Check for updates

Psychometric evaluation of the Chinese version of the BENEFITS-CCCSAT based on CTT and IRT: a cross-sectional design translation and validation study

Chuang Li[1,2], Youbei Lin[2], Betul Tosun[3], Pin Wang[1],
Hong Ye Guo[4], Cheng Rong Ling[5], Ran Qi[6], Qing Yue Luo[6],
Yan Wang[7], Fang Huang[8], Jiaqi Wang[2], Shu Hui Ma[2],
Dan Feng Xu[2], Shu Zhen Wu[2] and Lan Zhang[1]*

[1]First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China, [2]School of Nursing, Jinzhou Medical University, Jinzhou, China, [3]Faculty of Nursing, Hacettepe University, Ankara, Türkiye, [4]The First People's Hospital of Shenyang, Shenyang, China, [5]Department of Nursing, The Second People's Hospital of Yibin, Yibin, China, [6]School of Nursing, Jinzhou Normal College, Jinzhou, China, [7]School of Nursing, Liaodong University, Dandong, China, [8]School of Nursing Taizhou University, Taizhou, China

Background: The importance of culturally competent care in multicultural environments is increasingly recognized; however, effective tools to assess nursing students' cross-cultural competence remain limited. This study aimed to validate the BENEFITS-CCCSAT for Chinese nursing students.

Methods: The original BENEFITS-CCCSAT was translated, back-translated, culturally adapted, and pre-tested using the Brislin model to form a Chinese version. A combined approach of classical test theory (CTT) and item response theory (IRT) was then used for multidimensional validation.

Results: The CTT analysis showed that the C-BENEFITS-CCCSAT had a Cronbach's α coefficient of 0.80, dimension reliability values ranging from 0.700 to 0.905, a test−retest reliability value of 0.881, and a scale-level content validity index (S-CVI) value of 0.928. The criterion-related validity value was 0.619. The confirmatory factor analysis (CFA) indicated a good model fit (CMIN/DF = 1.071, RMSEA = 0.08), with factor loadings $\geq$0.50. The Rasch analysis showed an item reliability value of 1, person reliability values ranging from 0.76 to 0.89, item separation index values ranging from 17.37 to 60.34, and person separation index values ranging from 1.76 to 2.89. The information-weighted fit statistic mean square (infit MNSQ) and outlier-sensitive fit statistic mean square (outfit MNSQ) values for all items ranged from 0.86 to 1.27. Overall, the scale demonstrated good reliability and validity for the Chinese nursing students.

Conclusion: The 25-item C-BENEFITS-CCCSAT demonstrates good reliability and validity and can be applied in educational settings to assess students' ability to provide culturally competent care. Future studies should test the scale in culturally diverse populations to further determine its applicability and generalizability.

KEYWORDS

cross-cultural nursing, nursing students and education, psychometrics (MeSH), classical test theory, item response theory

1 Introduction

The world is currently experiencing an unprecedented rate of population mobility, driven by rapid advancements in the information age and the accelerated movement of people worldwide (1). In this context, the "Belt and Road" initiative has been introduced and widely promoted, further accelerating the internationalization of Chinese society. As a result, an increasing number of expatriates are coming to China for work, academic exchanges, and study (2). According to China's seventh population census, approximately 376 million people now constitute the floating population (3). Therefore, significant challenges and unique opportunities for transcultural nursing are posed by this increase in cultural diversity. Introduced by Madeleine Leininger, a renowned American nursing theorist, in the 1960s, the concept of transcultural nursing emphasizes its core principle of cultural care. Advocated by Leininger, this concept stresses that nurses should provide care that is safe, effective, and aligned with the values, beliefs, customs, and lifestyles inherent to the cultural backgrounds of their patients—elements shared, preserved, and transmitted across generations within each cultural group (4). Leininger's Sunrise Model conceptualizes transcultural nursing through four interconnected levels, each elucidating a distinct theoretical component and its relationships with the others. Encompassing worldviews, cultural norms, and social structures, the highest level addresses individuals' diverse perspectives and unique ways of life. Referred to as the service-object level, the second level highlights how individuals from specific cultures express their thoughts, emotions, and practices related to health and care. Focused on the healthcare system, the third level includes unique folk care practices within cultural groups. Finally, representing a targeted approach to nursing care, the fourth level—the decision-making and action level—emphasizes culturally congruent interventions (5).

As a critical workforce in the healthcare industry, nurses must possess cultural competence to address the complexities arising from diverse cultures, traditions, dietary practices, religious beliefs, and thought processes encountered in daily nursing practice (6). Underscoring this need, the urgent call for transcultural nursing education arises, aiming to equip nursing students with the necessary knowledge, skills, and assessment tools to deliver culturally competent care (7). However, a systematic review examining the relationship between educational strategies promoting cultural competence and patient treatment outcomes revealed a lack of standardized approaches. The absence of systematic educational strategies and assessment methods has led to inconclusive effects on improving cultural competence among healthcare professionals (8). Many existing assessment tools are based on specific theoretical models. For instance, Choi and Kim (9) utilized the "Nursing Students' Cultural Competence Scale" in their study, while Ge Yunyun developed the Cross-Cultural Sensitivity Scale for nursing students in 2006 (10). These tools highlight efforts to measure and enhance cultural competence but also reflect the need for more consistent and validated methodologies.

In the context of a multicultural environment, effective nursing education aimed at enhancing transcultural nursing skills requires robust evaluation tools. The better and effective nursing education for improving transcultural nursing skills cultural competence and cultural sensitivity assessment tool (BENEFITS-CCCSAT) integrates the strengths and features of existing instruments, providing a comprehensive framework for assessing the outcomes of cross-cultural nursing education, as well as cultural competence and sensitivity (7). In contrast, transcultural sensitivity measurement tools developed independently by Chinese scholars remain limited, both in their scope of application and in the scientific validation of their effectiveness (11). To address this gap, the present study aimed to adapt the BENEFITS-CCCSAT for the Chinese context, evaluate its reliability and validity, and explore its applicability and potential value in assessing and fostering cultural competence among nursing students in China.

