RESEARCH

BMC Nursing

Open Access



Cross-cultural adaptation and psychometric evaluation of the Chinese version of the sickness presenteeism scale- nurse (C-SPS-N): a cross-sectional study

Chuang Li^{1,2}, Zhi Xing Meng^{2,3}, You Bei Lin² and Lan Zhang^{1*}

Abstract

Background Currently, no standardized evaluation instrument exists to assess the impact of presenteeism on nurses' productivity and the quality of care they provide. This study aimed to translate the Sickness Presenteeism Scale-Nurse (SPS-N) into Chinese and evaluate its reliability and validity among Chinese nurses.

Methods This study first translated the 21-item English version of the SPS-N scale into Chinese using Brislin's model. Then, six experts in the relevant field were invited to evaluate the item content validity index (I-CVI) of the translated scale. Using a convenience sampling method, 503 clinical nurses meeting the inclusion criteria were recruited from tertiary hospitals in Jinzhou, Liaoning Province, China. The reliability of the scale was assessed through internal consistency, split-half reliability, and test-retest reliability. To examine the structural validity of the Chinese version of the SPS-N (C-SPS-N), exploratory factor analysis (EFA) was conducted first, followed by confirmatory factor analysis (CFA) to further assess its construct validity.

Results The C-SPS-N demonstrated strong psychometric properties, with a Cronbach's *a* coefficient of 0.924. The item content validity index (I-CVI) for individual items ranged from 0.830 to 1.000. The split-half reliability was 0.750, and the test-retest reliability was 0.895. The four-factor exploratory factor model explained 78.354% of the total variance, indicating a robust factor structure. Confirmatory factor analysis (CFA) produced model fit indices of CMIN/ DF = 2.527, RMSEA = 0.067, AGFI = 0.857, TLI = 0.941, IFI = 0.950, CFI = 0.949, GFI = 0.900, and PGFI = 0.692. All indices fell within acceptable ranges, confirming a satisfactory model fit. Both convergent validity and discriminant validity were adequately supported.

Conclusion This study strictly adhered to the Brislin translation model and successfully introduced the SPS-N scale, which demonstrated strong reliability and validity in the Chinese cultural context. The Chinese version of the SPS-N (C-SPS-N) serves as an effective and reliable tool for assessing nurses' presenteeism behaviors.

Keywords Nurses, Reliability, Validity, Presenteeism, Sickness

*Correspondence: Lan Zhang zhang800519@126.com ¹Nursing Department, First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China

²School of Nursing, Jinzhou Medical University, Jinzhou, China
³General Surgery, Cancer Center, Department of Vascular Surgery,
Zhejiang Provincial People's Hospital (Affiliated People's Hospital),
Hangzhou Medical College, Hangzhou, Zhejiang, China

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or provide are included in the article's Creative Commons licence, unless indicated otherwise in a credit to the original in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit ine to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

With ongoing advancements in the medical service system, patient demand for high-quality care continues to rise. Nurses play a critical role as primary caregivers and key implementers of nursing interventions [1]. Nonetheless, presenteeism is highly prevalent among nurses, with rates three to four times higher than in other industries. This heightened prevalence is driven by factors such as shift work, heavy workloads, job insecurity, unfavorable working conditions, and additional stressors [2, 3]. Presenteeism refers to the phenomenon in which employees go to work despite being ill, even when they believe they should take sick leave [4]. Nurses' presenteeism not only adversely impacts patients' physical and mental health but also diminishes the quality of nursing care, disrupting patient treatment and recovery [5]. Moreover, nurses' job performance, motivation, job satisfaction, and work commitment are negatively affected by presenteeism [6, 7]. Diminished competence leads to a decline in organizational productivity, which can ultimately result in significant financial losses [8, 9]. According to reports from head nurses and nurses, annual economic losses due to presenteeism in Henan Province are estimated at 2.88 billion yuan and 4.38 billion yuan, respectively, as reported by scholars [10]. Additionally, Letvak and colleagues conducted a study on nurses in North Carolina, United States, and found that annual per capita losses due to nurses' presenteeism ranged from \$1,346 to \$9,000 [11]. Therefore, it is crucial to identify suitable tools to evaluate the presenteeism behavior of nurses and mitigating the negative impact of presenteeism.

Currently, the tools available for evaluating presenteeism include Stanford Presenteeism Scale [12], Endicott Work Productivity Scale [13], Health and Work Questionnaire [14] and the Luo Lu version of the presenteeism scale [15]. The above four assessment tools are applicable to the general occupational group and do not involve the transformation and measurement of productivity loss [16]. Furthermore, the Nurses Work Functioning Questionnaire (NWFQ) and Nurse Presenteeism Questionnaire (NPQ) can also be used to evaluate presenteeism. However, the NWFQ focuses on the impairment of work function resulting from common mental disorders [17], whereas the NPQ evaluates whether nurses report working while experiencing various health issues [18]. Although the NWFQ and NPQ are tools developed for evaluating nurses, neither can be used effectively to access the effect of presenteeism on nurses' productivity and work output.

The concept of "transformation of productivity loss" refers to converting subjective experiences such as fatigue, decreased work efficiency, and distraction into quantifiable indicators. These indicators include factors like increased work hours, reduced task completion

quality, and a higher rate of medical errors. "Measurement of productivity loss" involves using standardized tools to collect data and quantify the extent of productivity loss through specific metrics [19]. The Health and Productivity Model (HPM) was proposed by Goetzel [20]. Its core idea of the model is that health problems not only directly lead to presenteeism but also affect employees' work efficiency and quality, ultimately impacting the overall productivity of the organization. The Work Ability Model (WAM) was proposed by Ilmarinen [21], the model suggests that work ability is a comprehensive concept that includes not only physical health but also psychological and social factors. As individuals age or experience changes in their health status, their work ability may decline, thereby affecting work performance and productivity. Through these two models, we can identify an important relationship between the transformation and measurement of productivity loss and presenteeism. Health issues, such as illness and fatigue, lead to a decline in work capacity, which in turn affects work efficiency and quality. Specifically, among nurses, the presence of presenteeism not only reduces work efficiency but also increases the risk of medical errors, lowers the quality of care, and poses a threat to patient safety. Moreover, productivity loss caused by attendance issues not only affects individual performance but also places a burden on the entire healthcare system. Therefore, quantifying productivity loss and accurately assessing the impact of nurses' health problems on productivity decline is crucial for implementing effective interventions, improving the work environment, and enhancing the quality of care.

