scored contractual item.

Strengths and limitations

Strengths include the explicit theoretical mapping, a multidisciplinary expert panel with extensive professional and LTC experience, separate assessment of relevance and clarity, use of complementary agreement indices and pre-testing with nurses from multiple Portuguese districts.

Several limitations require emphasis. Item generation was predominantly deductive and did not include qualitative concept elicitation with the target population; aspects experienced by LTC nurses may therefore not have been fully captured. Comprehensiveness was not formally rated as a distinct content-validity component, and the opportunity to provide written comments cannot substitute for a structured assessment of omitted concepts. The expert panel and pre-test samples were purposive or convenience samples and may not represent all Portuguese LTC contexts. Only one expert-assessment round was completed; although the predefined criteria were met, the process should not be interpreted as a full multi-round Delphi consensus. Comprehensibility was explored through a single final open-ended question rather than formal cognitive interviews, and the written responses were used descriptively and not formally coded. The pre-test sample of 39 was suitable for preliminary field administration but insufficient for structural validation of a 43-item instrument. The four domains and scoring rules are therefore theoretical and provisional. Internal consistency estimates cannot establish reliability before dimensionality is supported. The decision to retain and reverse-score the negatively worded contractual item has not been confirmed in an independent sample. The study did not assess test-retest reliability, measurement error, construct validity, criterion validity, responsiveness or measurement invariance. Finally, some system-level items may be difficult for frontline nurses to observe directly.

Conclusions

The study produced a 42-item candidate NPES-LTC intended to capture nursing practice environment attributes relevant to long-term care. Expert assessment supported the relevance and clarity of the retained items, and preliminary field administration provided information on comprehensibility and item behaviour. The four-domain structure, scoring model and reverse-scored item remain to be validated. Further cognitive testing and large-scale multicentre psychometric evaluation are required before the scale can be recommended for research, benchmarking or organisational decision-making.

Abbreviations

I-CVI
Item-level Content Validity Index
LTC
Long-term care
NPE
Nursing practice environment
NPES-LTC
Nursing Practice Environment Scale for Long-Term Care
PES-NWI
Practice Environment Scale of the Nursing Work Index
S-CVI/Ave
Scale-level Content Validity Index, average method.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Council of the Nursing School of Lisbon (Escola Superior de Enfermagem de Lisboa), process no. 4147/2023. Written informed consent was obtained from all participants before data collection.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analysed during the current study are not publicly available because participants did not consent to unrestricted public sharing. Pseudonymised data may be made available by the corresponding author upon reasonable request, subject to ethical approval, the approved protocol and applicable data-protection requirements. Data will be retained for five years after study completion and then securely destroyed.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors' contributions

VB was the principal investigator and first author and led the conceptualisation and design of the study, theoretical development of the construct, generation and refinement of the item pool, data collection, interpretation of the findings, and preparation and revision of the manuscript. AR contributed to the methodological development of the instrument-validation process and provided critical methodological appraisal. RB contributed to the organisation, analysis and interpretation of the results and to the critical revision of the manuscript. HJ contributed to study supervision, critical scientific appraisal, overall conceptual coherence and final intellectual review of the manuscript. All authors critically reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

The authors thank the expert panel and the nurses who participated in the pre-test.

References

1. World Health Organization Regional Office for Europe. Country assessment framework for the integrated delivery of long-term care: working document. Copenhagen: WHO Regional Office for Europe; 2021. Report no. WHO/EURO:2019-3522-43281-60659.
2. European, Commission. Directorate-General for Employment, Social Affairs and Inclusion. Long-term care report: trends, challenges and opportunities in an ageing society. Volume II, Country profiles. Luxembourg: Publications Office of the European Union; 2021. 10.2767/183997.
3. Fonseca C, Ramos A, Morgado B, Quaresma P, Garcia-Alonso J, Coelho A, et al. Long-term care units: a Portuguese study about the functional profile. *Front Aging*. 2023;4:1192718. 10.3389/fragi.2023.1192718.
4. Organisation for Economic Co-operation and Development. The economics of patient safety part III: long-term care. Paris: OECD.
5. Lake ET. Development of the Practice Environment Scale of the Nursing Work Index. *Res Nurs Health*. 2002;25:176–88. 10.1002/nur.10032.
6. Lake ET, Sanders J, Duan R, Riman KA, Schoenauer KM, Chen Y. A meta-analysis of the associations between the nurse work environment in hospitals and four sets of outcomes. *Med Care*. 2019;57:353–61. 10.1097/MLR.0000000000001109.

7. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*. 2002;288:1987–93. 10.1001/jama.288.16.1987.
8. Aiken LH, Sloane DM, Bruyneel L, Van den Heede K, Griffiths P, Busse R, et al. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. *Lancet*. 2014;383:1824–30. 10.1016/S0140-6736(13)62631-8.
9. de Brouwer BJM, Kaljouw MJ, Schoonhoven L, van Achterberg T. Essential elements of the nursing practice environment in nursing homes: psychometric evaluation. *Int J Older People Nurs*. 2017;12(2). 10.1111/opn.12137.
10. Min D, Cho E, Kim GS, Lee KH, Yoon JY, Kim HJ, et al. Factors associated with retention intention of registered nurses in Korean nursing homes. *Int Nurs Rev*. 2022;69:459–69. 10.1111/inr.12754.
11. Cho E, Min D, Heo S, Lee K, Kim H. Effects of registered nurses' staffing levels, work environment and education levels on nursing home residents' quality of life and nurse outcomes. *J Clin Nurs*. 2023;32:6494–503. 10.1111/jocn.16689.
12. Anunciada S, Benito P, Gaspar F, Lucas P. Validation of psychometric properties of the Nursing Work Index-Revised Scale in Portugal. *Int J Environ Res Public Health*. 2022;19:4933. 10.3390/ijerph19094933.
13. Sousa E, Lin CF, Gaspar F, Lucas P. Translation and validation of the Indicators of Quality Nursing Work Environments in the Portuguese cultural context. *Int J Environ Res Public Health*. 2022;19:12313. 10.3390/ijerph191912313.
14. Lucas P, Jesus É, Almeida S, Araújo B. Relationship of the nursing practice environment with the quality of care and patients' safety in primary health care. *BMC Nurs*. 2023;22:413. 10.1186/s12912-023-01571-8.
15. McGilton KS, Boscart VM, Brown M, Bowers B. Making tradeoffs between the reasons to leave and reasons to stay employed in long-term care homes: perspectives of licensed nursing staff. *Int J Nurs Stud*. 2014;51:917–26. 10.1016/j.ijnurstu.2013.10.015.
16. Skinner MS, Veenstra M, Sogstad M. Nurses' assessments of horizontal collaboration in municipal health and care services for older adults: a cross-sectional study. *Res Nurs Health*. 2021;44:704–14. 10.1002/nur.22144.
17. Midje HH, Torp S, Øvergård KI. The role of working environment and employee engagement in person-centred processes for older adults in long-term care services. *Int Pract Dev J*. 2022;12:1–19. 10.19043/ipdj.122.007.
18. Tummers LG, Groeneveld SM, Lankhaar M. Why do nurses intend to leave their organization? A large-scale analysis in long-term care. *J Adv Nurs*. 2013;69:2826–38. 10.1111/jan.12249.
19. Meleis AI. *Theoretical nursing: development and progress*. 6th ed. Philadelphia: Wolters Kluwer; 2017.
20. McCance T, McCormack B, Slater P, McConnell D. Examining the theoretical relationship between constructs in the person-centred practice framework: a structural equation model. *Int J Environ Res Public Health*. 2021;18:13138. 10.3390/ijerph182413138.

21. Donabedian A. The quality of care: how can it be assessed? *JAMA*. 1988;260:1743–8. 10.1001/jama.260.12.1743.
22. Coluci MZO, Alexandre NMC, Milani D. Construction of measurement instruments in the area of health. *Cien Saude Colet*. 2015;20:925–36. 10.1590/1413-81232015203.04332013.
23. DeVellis RF, Thorpe CT. Scale development: theory and applications. 5th ed. Thousand Oaks: SAGE; 2021.
24. Souza AC, Alexandre NMC, Guirardello EB. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol Serv Saude*. 2017;26:649–59. 10.5123/S1679-49742017000300022.
25. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. 6th ed. Oxford: Oxford University Press; 2024.
26. American Nurses Credentialing Center. 2023 Magnet Application Manual Sources of Evidence—2020 Pathway to Excellence Manual Elements of Performance Crosswalk. Silver Spring: ANCC; 2023.
27. Jasper MA. Expert: a discussion of the implications of the concept as used in nursing. *J Adv Nurs*. 1994;20:769–76. 10.1046/j.1365-2648.1994.20040769.x.
28. Gwet KL. Large-sample variance of Fleiss generalized kappa. *Educ Psychol Meas*. 2021;81:781–90. 10.1177/0013164420973080.
29. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data—which coefficients and confidence intervals are appropriate? *BMC Med Res Methodol*. 2016;16:93. 10.1186/s12874-016-0200-9.
30. Vanacore A, Pellegrino MS. Robustness of κ -type coefficients for clinical agreement. *Stat Med*. 2022;41:1986–2004. 10.1002/sim.9341.
31. Gagnier JJ, de Arruda GT, Terwee CB, Mokkink LB, Consensus Group. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures: version 2.0. *Qual Life Res*. 2025;34:1901–11. 10.1007/s11136-025-03950-x.
32. de Arruda GT, Terwee CB, Elsman EBM, Avila MA, Gagnier JJ, Mokkink LB, PROM Reporting Group. Explanation and elaboration document of the COSMIN Reporting Guideline 2.0 for studies on measurement properties of patient-reported outcome measures. *Qual Life Res*. 2025;34:1891–9. 10.1007/s11136-025-03949-4.
33. Braga VLRM. Construção de um Instrumento de Avaliação do Ambiente da Prática de Enfermagem nos Cuidados Continuados Integrados [Construction of an Instrument for Assessment of the Nursing Practice Environment in Integrated Continuous Care] [master's dissertation]. Lisbon: Escola Superior de Enfermagem de Lisboa; 2024. Available from: <http://hdl.handle.net/10400.26/58594>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [BMCNursingSupplementaryFile1NPESLTC42itemcandidate.docx](#)