2 Method

2.1 Study design

This research was a cross-sectional, multicenter study.

2.2 Data collection

The conclusions of this study were drawn from the data collected between April and August 2024. A total of 1,074 nursing students, recruited through convenience sampling from eight medical universities in Northeast, Southwest, and North China, participated. The survey was distributed via the China Questionnaires Star platform[1] and administered under standardized instructions. The students were informed about the study's purpose, the scale's instructions, and the estimated time for completion, and those who provided consent were included. To ensure data integrity, the e-questionnaire could only be submitted after all items were fully completed, and only one submission per participant was allowed. The sample size was estimated using the Kendall method, based on the recommendation of 5–10 times the number of items in the questionnaire, while accounting for a potential sample loss rate of 10–20% (12). With 26 items in the questionnaire, the required sample size was calculated to range from 143 to 312 participants. To meet the sample size threshold for confirmatory factor analysis (CFA) (minimum of 200 cases) and ensure the study's generalizability (13), approximately 1,200 questionnaires were distributed. After excluding the responses with completion times under 3 min and those with obviously patterned answers, 1,074 valid questionnaires were retained, resulting in a recovery rate of 89.5%.

2.3 Participants

The inclusion criteria were as follows: (1) Full-time college degree or above; (2) voluntary participation in the research; (3) previous clinical apprenticeship or internship experience, and (4) knowledge about basic nursing skills and intervention measures. The exclusion criteria were as follows: (1) Students with mental illness, such as emotional disorder and depression, as it is considered that students with mental illness may experience serious emotional or cognitive difficulties, which make it difficult to answer questions.

---

1 https://www.wjx.cn/

2.4 Tools

2.4.1 The researcher created a general information questionnaire

The data collected included information on educational level, grade, sex, age, ethnicity, home residence, and whether participants had attended any courses or training related to cross-cultural care.

2.4.2 Chinese version of the BENEFITS−CCCSAT (C-BENEFITS-CCCSAT)

It consists of five dimensions with 25 items, as follows: respect for cultural diversity (6 items), challenges and barriers providing culturally competent care (3 items), achieving cultural competence (3 items), culturally sensitive communication (5 items), and perceived meaning of cultural care (8 items). Seven items (items 7, 8, 9, 14, 15, 17, and 18) are reverse scored. A seven-point Likert scale is used, with one representing "strongly disagree" and seven representing "strongly agree." The overall score ranges between 25 and 175, with higher scores indicating greater cultural competence, cultural sensitivity, and intercultural nursing skills among nursing students.

2.4.3 Chinese version of the short form of the cultural competence scale for nurses (C-SFCCSN)

The instrument, developed by Chae et al. (14) and translated by Yajin Zhu (15), was designed to assess the cultural competence of clinical nurses. It consists of four dimensions: cultural awareness (6 items), cultural knowledge (7 items), cultural sensitivity (13 items), and cultural skills (7 items), for a total of 33 items. Each item is scored on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The total score can range from 33 to 231, with higher scores indicating greater cultural competence in clinical nurses.

2.5 Translation and cultural adaptation of the BENEFITS-CCCSAT

The BENEFITS-CCCSAT was developed by Ayla (7). It consists of five dimensions with 26 items, as follows: respect for cultural diversity (1–6 items), challenges and barriers providing culturally competent care (7–10 items), achieving cultural competence (11–13 items), culturally sensitive communication (14–18 items), and perceived meaning of cultural care (19–26 items). A seven-points Likert scale is used, with one representing "strongly disagree" and seven representing "strongly agree." The overall score is between 26 and 182, with higher scores indicating greater cultural competence, cultural sensitivity, and intercultural nursing skills among nursing students. In addition, eight items (items 7, 8, 9, 10, 14, 15, 17, and 18) are reverse scored. The BENEFITS-CCCSAT has good construct validity and reliability, with an internal consistency Cronbach's α coefficient of 0.828.

To obtain permission to use the scale, the research team contacted the original developers via email. The translation process closely followed the Brislin translation model (16).

2.5.1 Step 1

The original English version of the scale was translated into two Chinese versions (B1 and B2) by two native Chinese speakers who were nursing master's degree candidates. The research team then reviewed and discussed the two Chinese versions, resolving any discrepancies and combining them into a preliminary Chinese version, B3.

2.5.2 Step 2

Two native Chinese educators, unfamiliar with the questionnaire, independently translated the Chinese version B3 back into English (producing versions TB1 and TB2). These versions were then combined to form the English version TB3. The original authors were asked to review TB3 and provide feedback. The research team further deliberated and made revisions before finalizing the Chinese version of the scale, F1.

2.5.3 Step 3

To culturally adapt the Chinese version F1, the Delphi method was employed. Six experts were recruited: two clinical nursing experts and four nursing education specialists, all with associate senior titles and at least a master's degree (two with master's degrees and four with doctoral degrees). The experts, with an average age of 35.43 ± 5.53 years and an average of 13.14 ± 4.02 years of relevant experience, assessed the cultural relevance, contextual appropriateness, and linguistic expression of the items in the Chinese version F1, comparing them with the original scale. Based on the experts' feedback, the research team finalized the questionnaire through necessary revisions.

2.5.4 Step 4

A total of 30 randomly selected nursing students were given a questionnaire for the presurvey, and all the students responded that all the items on the scale were comprehensible.

2.6 Ethical statement

The study was approved by the Ethics Review Committee of Jinzhou Medical University (JZMULL2025043) and adhered to their ethical guidelines. Informed consent was obtained from all participating students to ensure their confidentiality and anonymity. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki.

2.7 Statistical analysis

Statistical analysis was performed using SPSS 27.0, AMOS 26.0, and Winsteps 3.72.3. Prior to the analysis, the data were cleaned to address invalid and missing values. Descriptive statistics, including mean and standard deviation (mean ± SD), were used to summarize the quantitative data following a normal distribution, while frequency and percentage (%) were used to describe the qualitative data.