Nurses' work is highly specialized, encompassing direct patient care, healthcare safety management, and intense emotional labor [22]. However, existing tools struggle to accurately measure the impact of these factors on sickness presenteeism [12-15, 18]. In 2023, Turkish scholar Veysel Karani Barış developed the Sickness Presenteeism Scale for Nurses (SPS-N) [23]using a systematic literature review and the Delphi method, both widely recognized approaches for scale development [24]. The SPS-N was designed to systematically review national and international studies on sickness presenteeism and extract key items closely related to nurses' experiences. Additionally, nurses were consulted to assess their work conditions and perceptions of presenteeism [23], leading to the development of a multidimensional assessment tool encompassing general performance, patient safety, relationships within the team, and emotions. Compared to existing assessment tools, the SPS-N provides more comprehensive coverage and a nurse-specific framework for evaluating sickness presenteeism. The SPS-N has undergone rigorous psychometric validation, demonstrating strong reliability and validity [23]. Therefore, the purpose of this study is to translate the SPS-N into

Chinese and evaluate its psychometric properties among Chinese clinical nurses. Through this research, we aim to provide nursing administrators and policymakers with a more accurate and culturally relevant measurement tool to optimize nurses' occupational health management and mitigate the impact of sickness presenteeism on the quality of care and patient safety.

Methods

Participants

From October 23, 2023 to February 2024, a cross-sectional study was conducted, using a convenience sampling method to select clinical nurses from the First Hospital of Jinzhou Medical University in Liaoning Province. The inclusion criteria were as follows: ①Licensed Nurse Practitioner with at least six months of experience working as hospital nurses. ② Voluntary participation in research on this topic. The exclusion criteria were as follows: ①Nurses who left their clinical posts due to reasons such as study abroad, vacation, maternity leave, or other circumstances during the study period. ②Nurses who had been not working with illness in the last month. ③Nurses in internships or undergoing advanced training at the surveyed hospitals.

The sample size was estimated using Kendall's method, which recommends a sample size of 5 to 10 times the number of questionnaire items [25]. Considering an expected attrition rate of 20%, preliminary calculations indicated that the required sample size ranged from 116 to 252 participants. Additionally, to meet the minimum sample requirements for exploratory factor analysis (EFA) (\geq 100 cases) and confirmatory factor analysis (CFA) (\geq 200 cases) [26]. We ultimately recruited 503 clinical nurses.

It is important to note that our initial estimated sample size (116–252) was based on general psychometric assessment guidelines. However, a larger sample was chosen to enhance the stability and reliability of the findings. Increasing the sample size also improved the robustness of the factor analyses, enhanced the psychometric validation of the assessment instrument, and strengthened the generalizability of the results.

Translation and cross-cultural adaptation

In this study, the original author was contacted by email for authorization, and then the SPS-N scale was translated into Chinese version according to Brislin model [27]:

The SPS-N was developed by the team of Professor Veysel Karani Baris based on a multidisciplinary theory, including general performance (items 1–5), patient safety (items 6–12), relationships within the team (items 13–15) and emotions (items 16–21), a total of 21 items, using the Likert 5-level scoring method, with responses ranging

from 1 "strongly disagree" to 5 "strongly agree". The total score is 21–105, with higher scores indicating higher sickness presenteeism among nurses. The original scale has good reliability and validity, and it tested the reliability and validity of 619 nurses living in 55 different cities in Turkey. The total Cronbach's α value was calculated as 0.928, the Cronbach's α value of the sub-dimension was calculated as 0.815 ~ 0.903, and the composite reliability value was calculated as 0.804 ~ 0.903 [23].

- *Step 1*: The original SPS-N was independently translated into two Chinese versions, S1 and S2, by two graduate nursing students who were native Chinese speakers with Level 6 English proficiency. The first author then integrated S1 and S2, conducted thorough discussions, and made necessary modifications to develop the final Chinese version of the scale, S.
- *Step 2*: The Chinese version of Scale S was independently back-translated into English by a Doctor of Nursing Science and a Master's degree holder in Medical English, both of whom had no prior exposure to the original scale. This process produced the English versions SS1 and SS2.
- *Step 3*: A professor of nursing management and an associate professor of clinical nursing integrated the back-translated versions to achieve a semantic consistency rate of over 95%, forming the final back-translated version, SS.
- *Step 4*: Following cultural adaptation guidelines, six experts were invited to evaluate the Chinese version of the SPS-N through two rounds of assessment via email and on-site consultation. This process aimed to balance idiomatic conceptual equivalence with cultural adaptation, ensuring that the language aligned with regional linguistic norms.

Measurement and instruments

- 2.3.1 After reviewing the literature, the researcher designed a questionnaire to collect demographic data of nurses, including: gender, age, department, working years, Marriage and childbearing situation, etc.
- 2.3.2 Chinese version of the Sickness Presenteeism Scale-Nurse(C-SPS-N), including 21 items in four dimensions: general performance (items 1–5), patient safety (items 6–12), relationships within the team(items 13–15) and emotions (items 16–21), The answers ranged from 1 for "strongly disagree" to 5 for "strongly agree," using Likert's 5-level scale. The final result was 21–105 scores, with higher scores indicating higher sickness presenteeism by nurses.

2.3.3 The Nurse Presenteeism Questionnaire (NPQ) was developed by Chinese scholar Geyan Shan in 2021 [18]. It is a unidimensional scale comprising 11 items, rated on a four-point Likert scale: 0 = never, 1 = once, 2 = 2-5 times, and 3 = more than five times. Higher scores indicate more frequent presenteeism. The NPQ has an internal reliability coefficient of 0.940.

Data collection

Pre-survey

In October 2023, 30 clinical nurses from the First Affiliated Hospital of Jinzhou Medical University, Liaoning Province, China, were selected as pre-survey participants using a convenience sampling method [28]. After receiving an introduction to the study's purpose and significance, all participants provided informed consent. The pre-survey results indicated that the scale was thematically clear, structurally complete, and logically coherent, with no reported difficulties in semantic comprehension. On average, participants completed the questionnaire in approximately three minutes. Consequently, no modifications were made, and the Chinese version of the SPS-N scale was finalized.

Formal investigation

Before the survey, informed consent was obtained from the hospital's nursing department. Additionally, the head nurses of all participants were contacted to explain the study's purpose and provide instructions for completing the questionnaire. The survey instructions emphasized that the data would be used exclusively for scientific research. Participants were assured that participation was anonymous and voluntary. Trained staff distributed paper questionnaires, and completed forms were carefully reviewed for accuracy. Responses completed in less than three minutes or displaying clear answer patterns were excluded. A total of 550 questionnaires were distributed, of which 503 were verified as valid, resulting in a response rate of 91.5% (503/550*100%). During the survey, participants could voluntarily provide their contact information for reliability retesting. Two weeks later, 40 nurses were randomly selected from the initial participants and completed the same questionnaire to assess test-retest reliability.

To ensure methodological rigor and avoid potential biases associated with using the same sample for both EFA and CFA, the total sample (N=503) was randomly divided into two independent subsamples. A total of 162 participants were allocated for EFA, while 341 participants were used for CFA. This approach allowed us to independently identify the factor structure in EFA and validate it in CFA, thereby enhancing the psychometric robustness of the scale.

Data analysis

The data from the paper questionnaires were independently entered into Excel by two researchers. Statistical analysis was conducted using SPSS 27.0 and AMOS 24.0. Qualitative data were reported as frequencies and percentages, while continuous variables were presented as means and standard deviations.

Prior to formal analysis, missing data analysis was performed to ensure data completeness and enhance study transparency. The results, obtained using SPSS 27.0, indicated that the dataset was complete and contained no missing values (Supplementary material 1). Consequently, no imputation or other missing data handling techniques were required.

Item analysis

The critical ratio and correlation coefficient methods were used to screen scale items. (1) Critical Ratio Method: An independent samples t-test was conducted on the high (top 27%) and low (bottom 27%) subgroups to assess whether the differences were statistically significant. A total of 503 questionnaires were ranked from highest to lowest based on total scores. Items with a critical ratio > 3 and statistical significance were retained [29]. (2) Correlation Coefficient Method: Pearson's correlation coefficient was used to evaluate the relationship between each of the 21 items and the total scale score. Items with correlation coefficients below 0.4 were excluded due to their weak correlation with the total score [30].

Validity analysis

- (1). Content validity: Six nursing experts were invited to evaluate the content validity of the C-SPS-N using the Delphi method. The assessment was conducted on a 4-point Likert scale, where each item was rated as follows: not relevant = 1, weakly relevant = 2, more relevant = 3, and strongly relevant = 4, based on its relevance to the topic. The item content validity index (I-CVI) was calculated as the proportion of experts who rated an item as 3 or 4 out of the total number of experts. The scale content validity index (S-CVI) was determined as the average I-CVI across all items [31].
- (2). Construct validity: The latent factor structure of the translated scale was examined using both EFA and CFA. For EFA, principal component analysis with orthogonal rotation (varimax) was performed. CFA was conducted using AMOS to assess the model's fit indices.
- (3). Convergent and discriminant validity: Based on the outcomes of CFA, correlation coefficients between observed variables, Average Variance Extracted (AVE), and Construct reliability (CR) were

measured. Discriminant validity was tested using the Fornell-Larcker criterion, which revealed that the square root of the AVE for each latent variable was greater than the correlation coefficients between that latent variable and the other latent variables.

(4). Calibration validity refers to the relationship between the target instrument and other measurement standards [32], The NPQ is used as the calibration standard in this study.

Reliability analysis

This study assessed reliability using test-retest reliability and internal consistency. To evaluate internal consistency, Cronbach's α coefficient was calculated for each dimension of the C-SPS-N. A total of 40 nurses who voluntarily provided their contact information during the first survey were randomly selected as the sample for the test-retest reliability analysis. The correlation between the two sets of scores was calculated to determine the stability of the measurement tool. Additionally, the scale items were split into two halves, and the correlation between the two halves was computed to assess split-half reliability.

Ethical consideration

The Jinzhou Medical University Ethics Committee (JZMULL2023133) approved this study, and all research procedures adhered to the committee's ethical guidelines. Informed consent was obtained from all participants before data collection.

Results

Cross-cultural adaptation

Taking into account the conventions of the Chinese language in our context and in accordance with expert opinions, items 1 to 15 were "Due to my problem," was revised to "Because of my problem," to better align with the everyday language habits of Chinese speakers. In both spoken and written Chinese, "Because of" is more commonly used than "Due to" [33] making the expression more natural and relatable for respondents. Additionally, "Because of" is a typical pair of correlative conjunctions in Chinese, often used to emphasize cause-and-effect relationships. By replacing "Due to" with "Because of" the phrasing aligns better with Chinese grammatical conventions and enhances the logical clarity of the causal relationship within the sentence [34]. Consideration of comprehensibility and the purpose of the scale and avoidance of ambiguity. Replace entry 16, "I am angry with my leader because I have to work even though I have health problems" with "I am unhappy with my leader because I have to work even though I have health problems".

Participants

A total of 503 research participants met the inclusion criteria. The participants' ages ranged from 22 to 55 years (33.24 ± 6.67) . For more details, see Table 1.

Item analysis

In this study, an independent samples t-test was conducted to assess the discriminative ability of the questionnaire between high and low scoring groups. The critical ratios for the 21 items ranged from 9.015 to 22.837 (all > 3, P < 0.01) [29]. Pearson correlation analysis was used to examine the relationship between individual item scores and the total score, yielding correlation coefficients of r = 0.440 to 0.733 (P < 0.01) [30] Table 2.

Validity

Content validity

Six experts were invited to assess the content validity of the C-SPS-N using the Delphi method. The I-CVI and S-CVI were calculated based on a 4-point Likert scale. The results indicated that the I-CVI ranged from 0.83 to 1.00(>0.78), while the S-CVI was 0.910(>0.90) [35].