2.7.1 Analysis of the CTT model

This study involved the analysis of classical test theory (CTT) data, including assessments of the reliability and validity of the scales. Reliability was evaluated through internal consistency, split-half reliability, and test–retest reliability (17). Regarding validity, content validity was established using the Delphi method (18), while structural validity was assessed through confirmatory factor analysis (CFA),

which evaluated convergent and discriminant validity. The validity of the calibration correlations was determined using the Pearson correlation test between the C-SFCCSN and the C-BENEFITS-CCCSAT [19].

2.7.2 Analysis of the IRT model

The validity of the Rasch analysis requires confirmation of unidimensionality, which was evaluated through principal component analysis (PCA) of residuals [20]. Reliability was assessed using person and item separation indices to determine the scale's discriminative capacity across dimensions, alongside person and item reliability metrics. Item fit was examined using the following indices [21]: ① Information-weighted fit statistic mean square (infit MNSQ), ② outlier-sensitive fit statistic mean square (outfit MNSQ), and ③ point-measure correlation (PT-Measure Corr). The residual patterns were further analyzed via PCA. Item-person fit was visualized using Wright maps, and the appropriateness of the response category thresholds was verified through item characteristic curves. Finally, differential item functioning (DIF) analysis was conducted to identify potential group-based measurement bias.

3 Results

3.1 Demographic characteristics

A total of 1,074 students participated in the validation study, with a mean age of 20.48 years (range:17–34; SD = 1.986). The sample consisted predominantly of female (68.5%) and undergraduate students (50.0%). Second-year students represented the largest group (55.8%). Detailed demographic information is provided in Table 1.

3.2 CTT model results

3.2.1 Item analysis

The recovered scales were sorted by total score, from high to low, with the top 27% defined as the high group and the bottom 27% as the low group. A normality test was then conducted, with skewness and kurtosis values between −2 and + 2 indicating that the data followed a normal distribution [22]. A test of 579 cases confirmed that the data adhered to a normal distribution. Item analysis was performed using the two independent samples t-test, which revealed critical ratios ranging from 5.180 to 19.579 (all >3) and a p-value of <0.001 [23], indicating statistical significance. As a result, the items were considered to be well differentiated (Table 2).

3.2.2 Reliability

The Cronbach's α coefficient for the C-BENEFITS-CCCSAT was 0.80, and the reliability of the individual dimensions were 0.880, 0.700, 0.759, 0.840, and 0.905, respectively. The split-half reliability was 0.837. Two weeks later, 40 students were randomly selected for a test–retest reliability assessment, and the test-retest reliability of the C-BENEFITS-CCCSAT was 0.881 (Table 3). All values were greater than the reference value of 0.7 [24], indicating good internal consistency and measurement invariance.

TABLE 1 Distribution of the participant attributes and demographic data.

Variables		Frequency (n = 1,074)	Percentage (%)
Sex	Male	338	31.5
	Female	736	68.5
Educational level	College students	328	30.5
	Undergraduate students	537	50.0
	Graduate students	209	19.5
Grade	First	134	12.5
	Second	599	55.8
	Third	307	28.6
	Fourth	34	3.2
Ethnicity	Han	712	66.3
	National minority	362	33.7
Home residence	Countryside	576	53.6
	Municipalities	498	46.4
WPHACOTCCC	Yes	493	45.9
	No	581	54.1

WPHACOTCCC, whether participants had attended any courses or training related to cross-cultural care.

3.2.3 Content validity

Six experts (two clinical experts and four nursing education experts) were invited to evaluate the cultural appropriateness and relevance of each item in the C-BENEFITS-CCCSAT. The expert panel consisted of four Ph.D. holders and two M.D. holders, all holding the title of Associate Professor or higher, with professional experience ranging from 10 to 21 years. A 4-point Likert scale was employed, where "Not Relevant" was scored as 1, "Weakly Relevant" as 2, "More Relevant" as 3, and "Very Relevant" as 4. The results indicated that the content validity index of the expert evaluation scale (S-CVI) was 0.928(>0.9) and the content validity index for each item (I-CVI) ranged from 0.83 to 1.00(>0.78) [25].

3.2.4 Structural validity

The prerequisite for conducting factor analysis was calculating the KMO value and performing Bartlett's test of sphericity. The results showed that the KMO value for the C-BENEFITS-CCCSAT was 0.884, which was greater than 0.7 [26]. The Bartlett's test of sphericity yielded a chi-square value of χ^2 = 10,503.709 with df = 300, indicating a statistically significant difference ($p < 0.001$) and confirming the suitability of the data for factor analysis. Confirmatory factor analysis (CFA) was performed to assess the item-factor structure of the scale, and the model was estimated using the robust maximum likelihood method. The model showed a chi-square/degrees of freedom (CMIN/DF) ratio of 1.071 (less than 3), a root mean square error of approximation (RMSEA) value of 0.08 (≤0.08), and the following fit indices: the Comparative Fit Index (CFI) = 0.998, Normed Fit Index (NFI) = 0.973, Incremental Fit Index (IFI) = 0.998, and Tucker-Lewis Index (TLI) = 0.998, all greater than 0.9 [27]. These results suggested that the model fit was

TABLE 2 The C-BENEFITS-CCCSAT values for skewness, kurtosis, and the critical ratio.

Items	Mean (SD)	Skewness/ Kurtosis	Critical ratio	95% CI	
Rfcd1	3.62(1.72)	0.144/−0.923	14.531	1.540	2.022
Rfcd2	4.56(1.70)	−0.354/−0.801	13.188	1.392	1.879
Rfcd3	3.44(1.72)	0.331/−0.836	12.951	1.387	1.883
Rfcd4	4.50(1.64)	−0.242/−0.730	12.959	1.320	1.792
Rfcd5	4.15(1.70)	−0.089/−0.886	12.442	1.318	1.812
Rfcd6	4.93(1.60)	−0.491/−0.648	12.664	1.266	1.731
CB7	1.68(0.89)	1.414//1.809	5.180	0.235	0.523
CB8	4.54(1.65)	−0.253/−0.782	5.755	0.506	1.031
CB9	5.52(1.38)	−0.893/0.269	5.210	0.365	0.806
Acc11	3.27(1.66)	0.385/−0.755	8.579	0.861	1.372
Acc12	4.09(1.70)	−0.005/−0.927	6.662	0.641	1.177
Acc13	2.71(1.48)	0.716/−0.197	6.378	0.525	0.993
Csc14	3.81(1.69)	0.168/−0.939	11.398	1.200	1.700
Csc15	4.01(1.74)	−0.022/−0.936	10.725	1.163	1.684
Csc16	4.65(1.62)	−0.306/−0.716	11.271	1.140	1.621
Csc17	2.83(1.55)	0.632/−0.415	9.386	0.893	1.365
Csc18	5.02(1.54)	−0.571/−0.391	10.236	0.980	1.445
Pmoc19	3.92(1.76)	0.013/−0.936	19.355	1.986	2.435
Pmoc20	3.64(1.79)	0.189/−1.044	17.926	1.912	2.383
Pmoc21	5.14(1.61)	−0.584/−0.684	16.555	1.611	2.045
Pmoc22	3.70(1.76)	0.252/−0.936	18.500	1.926	2.383
Pmoc23	3.30 (1.77)	0.416/−0.893	18.669	1.942	2.399
Pmoc24	4.44(1.71)	−0.202/−0.928	17.417	1.785	2.239
Pmoc25	3.74(1.81)	0.076/−1.024	19.579	2.063	2.523
Pmoc26	5.01(1.67)	−0.568/−0.614	17.144	1.725	2.171

good and that the factor loadings for all items were ≥ 0.50 (28) (Figure 1).

3.2.5 Convergent and discriminant validity

Based on the standardized factor loadings from the CFA, the composite reliability (CR) values ranged from 0.700 to 0.880, all ≥0.7 (29). The average variance extracted (AVE) values ranged from 0.420 to 0.55, all >0.4 (30). Discriminant validity was tested using the Fornell–Larcker criterion, which revealed that the square root of the AVE value for each latent variable was greater than the correlation coefficients between that latent variable and the other latent variables (31). This indicated that the C-BENEFITS-CCCSAT demonstrated good convergent and discriminant validity (Table 3).

3.2.6 Calibration correlation validity

Calibration correlation validity assesses the degree of correlation between a new instrument and an existing, authoritative, validated scale (32). A higher correlation coefficient indicates better validity for the new instrument. In this study, the C-SFCCSN was used as the reference standard for intercultural competence. The results showed that the calibration correlation validity of the C-BENEFITS-CCCSAT

was 0.619 (33), which was greater than 0.5 but less than 0.7, indicating that the calibration correlation validity of this questionnaire for use with a population of Chinese nursing students is of moderate relevance.

3.3 IRT model results

3.3.1 Unidimensionality test

The PCA of the residuals indicated a first principal component standardized residual value of 5.6 (>3.0) (34), suggesting multidimensionality, which violates the assumptions of the Rasch model. This could lead to a poorer fit and inaccurate estimation of the scale. However, Van der Linden argued that despite the potential multidimensionality of the overall scale (35), the Rasch model can still be applied, particularly when the dimensions are clearly defined. Specifically, the Rasch model not only evaluates the fit of each item but also analyzes the independence and validity of each dimension when the model is appropriately configured. Therefore, even if the scale is multidimensional, analyzing each dimension separately ensures that it accurately measures the underlying concept while avoiding inter-dimensional interference, thereby maintaining the validity of the measurement. In addition, related studies (36, 37) have indicated that dimensional analyses using the Rasch model can effectively capture independent information across multiple dimensions of the scale, successfully identifying underlying constructs. Drawing from these perspectives, we argue that when a scale exhibits multidimensionality, it suggests that the construct being measured encompasses multiple independent yet related sub-concepts. Each dimension may reflect a distinct aspect or domain, with items within each dimension measuring specific characteristics of that domain. Although the overall scale does not meet the unidimensionality assumption, it remains suitable for measurement as long as each individual dimension satisfies Rasch's unidimensionality requirement. Therefore, based on the five dimensions classified by the original authors, a unidimensionality test was conducted to isolate the contribution of each dimension and ensure that each dimension's effect was accurately estimated. The results showed that the standardized residual value of the first component for each dimension was less than 3.0 (Table 4), indicating that each dimension met the unidimensionality assumption and that there was no cross-dimensional overlap. This suggests that the scale remains a valuable tool for assessing transcultural nursing competencies.

3.3.2 Reliability and separation indices

After conducting the Rasch analysis, the item reliability for each dimension of the C-BENEFITS-CCCSAT was 1.00, and the person reliability ranged from 0.76 to 0.86, both surpassing the critical threshold of 0.7 (38). The item separation indices ranged from 17.37 to 60.34, while the person separation indices ranged from 1.76 to 2.89, both exceeding the minimum acceptable value of 1.5 (39). These results indicate that both the sample and the items were well represented (Table 5).

3.3.3 Fit of the items

Some researchers have proposed that the ideal criteria for the fit of item analysis include infit MNSQ and outfit MNSQ values between 0.6 and 1.4 (34), which indicate a good fit to the model in item analysis. In this study, the infit MNSQ values for each item of the

TABLE 3 Reliability, convergent, and discriminant validity of the C-BENEFITS-CCCSA.