Construct validity

Exploratory factor analysis Before conducting EFA, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were performed. A KMO value greater than 0.7 and P < 0.05 were generally considered suitable for factor analysis [36]. In this study, the KMO value was 0.890, and Bartlett's test of sphericity yielded an approximate chi-square value of 3706.134 (df = 210, P < 0.05). Principal component analysis (PCA) was used to extract factors with eigenvalues greater than 1 [37]. The component matrix was obtained through orthogonal varimax rotation, and only factors with loadings greater than 0.5 were retained [38] (Table 3). After 6 iterations of rotation and convergence, a total of 4 metrics were extracted to agree with the original scale, with a cumulative explained variance of 78.354% (Fig. 1).

Confirmatory factor analysis The goal of CFA is to verify whether the relationships between questionnaire items and factors align with the proposed hypotheses. Model fit indices include CMIN/DF, RMSEA, AGFI, GFI, TLI, IFI, CFI, and PGFI. The initial model did not meet the desired criteria. Based on modification indices (MI) [39], the initial model was adjusted by correlating error terms e6 and e7, e16 and e17, and e11 and e12 sequentially (Table 4). Figure 2 presents the final model fit indices: CMIN/DF = 2.527(<3), RMSEA = 0.067(<0.08), AGFI = 0.875, TLI = 0.941, IFI = 0.950, CFI = 0.949, GFI = 0.900, and PGFI = 0.692(>0.5) [40]. Table 5 shows that CR ranged

Table 1 Distribution of demographic characteristics (N = 503)

Variables		Frequency	Percentage%
Sex	Males	33	6.6
	Females	470	94.4
Department	Medical department	207	41.2
	Surgical department	160	31.8
	Maternity ward	15	3.0
	ICU	45	8.9
	Operating rooms	47	9.3
	Emergency department	18	3.6
	OPD	6	1.2
	Paediatrics department	5	1.0
Marriage and childbearing situation	Unmarried and childless	150	29.8
	Married and childless	71	14.1
	Married and pregnant	278	55.3
	Other	4	0.8
Forms of employment	Formal	87	17.3
	Contractual	407	80.9
	Labor dispatch	9	1.8
Working time	6–12 months	32	6.4
	1–5 years	110	21.9
	5–10 years	150	29.8
	More than 10 years	211	41.9
Occupational level	Nurse	67	13.3
	Nurse Practitioner	307	61.0
	Nurse-in-charge	117	23.3
	Co-chief nurse and above	12	2.4
	Technical secondary school	1	0.2
Educational level	College degree	57	11.3
	Bachelor's degree	399	79.3
	Master's degree	46	9.1
Income situation	2000–3000 yuan	54	10.7
	3001–5000 yuan	198	39.4
	5001–7000 yuan	187	37.2
	7001–10,000 yuan	62	12.3
	More than 10,000 yuan	2	0.4

Abbreviations: ICU-Intensive Care Unit; POD-Pediatric Outpatient Department

from 0.854 to 0.927(>0.7), while the AVE values ranged from 0.548 to 0.688(>0.5) [41].

Criterion validity NPQ was used as a criterion in this study. By correlation analysis, C-SPS-N was highly correlated with it with a correlation coefficient of 0.867 > 0.7, p < 0.001 [42].

Reliability

As shown in Table 6, the overall Cronbach's α coefficient for C-SPS-N was 0.924. The Cronbach's α coefficient for the four factors were 0.854, 0.939, 0.870, and 0.928, all exceeding the threshold of 0.7 [43]. Additionally, the testretest reliability after a two-week interval was 0.895 and the calculated split-half reliability was 0.750, both meeting the minimum reference standards [44].

Discussion

Advantages of C-SPS-N

In the context of Chinese culture, presenteeism among nurses is profoundly influenced by collectivist values. Collectivism prioritizes the interests of the group over individual needs, leading nurses to exhibit a strong sense of responsibility and professional commitment in their work [45]. Even when facing health issues, they may choose to continue working to avoid disrupting team operations or compromising patient care due to their absence [46]. The significant difference between the C-SPS-N and other related assessment tools (e.g., NPQ) [18] lies in its ability to capture the influence of collectivist values within the Chinese cultural context on nurses' behavior. The NPQ primarily focuses on the direct impact of health issues on nurses' presenteeism, with items 1–11 exploring scenarios where nurses persist in working

Table 2 Critical ratios of C-SPS-N, item-total correlation coefficients, and Cronbach's alpha values after item deletion (n = 503)

ltem	Critical	Correlation item-	Р	Cronbach'sa
	ratio	total score		after delet-
				ing the item
gp1	11.711	0.440	P<0.01	0.927
gp2	22.837	0.582	P<0.01	0.922
gp3	17.916	0.587	P<0.01	0.922
gp4	19.457	0.659	P<0.01	0.920
gp5	19.361	0.691	P<0.01	0.919
ps6	9.015	0.498	P<0.01	0.923
ps7	10.365	0.601	P<0.01	0.921
ps8	13.716	0.660	P<0.01	0.920
ps9	15.150	0.697	P<0.01	0.920
ps10	15.316	0.689	P<0.01	0.920
ps11	14.455	0.679	P<0.01	0.920
ps12	14.104	0.689	P<0.01	0.921
tr13	15.408	0.821	P<0.01	0.921
tr14	11.232	0.607	P<0.01	0.921
tr15	12.184	0.644	P<0.01	0.920
em16	12.002	0.698	P<0.01	0.919
em17	15.932	0.733	P<0.01	0.918
em18	17.156	0.725	P<0.01	0.919
em19	15.356	0.706	P<0.01	0.919
em20	17.399	0.702	P<0.01	0.919
em21	16.128	0.704	P<0.01	0.919

Table 3 Factor loadings of exploratory factor analysis for the C-SPS-N (n = 162)