Dimensions	Cronbach's alpha coefficient	Split-half reliability	Test–retest reliability	Convergent Validity		Fornell–Larcker test				
				CR	AVE	F1	F2	F3	F4	F5
F1	0.880	0.837	0.881	0.880	0.550	**0.741**	–	–	–	–
F2	0.700	–	–	0.700	0.420	0.004	**0.648**	–	–	–
F3	0.759	–	–	0.761	0.515	−0.018	−0.023	**0.717**	–	–
F4	0.830	–	–	0.841	0.514	0.003	0.000	0.039	**0.716**	–
F5	0.905	–	–	0.870	0.500	−0.014	−0.015	0.067	0.000	**0.707**

The square root of the AVE is the value on the diagonal of the Fornell–Larcker test; F1: Respect for cultural diversity; F2: Challenges and barriers providing culturally competent care; F3: Achieving cultural competence; F4: Culturally sensitive communication; F5: Perceived meaning of cultural care. Bold values represents the square root of the AVE.
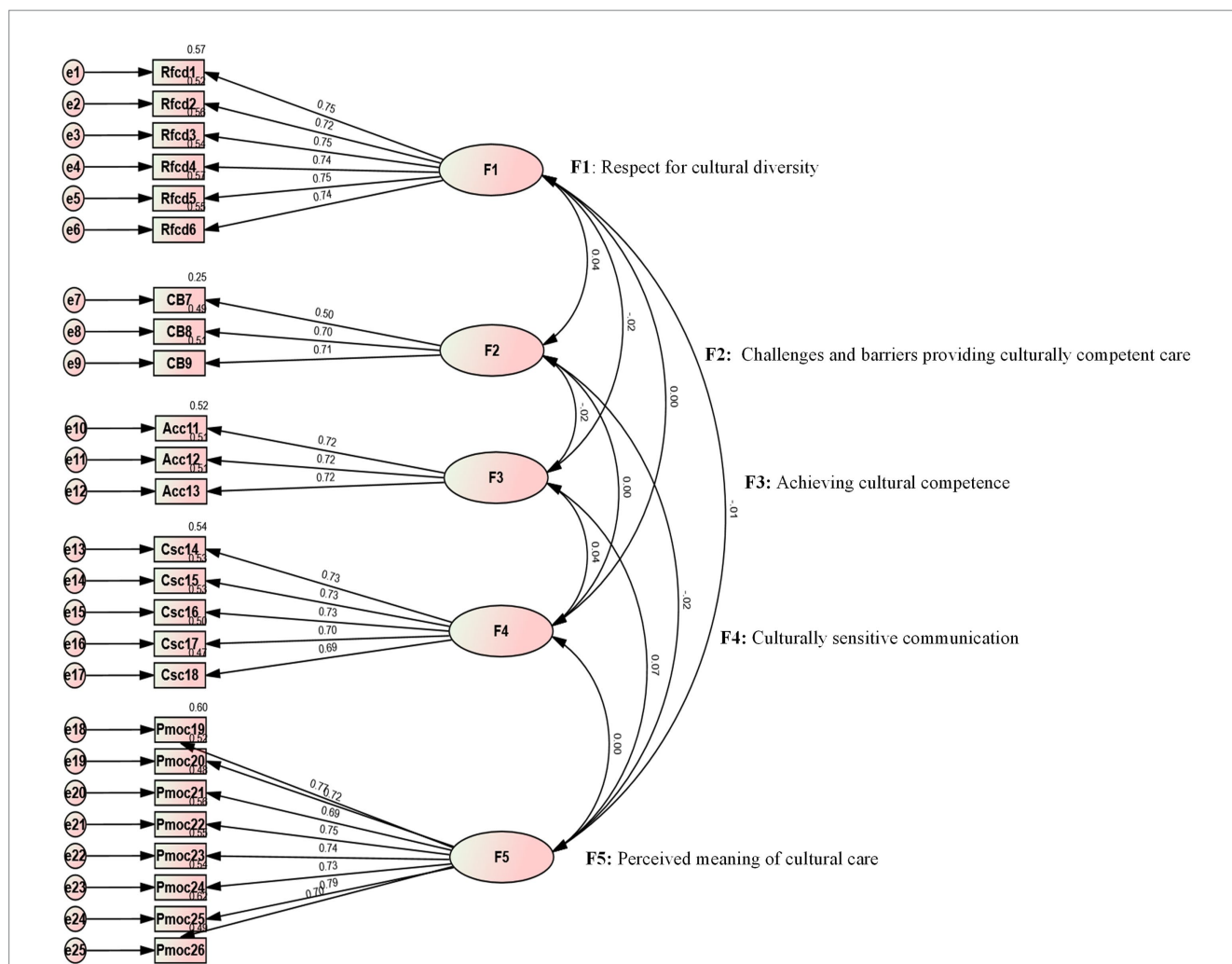

FIGURE 1
Five-factor standardized factor loadings plot of the C-BENEFITS-CCCSAT ($n = 1,074$).

C-BENEFITS-CCCSAT ranged from 0.87 to 1.08, and the outfit MNSQ values ranged from 0.86 to 1.27. The correlation coefficient measures how closely the items align with the measurement target, with an acceptable minimum value of 0.5 (32). In this study, the PT-Measure Corr values ranged from 0.65 to 0.84, exceeding the minimum reference value, suggesting that the scale items were closely aligned with the measurement target and that the study data fit well with the model (Table 5).

3.3.4 Item-person matching

Wright's map, which converts the original scores of individual ability and item difficulty into logit values on the same scale, visually illustrates the suitability of items for individuals (40). It serves as one of the key indicators of the overall quality of the scale. The map simultaneously displays the ability levels of the nursing students who took the test and the difficulty of all the items on the C-BENEFITS-CCCSA scale. Ideally, the mean value for both sides should be close to 0, with the difference

between them being less than 1 logit (41). A difference greater than 1 logit typically indicates a mismatch between individual ability and item difficulty, suggesting that the individual may not be a good fit for the test.

TABLE 4 Overall scale and Rasch residual analyses of each dimension.