Item	Factor 1	Factor 2	Factor 3	Factor 4
gp1	-	_	0.717	-
gp2	-	-	0.815	-
gp3	-	-	0.792	-
gp4	-	-	0.724	-
gp5	-	-	0.674	-
ps6	0.734	-	-	-
ps7	0.830	-	-	-
ps8	0.844	-	-	-
ps9	0.889	-	-	-
ps10	0.888	-	-	-
ps11	0.756	-	-	-
ps12	0.739	-	-	-
tr13	-	-	-	0.813
tr14	-	-	-	0.736
tr15	-	-	-	0.661
em16	-	0.584	-	-
em17	-	0.788	-	-
em18	-	0.857	-	-
em19	-	0.834	-	-
em20	-	0.881	-	-
em21	-	0.807	-	-

despite experiencing physical discomforts such as fever, dizziness, or abdominal pain. However, this design is relatively generic and fails to fully reflect the complexity of nurses' behaviors within different cultural contexts. In contrast, the C-SPS-N incorporates the cultural characteristics of nursing practice in China. From the perspective of cultural background, such as item13"Because of my health problem, I had a conflict with the healthcare team members I worked with "and item 15" I felt unhappy because my colleagues, who had to do my work due to my health problem, were angry with me" It highlights the team pressure and psychological burden nurses may face when their health is compromised within a collectivist culture. In Chinese culture, nurses often prioritize their sense of responsibility and the importance of maintaining team harmony, leading them to persist in their work despite poor health. These cultural factors are not adequately reflected in the NPQ.

In terms of nursing performance, the NPQ items primarily focus on the direct relationship between health status and attendance behavior, whereas the C-SPS-N places greater emphasis on the broader impact of health issues on care quality and team collaboration. For instance, item 6 in the C-SPS-N, "I made medication errors because of my health problem" and item 8, "I could not implement infection control interventions because of my health problem" explicitly assess the specific effects of health issues on the quality of care and patient safety. In contrast, the NPQ does not cover similar content, which limits its applicability in evaluating nursing performance. Item 21, "Because I had to work despite my health problem, I could not feel successful in my job." delves into the negative impact of health problems on nurses' psychological well-being. These items address the NPQ's lack of attention to nurses' emotional states, enabling the C-SPS-N to provide a more comprehensive assessment of the multifaceted effects of health issues on nursing performance.

The C-SPS-N has the suitable distinction

In addition to its previously mentioned advantages, C-SPS-N demonstrates strong performance in quantitative research. The primary objective of item analysis is to assess the discriminatory power of the scale and individual items. To reduce bias associated with a single testing method, this study employed both the critical ratio method and the correlation coefficient method to evaluate item inclusion or exclusion. Independent samples t-tests were conducted for high and low scoring subgroups, yielding t-values ranging from 9.015 to 22.837 (all > 3.0, P < 0.01) [31], indicating strong item discrimination. The correlation coefficients between each item and the total score, calculated using the Pearson correlation method, ranged from 0.440 to 0.733 (all > 0.4, P < 0.01) [30], indicating a significant association between each item and the overall scale. We found that the item-total correlation coefficients of the slightly C-SPS-N were higher than those of the original scale, ranging from 0.430 to 0.730. This improvement may be attributed to appropriate linguistic adaptations made during the Sinicization process, which effectively avoided potential cultural ambiguities or translation biases present in the



Fig. 1 Scree plot for the C-SPS-N exploratory factor analysis(n = 162)

Table 4	Model fit ind	lices of C-SPS	-N before and	l after moo	dification ir	n confirmatory	/ factor anal	ysis
---------	---------------	----------------	---------------	-------------	---------------	----------------	---------------	------

model	CMIN/DF	RMSEA	AGFI	TLI	IFI	CFI	GFI	PGFI
M1	4.205	0.097	0.776	0.876	0.893	0.892	0.823	0.652
M2	2.527	0.067	0.857	0.941	0.950	0.949	0.900	0.692
Standard	< 3	< 0.08	>0.9	> 0.9	> 0.9	>0.9	> 0.9	> 0.5
Fitting effect	good	good	acceptable	good	good	good	good	good

Note: M1: Before Modification; M2: After Modification

original version. These adjustments likely reduced variability in participants' interpretations of the items and enhanced the items' representativeness of the overall construct. Furthermore, the controlled data collection environment may have minimized comprehension bias, thereby strengthening the associations between individual items and the total score. The total Cronbach's α coefficient of the scale was 0.924. Although removing the first item increased Cronbach's α coefficient to 0.927, following the criteria established by Hanyi Wang [47], items were retained unless their removal increased Cronbach's α coefficient by more than 0.5. As the reliability of other items remained unaffected, all 21 items were preserved. These results suggest that the C-SPS-N maintains all 21 items with high homogeneity and strong discriminative power.

The C-SPS-N has suitable validity

Content validity assesses whether the items accurately represent the construct being measured. Experts in the field comprehensively evaluate the scale's content to ensure its appropriateness. In this study, six experts were invited to assess the scale's content validity and perform cultural adaptations. The results indicated that the S-CVI of the translated scale was 0.910, which is slightly lower than that of the original scale (0.963). This discrepancy may be attributed to variations in expert interpretation due to differences in domain expertise or professional experience, potentially resulting in inconsistent scoring across certain items. Nonetheless, the I-CVI values ranged from 0.83 to 1.00, closely aligning with those of the original scale, and all exceeded the recommended threshold, supporting the content validity of the translated version [37]. These findings suggest that the C-SPS-N is highly regarded by professionals and that its language is culturally appropriate, aligns with Chinese linguistic norms, and is easy to comprehend.

Structural validity is a theoretical form of validity that reflects the conceptual framework under study. Principal component analysis with varimax rotation identified four latent factors: general performance, patient safety, relationships within the team, and emotions. These factors were consistent with those in the English version of the scale. All rotated factor loadings exceeded 0.5, with double-loading phenomena, meeting psychometno ric requirements. The results of EFA indicated that the cumulative variance explained by the Chinese version of the scale was 78.354%, significantly higher than the 57.9% reported for the original version. This discrepancy may be attributed to several factors. First, during the localization process, certain items were semantically adapted to align with the Chinese cultural context, enhancing their relevance to the target population's linguistic habits and cognitive styles. This likely improved the consistency in



Fig. 2 Hypothesized confirmatory factor analysis model of the C-SPS-N (n = 341)

participants' interpretation of the items. Second, the translated version may have achieved greater clarity and contextual relevance in its phrasing, which enhanced item cohesion, reduced measurement error, and improved the efficiency of factor extraction. Additionally, variations in sample characteristics could have influenced the stability of the factor structure and the variance explained. Overall, the higher cumulative variance suggests that the Chinese version of the scale demonstrates strong structural validity. CFA results showed that CMIN/DF was below 3,

while GFI, TLI, IFI, and CFI all exceeded 0.9, and RMSEA was below 0.08 [48, 49]. Although the adjusted goodnessof-fit index (AGFI = 0.857) fell short of the ideal threshold of 0.9, it remained within an acceptable range. This minor deviation may be due to sample size limitations. Overall, the remaining indices met ideal thresholds, and the model demonstrated satisfactory fit, confirming that the scale possesses strong structural validity. Interestingly, the CFA revealed that the standardized factor loading of the first item was identical for both the Chinese-adapted

Discrimin	ant Validity					Convergent Va	lidity			
Factors	F1	F2	F3	F4	Items	Std.Estimate	SE	P	CR	AVE
F1	0.740	0.421**	0.361**	0.362**	gp1	0.444			0.854	0.548
					gp2	0.805	0.216	P<0.01		
					gp3	0.828	0.198	P<0.01		
					gp4	0.769	0.198	P<0.01		
					gp5	0.787	0.188	P<0.01		
F2		0.807	0.598**	0.342**	ps6	0.571			0.927	0.651
					ps7	0.750	0.092	P<0.01		
					ps8	0.855	0.139	P<0.01		
					ps9	0.957	0.154	P<0.01		
					ps10	0.915	0.155	P<0.01		
					ps11	0.741	0.158	P<0.01		
					ps12	0.796	0.152	P<0.01		
F3			0.830	0.388**	tr13	0.826			0.869	0.688
					tr14	0.837	0.053	P<0.01		
					tr15	0.826	0.056	P<0.01		
F4				0.815	em16	0.659			0.921	0.664
					em17	0.721	0.074	P<0.01		
					em18	0.832	0.117	P<0.01		
					em19	0.877	0.126	P<0.01		
					em20	0.925	0.125	P<0.01		
					em21	0.854	0.117	P<0.01		

 Table 5
 Discriminant and convergent validity of the C-SPS-N (n = 341)

Note: Bold text is the square root of AVE; **P < 0.01

Table 6 Total reliability, split-half reliability, and test-retest reliability of C-SPS-N (n = 503)

Factors	Cronbach's α coefficient	split-half reliability	test- retest reliability
General performance Patient safety	0.854 0.939	0.750	0.895
Relationships within the team Emotions	0.870 0.928		

and original versions of the scale, with a value of 0.44. This consistency may be attributed to the high semantic equivalence maintained during the translation process, which preserved the original meaning without introducing cultural bias or alterations in presentation. As a result, despite the difference in language, participants likely interpreted the item similarly, leading to equivalent psychological responses and identical factor loading values.

Convergent validity assesses whether items measuring the same underlying construct are appropriately grouped. The C-SPS-N demonstrated CR values exceeding 0.6 and AVE values above 0.5 for all four factors [30]. Specifically, the AVE for Factor 1, was 0.548 (compared to 0.462 in the original scale); for Factor 2, it was 0.651 (original: 0.572); for Factor 3, it was 0.688 (original: 0.644); and for Factor 4, it was 0.664 (original: 0.540). The observed increases in AVE values may be attributed to several factors. First, during the translation process, the research team not only preserved the fidelity of the original content but also optimized culturally ambiguous items by contextualizing them appropriately. This enhanced alignment with Chinese linguistic habits and cognitive styles, thereby improving item cohesion. Second, the sample used in this study may have had higher compatibility with the adapted content, potentially showing greater consistency in educational background, professional experience, and cultural understanding, which contributed to stronger inter-item correlations. Third, the expert review process involved refining the wording of items to ensure clarity and conciseness, which facilitated accurate comprehension by respondents and minimized interpretation bias. Collectively, these factors contributed to the enhanced convergent validity of the Chinese version, allowing for a more precise measurement of the intended latent constructs.Therefore, the scale exhibits good convergent validity. The square root of the average variance extracted for each dimension of the C-SPS-N exceeds the correlation coefficients between subscales, indicating satisfactory discriminant validity. This test assesses whether items representing different constructs are properly distinguished and not incorrectly classified together [31].

The C-SPS-N has suitable reliability

Internal consistency reliability reflects the degree of homogeneity among all test items. Cronbach's α coefficient below 0.6 indicate insufficient internal consistency,

while values between 0.7 and 0.8 suggest moderate reliability. A Cronbach's α coefficient between 0.8 and 0.9 signifies good reliability [29]. In this study, the Cronbach's α coefficient of the total scale was 0.943, which is higher than that of the original scale (0.928). This improvement may be attributed to the enhanced clarity and precision achieved during the localization process. By employing a rigorous "translation-back-translation-expert revision" procedure, the study preserved the original meanings while adapting certain semantic expressions, making the items easier for participants to understand and respond to accurately. Furthermore, the Cronbach's α coefficients for each subscale ranged from 0.843 to 0.944, indicating strong internal consistency across the 21 items of the translated version. Test-retest reliability measures the stability and consistency of a scale's results over time, expressed as a correlation coefficient ranging from 0 to 1, with values closer to 1 indicating higher reliability [30]. The overall test-retest reliability in this study was 0.896, with individual dimension reliability ranging from 0.854 to 0.939, demonstrating strong stability and consistency. Split-half reliability assesses internal consistency by dividing the questionnaire items into two halves, treating them as separate measurements taken within a short time frame. The correlation coefficient between the two halves serves as the measure of split-half reliability. A spearman correlation coefficient of ≥ 0.7 indicates good split-half reliability. In this study, the split-half reliability of the translated scale was 0.750, which is lower than that of the original English version (0.867). This discrepancy may be attributed to differences in item comprehension between the two linguistic and cultural contexts. While the original items may have been uniformly understood in the native English context, certain translated itemsparticularly those reflecting emotions, beliefs, or cognitive experiences-might have allowed for more subjective interpretation in Chinese, thereby reducing consistency between the two split halves. Nevertheless, the current level of split-half reliability remains within an acceptable range, indicating that the translated scale maintains reasonable internal stability.