Dimension	Eigenvalue of first contrast residual	Explanatory variance (%)
Overall Scale	5.6	35.0
Respect for cultural diversity,	1.3	66.6
Challenges and barriers providing culturally competent care	1.7	87.7
Achieving cultural competence	1.6	71.7
Culturally sensitive communication	1.3	69.5
Perceived meaning of cultural care	1.3	66.8

A difference of 1 logit often signifies that the individual's ability does not align with the difficulty level of the items. The "S" and" T" represent one and two times the standard deviation, respectively. On the left side of the map, the distribution of individual abilities is shown, with higher positions indicating greater ability. On the right side, the distribution of item difficulty is displayed, with higher positions reflecting more difficult items. Figure 2 demonstrates that the overall fit between individual ability and item difficulty across the five dimensions was good, with a similar distribution. The difference in the mean values of the measures did not exceed 1 logit, and the distribution of individual abilities was nearly normal, with most participants concentrated in the middle ability range. The majority of the items on the scale were also located in this region, suggesting that the scale is appropriate for most nursing students. However, the scale's overall difficulty did not fully accommodate nursing students across all ability ranges. Future studies could consider adding items with higher and lower difficulty levels, as well as adjusting the difficulty intervals between items.

3.3.5 Response category fit

The C-BENEFITS-CCCSAT includes seven response categories: "Strongly Disagree," "Disagree," "Somewhat Disagree," "Neutral," "Somewhat Agree," "Agree," and "Strongly Agree." Table 6 presents the

TABLE 5 Analysis of the fit of the C-BENEFITS-CCCSAT and separation indices and reliability values for each dimension.

| Items	INFIT MNSQ	OUTFIT MNSQ	PT-Measure Corr	Person		Item	
				Reliability	Separation	Reliability	Separation
Rfcd1	0.97	0.97	0.79	0.86	2.58	1.00	17.37
Rfcd2	1.05	1.03	0.77				
Rfcd3	0.96	0.95	0.79				
Rfcd4	0.99	1.01	0.78				
Rfcd5	0.97	0.97	0.79				
Rfcd6	0.96	0.97	0.77				
CB7	0.98	1.27	0.65	0.76	1.76	1.00	60.34
CB8	0.94	0.93	0.84				
CB9	0.91	0.97	0.81				
Acc11	1.00	0.99	0.81	0.76	1.80	1.00	21.08
Acc12	0.95	0.94	0.83				
Acc13	0.99	0.99	0.79				
Csc14	0.97	0.97	0.78	0.83	2.24	1.00	24.95
Csc15	1.01	1.01	0.78				
Csc16	0.97	0.96	0.78				
Csc17	0.96	0.97	0.76				
Csc18	1.00	1.05	0.75				
Pmoc19	0.93	0.94	0.79	0.89	2.89	1.00	20.95
Pmoc20	1.04	1.03	0.76				
Pmoc21	1.02	1.02	0.72				
Pmoc22	0.97	0.97	0.77				
Pmoc23	1.00	0.99	0.77				
Pmoc24	1.00	0.99	0.76				
Pmoc25	0.87	0.86	0.81				
Pmoc26	1.08	1.13	0.73				
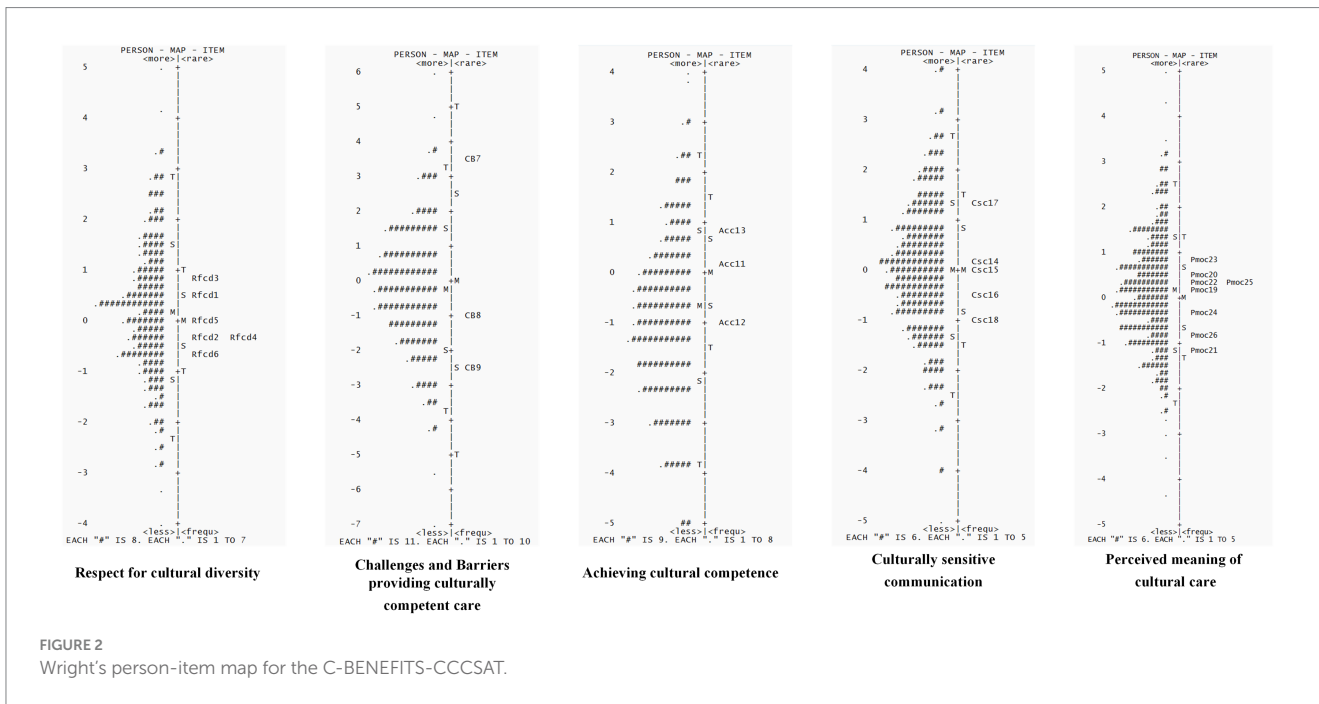
FIGURE 2
Wright's person-item map for the C-BENEFITS-CCCSAT.