With its strong reliability and validity, the C-SPS-N integrates cultural adaptation and nursing-specific work characteristics, providing greater specificity and scientific rigor in assessing nurse performance. By addressing the limitations of traditional scales in the nursing context, it offers enhanced practical applicability.

Limitations

First, a convenience sampling method was used to select 503 nurses from a tertiary hospital in Jinzhou, Liaoning Province. This approach may introduce selection bias, limiting the representativeness of the sample and affecting the generalizability of the findings. Future studies should consider using random sampling or expanding the sample to multiple healthcare institutions to enhance the generalizability of the results. Second, this study may be subject to sampling bias and confounding bias. For example, factors such as participants' years of professional experience, department, and personal health status may influence their understanding of the scale and response tendencies, potentially affecting the results. Future research could employ stratified sampling or adjust statistical analysis methods (e.g., multivariate regression analysis) to control for potential confounding factors and improve the internal validity of the study. Despite these limitations, this study followed a rigorous process of translation, cultural adaptation, and reliability and validity testing, confirming the applicability and measurement quality of the scale. Future studies could further validate the scale's applicability across different regions and populations and use longitudinal research methods to examine its long-term stability.

Conclusion

This study strictly adhered to the Brislin translation model and successfully introduced the SPS-N scale, demonstrating strong reliability and validity within the Chinese cultural context. The scale serves as an effective and reliable tool for assessing nurses' presenteeism behaviors. Furthermore, the C-SPS-N provides a foundation for developing targeted interventions and nursing management strategies to mitigate presenteeism in clinical settings.

Abbreviations

Sickness presenteeism scale- nurses
chi-square/degree of freedom
Root-mean-square error of approximation
Comparative fit index
Tucker lewis index
Incremental fit index
Goodness-of-fit index
Adjusted goodness-of-fit index
Parsimony goodness-of-fit index
Scale-level content validity index
Item-level content validity index
The Kaiser–Meyer–Olkin
Exploratory factor analysis
Confirmatory factor analysis
The modification indices
Critical ration

AVE Average variance extracted

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12912-025-03113-w.

Supplementary Material 1

Acknowledgements

We express our great gratitude to the participants in the study.

Author contributions

CL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing-original draft, Writing-review & editing. ZM: Investigation, Methodology, Resources.YL: Investigation, Methodology, Software. LZ: Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing-review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Data availability

The experimental data of this study are available from the authors upon request. Data will be provided by the authors upon reasonable request.

Declarations

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 July 2024 / Accepted: 24 April 2025 Published online: 07 May 2025

References

- Smith CM, Horne CE, Wei H. Nursing practice in modern healthcare environments: A systematic review of attributes, characteristics, and demonstrations. J Adv Nurs. 2024;80:3481–98.
- Li Y, Guo B, Wang Y, Lv X, Li R, Guan X, et al. Serial-Multiple mediation of job burnout and fatigue in the relationship between sickness presenteeism and productivity loss in nurses: A multicenter Cross-Sectional study. Front Public Health. 2021;9:812737.
- Allemann A, Siebenhüner K, Hämmig O. Predictors of presenteeism among hospital Employees-A Cross-Sectional Questionnaire-Based study in Switzerland. J Occup Environ Med. 2019;61:1004–10.
- Min A, Kang M, Park H. Global prevalence of presenteeism in the nursing workforce: A meta-analysis of 28 studies from 14 countries. J Nurs Manag. 2022;30:2811–24.
- Homrich PHP, Dantas-Filho FF, Martins LL, Marcon ER. Presenteeism among health care workers: literature review. Rev Bras Med Trab. 2020;18:97–102.
- Rainbow JG, Drake DA, Steege LM. Nurse health, work environment, presenteeism and patient safety. West J Nurs Res. 2020;42:332–9.
- Rainbow JG, Steege LM. Presenteeism in nursing: an evolutionary concept analysis. Nurs Outlook. 2017;65:615–23.
- Aysun K, Bayram Ş. Determining the level and cost of sickness presenteeism among hospital staff in Turkey. Int J Occup Saf Ergon. 2017;23:501–9.
- 9. Umann J, Guido L, de Grazziano A. E da S. Presenteeism in hospital nurses. Rev Lat Am Enfermagem. 2012;20:159–66.
- Shan G, Wang S, Wang W, Guo S, Li Y. Presenteeism in nurses: prevalence, consequences, and causes from the perspectives of nurses and chief nurses. Front Psychiatry. 2020;11:584040.
- 11. Letvak SA, Ruhm CJ, Gupta SN. Nurses' presenteeism and its effects on selfreported quality of care and costs. Am J Nurs. 2012;112:30–8. quiz 48, 39.
- Cicolini G, Della Pelle C, Cerratti F, Franza M, Flacco ME. Validation of the Italian version of the Stanford presenteeism scale in nurses. J Nurs Manag. 2016;24:598–604.
- 13. Endicott J, Nee J. Endicott work productivity scale (EWPS): a new measure to assess treatment effects. Psychopharmacol Bull. 1997;33:13–6.
- Ospina MB, Dennett L, Waye A, Jacobs P, Thompson AH. A systematic review of measurement properties of instruments assessing presenteeism. Am J Manag Care. 2015;21:e171–185.
- 15. Lu L, Cooper C, Yin LH. A cross-cultural examination of presenteeismand supervisory support. Career Dev Int. 2013;5:440–56.
- Li Y, Zhang J, Wang S, Guo S. The effect of presenteeism on productivity loss in nurses: the mediation of health and the moderation of general Self-Efficacy. Front Psychol. 2019;10:1745.

- Gärtner FR, Nieuwenhuijsen K, van Dijk FJH, Sluiter JK. Psychometric properties of the nurses work functioning questionnaire (NWFQ). PLoS ONE. 2011;6:e26565.
- Shan G, Wang S, Feng K, Wang W, Guo S, Li Y. Development and validity of the nurse presenteeism questionnaire. Front Psychol. 2021;12:679801.
- Wang S, Fan L. Research progress of nurses' attendance behavior. Chin J Nurs. 2024;59:634–40.
- Goetzel RZ, Hawkins K, Ozminkowski RJ, Wang S. The health and productivity cost burden of the top 10 physical and mental health conditions affecting six large U.S. Employers in 1999. J Occup Environ Med. 2003;45:5.
- 21. Ilmarinen J. Work ability–a comprehensive concept for occupational health research and prevention. Scand J Work Environ Health. 2009;35:1–5.
- Santo LD, Marognoli O, Previati V, Gonzalez CIA, Melis P, Galletta M. Providing personal care to patients: the role of nursing students' emotional labor. Int J Nurs Educ Scholarsh. 2019;16.
- Baris VK, Intepeler SS, Unal A. Development and psychometric validation of the sickness presenteeism Scale-Nurse. Int J Nurs Pract. 2023;29:e13168.
- 24. Furtado L, Coelho F, Pina S, Ganito C, Araújo B, Ferrito C. Delphi technique on nursing competence studies: A scoping review. Healthcare. 2024;12:1757.
- 25. Kong L, Lu T, Zheng C, Zhang H. Psychometric evaluation of the Chinese version of the positive health behaviours scale for clinical nurses: a cross-sectional translation. BMC Nurs. 2023;22:296.
- Sharif Nia H, Kaur H, Fomani FK, Rahmatpour P, Kaveh O, Pahlevan Sharif S, et al. Psychometric properties of the impact of events Scale-Revised (IES-R) among general Iranian population during the COVID-19 pandemic. Front Psychiatry. 2021;12:692498.
- Yu Y, Wan C, Huebner ES, Zhao X, Zeng W, Shang L. Psychometric properties of the symptom check list 90 (SCL-90) for Chinese undergraduate students. J Ment Health. 2019;28:213–9.
- Gunawan J, Marzilli C, Aungsuroch Y. Establishing appropriate sample size for developing and validating a questionnaire in nursing research. Belitung Nurs J. 2021;7:356–60.
- Shao Y, Zhang H, Zhang X, Liang Q, Zhang H, Zhang F. Chinese version of exercise dependence scale-revised: psychometric analysis and exploration of risk factors. Front Psychol. 2023;14:1309205.
- Tong L-K, Zhu M-X, Wang S-C, Cheong P-L, Van I-K. A Chinese version of the caring dimensions inventory: reliability and validity assessment. Int J Environ Res Public Health. 2021;18:6834.
- Liu Y, Zhang L, Li S, Li H, Huang Y. Psychometric properties of the Chinese version of the oncology nurses health behaviors determinants scale: a crosssectional study. Front Public Health. 2024;12:1349514.
- 32. Levin SF. Calibration and verification of measuring instruments: conceptual transformation. Meas Tech. 2022;64:871–82.
- Chen H. A Comparative Study on the Expressions of Complex Sentences of Causality between Chinese and Uygur. 2023;S1:75–7.
- Huang Z. Implicit cohesion in Chinese translation texts —— taking related words as an example. Foreign Lang Stud. 2021;07:55–64.
- Lu X, Wang L, Xu G, Teng H, Li J, Guo Y. Development and initial validation of the psychological capital scale for nurses in Chinese local context. BMC Nurs. 2023;22:28.
- Gebremedhin M, Gebrewahd E, Stafford LK. Validity and reliability study of clinician attitude towards rural health extension program in Ethiopia: exploratory and confirmatory factor analysis. BMC Health Serv Res. 2022;22:1088.
- Ferreira LK, Filgueiras Meireles JF, de Oliveira Gomes GA, Caputo Ferreira ME. Development and psychometric evaluation of a lifestyle evaluation instrument for older adults. Percept Mot Skills. 2023;130:1901–23.
- Ding J, Yu Y, Kong J, Chen Q, McAleer P. Psychometric evaluation of the student nurse stressor-14 scale for undergraduate nursing interns. BMC Nurs. 2023;22:1–10.
- Yang Z, Chen F, Lu Y, Zhang H. Psychometric evaluation of medication safety competence scale for clinical nurses. BMC Nurs. 2021;20:165.
- Hu W, Shang K, Wang X, Li X. Cultural translation of the ethical dimension: a study on the reliability and validity of the Chinese nurses' professional ethical dilemma scale. BMC Nurs. 2024;23:711.
- Cui Y-Y, Zhong X, Wen L-Y, Chen X-Y, Bai X-H. Cross-cultural adaptation and psychometric validation of the Chinese version of career success in nursing scale (CSNS). BMC Nurs. 2023;22:250.
- 42. Jenkinson C, Wright L, Coulter A. Criterion validity and reliability of the SF-36 in a population sample. Qual Life Res. 1994;3:7–12.
- 43. Hoseinzadeh E, Sharif-Nia H, Ashktorab T, Ebadi A. Development and psychometric evaluation of Nurse's intention to care for patients with infectious

disease scale: an exploratory sequential mixed method study. BMC Nurs. 2024;23:65.

- 44. Li J-Y, Wu X-X, Fan Y-R, Shi Y-X. Valuation of the cultural adaptation and psychometric properties of the Chinese version of the hidden curriculum evaluation scale in nursing education. Nurse Educ Pract. 2024;75:103880.
- Ma C, Zhao S. The myth and reality of employee wellbeing in China. In: Oruh ES, Adisa TA, editors. Employee wellbeing in the global South: A critical overview. Cham: Springer Nature Switzerland; 2024. pp. 145–77.
- 46. Feng H, Zhang M, Li X, Shen Y, Li X. The level and outcomes of emotional labor in nurses: A scoping review. J Nurs Adm Manag. 2024;2024:5317359.
- Wang H, Wang Z, Chen C, Wei W. Cross-Cultural adaptation and psychometric evaluation of the Chinese version of the authentic nurse leadership questionnaire. J Nurs Adm Manag. 2024;2024:1–10.

- Asadizaker M, Ebadi A, Molavynejad S, Yadollahi S, Saki Malehi A. Development and psychometric evaluation of the clinical nursing cultural competence scale. J Nurs Meas. 2023;31:615–25.
- Li C, Lin Y, Tosun B, Wang P, Guo HY, Ling CR, et al. Psychometric evaluation of the Chinese version of the BENEFITS-CCCSAT based on CTT and IRT: a cross-sectional design translation and validation study. Front Public Health. 2025;13:1532709.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.