TABLE 6 Response category fit for the C-BENEFITS-CCCSAT.

Response category	Score	F1				F2				F3			
		PACO	RCDP	Infit	Outfit	PACO	RCDP	Infit	Outfit	PACO	RCDP	Infit	Outfit
Strongly Disagree	1	7%	−1.88	1.03	1.03	20%	−4.50	1.07	1.06	15%	−2.80	1.08	1.05
Disagree	2	12%	−1.25	0.96	0.97	14%	−2.77	1.05	1.38	21%	−1.85	0.98	0.99
Somewhat Disagree	3	17%	−0.54	1.00	0.97	10%	−1.16	0.90	1.09	20%	−1.01	0.96	0.95
Neutral	4	19%	0.01	0.99	0,97	13%	0.27	0.91	0.90	17%	−0.19	0.94	0.92
Somewhat agree	5	19%	0.65	0.96	0,97	13%	1.36	0.87	0.86	14%	0.61	0.91	0.90
Agree	6	17%	1.35	1.00	0.99	16%	2.29	0.90	1.03	9%	1.42	1.06	1.07
Strongly agree	7	10%	2.17	0.98	0.99	14%	3.58	0.94	0.96	4%	2.39	0.92	0.95

Response Category	Score	F4				F5				−	−	−	−
		PACO	RCDP	Infit	Outfit	PACO	RCDP	Infit	Outfit	−	−	−	−
Strongly Disagree	1	9%	−2.22	1.03	1.02	8%	−1.81	1.00	1.00	−	−	−	−
Disagree	2	15%	−1.32	0.94	0.97	14%	−1.13	0.97	0.95	−	−	−	−
Somewhat Disagree	3	17%	−0.65	0.97	0.99	16%	−0.58	0.93	0.95	−	−	−	−
Neutral	4	18%	0.05	0.91	0.86	18%	0.02	0.99	1.03	−	−	−	−
Somewhat Agree	5	18%	0.70	0.97	1.00	18%	0.61	0.96	0.96	−	−	−	−
Agree	6	14%	1.35	1.02	1.06	16%	1.22	1.06	1.05	−	−	−	−
Strongly Agree	7	10%	2.14	1.05	1.04	10%	2.08	1.00	1.00	−	−	−	−

PACO, percentage of the appearance of the category in the observed data; RCDP, Response category difficulty parameter; F1, Respect for cultural diversity; F2, Challenges and barriers providing culturally competent care; F3, Achieving cultural competence; F4, Culturally sensitive communication; F5, Perceived meaning of cultural care.

percentage of occurrences for each response category, the selection difficulty parameters, and the mean square (MNSQ) values for both the infit and outfit statistics. As shown in the table, the difficulty levels of the response categories were calibrated to follow the expected ascending order. According to the infit and outfit statistics, all response categories were statistically appropriate, with the MNSQ values falling within the acceptable range (0.6–1.4) (34). Given that these values represent the ideal fit for infit and outfit statistics, the C-BENEFITS-CCCSAT demonstrated a high degree of fit for the individuals, items, and categories within the rating scale model. Figure 3 (① F1–F5) illustrates the thresholds of the C-BENEFITS-CCCSAT ordered by difficulty, with nearly identical discrimination across the response options. This aligns with the assumptions of the rating scale model, where the difficulty parameter is a key factor influencing the probability of a correct response. Specifically, as difficulty increases, the likelihood of an individual answering a question correctly decreases, while the discrimination parameter remains consistent across all items, indicating that the scale's ability to differentiate across various difficulty levels is well-balanced (42). In addition, Figure 3 (② F1–F5) demonstrates that the difficulty of the response options increased progressively, with the most difficult option being the correct one. This suggests that the scale effectively reflected the varying ability levels of the respondents. As the ability level increased, the probability of selecting more difficult options also increased, with the most difficult option (typically representing "Strongly Agree" or similar) ultimately becoming the correct answer. This indicates that the scale can effectively differentiate between different trait levels, supporting the assumptions of the scale's design. These findings demonstrate that the C-BENEFITS-CCCSAT is designed to effectively distinguish between varying levels of ability and item difficulty, making it suitable for a wide range of nursing student populations (43). The peak values of the five-dimensional item characteristic curves were all differentiated from each other and in the same order, indicating that there is a good degree of differentiation among the dimensional items.

3.3.6 Item function test (DIF)

Differences in item functioning refer to the variation in responses to a given item between individuals of the same ability level across different subgroups (44). When the absolute difference between two groups exceeds 1 logit, it is considered a substantial difference in contrast, indicating that the item is biased (32). The present study found that the DIF contrast values below 0.5 logit indicated measurement invariance in the item functioning test performed among the nursing students from the subgroups with or without cross-culturally relevant training or education (45). These results suggest that the C-BENEFITS-CCCSAT is unbiased in measuring populations with different characteristics (Table 7).

4 Discussion

Cross-cultural care aims to address the unique needs of patients from diverse cultural backgrounds by providing culturally competent care (46). However, the implementation of intercultural care in healthcare settings faces numerous challenges (47), one of which is the lack of effective assessment tools. This study translated advanced foreign tools for assessing transcultural nursing skills, cultural competence, and sensitivity, with the goal of providing a scientifically valid and effective assessment instrument for the healthcare field. By doing so, it aimed to promote the development of transcultural nursing, enhance the quality of care, and better meet the diverse nursing needs of patients. However, we encountered several challenges during the study. Differences in expression between languages may have caused some items on the scale to fail in accurately conveying the original measurement intent after translation, potentially affecting their validity and reliability across cultural contexts. To address this, we employed a "direct translation-back translation" method and conducted two rounds of expert review to ensure the scale's applicability and cultural sensitivity across various languages and cultures (48). Despite these efforts, during the first validation,

FIGURE 3
The item characteristic curve of the five dimensions of the C-BENEFITS-CCCSAT (F1: Respect for cultural diversity; F2: Challenges and barriers providing culturally competent care; F3: Achieving cultural competence; F4: Culturally sensitive communication; F5: Perceived meaning of cultural care).
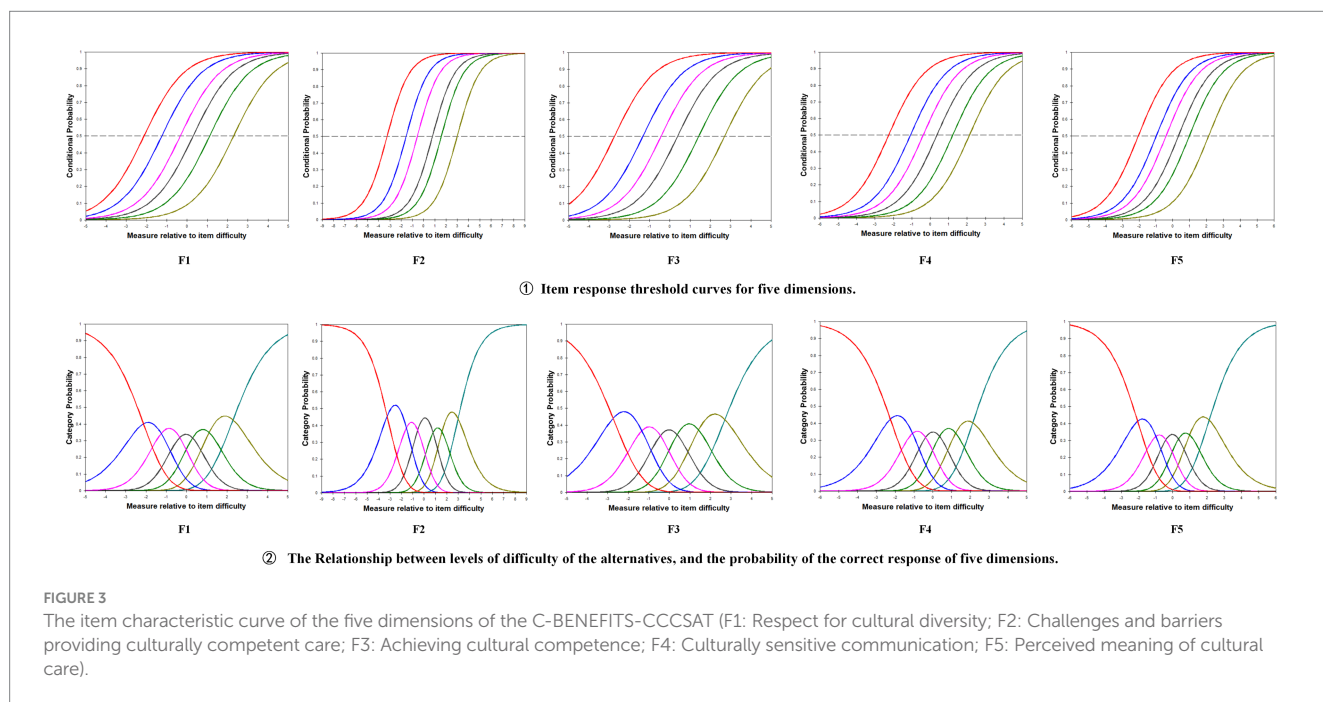
TABLE 7 DIF analysis for the C-BENEFITS-CCCSAT based on whether the participants had attended any courses or training related to cross-cultural care.

WPHACOTCCC	Items	DIF Measure	DIF score	DIF size	DIF SE	DIF t	DIF Contrast	*P*-value
Yes	Rfcd1	0.15	−0.01	0.00	0.03	0.00	0.00	1.000
No	Rfcd1	0.15	0.02	0.00	0.03	0.00	0.00	1.000
Yes	Rfcd2	−0.23	−0.07	0.03	0.03	1.12	0.07	0.261
No	Rfcd2	−0.30	0.08	−0.04	0.03	−1.17	−0.07	0.242
Yes	Rfcd3	0.24	−0.03	0.00	0.03	0.00	0.00	1.000
No	Rfcd3	0.24	0.04	0.00	0.03	0.00	0.00	1.000
Yes	Rfcd4	−0.22	−0.02	0.00	0.03	0.00	0.00	1.000
No	Rfcd4	−0.22	0.02	0.00	0.03	0.00	0.00	1.000
Yes	Rfcd5	−0.08	0.03	0.00	0.03	0.00	0.00	1.000
No	Rfcd5	−0.08	−0.03	0.00	0.03	0.00	0.00	1.000
Yes	Rfcd6	−0.45	0.05	−0.02	0.03	−0.85	−0.06	0.394
No	Rfcd6	−0.39	−0.07	0.03	0.03	1.01	0.06	0.312
Yes	CB7	1.40	0.00	0.00	0.05	0.00	0.00	1.000
No	CB7	1.40	0.01	0.00	0.05	0.00	0.00	1.000
Yes	CB8	−0.25	−0.04	0.00	0.03	0.00	0.02	1.000
No	CB8	−0.27	0.05	−0.02	0.03	0.70	−0.02	0.483
Yes	CB9	−0.71	−0.05	0.03	0.03	0.92	0.06	0.360
No	CB9	−0.77	0.05	−0.03	0.04	−0.87	−0.06	0.382
Yes	Acc11	0.31	−0.03	0.00	0.03	0.00	0.00	1.000
No	Acc11	0.31	0.04	0.00	0.03	0.00	0.00	1.000
Yes	Acc12	−0.06	−0.01	0.00	0.03	0.00	0.00	1.000
No	Acc12	−0.06	0.01	0.00	0.03	0.00	0.00	1.000
Yes	Acc13	0.06	0.06	−0.03	0.03	−1.00	−0.06	0.316
No	Acc13	0.57	−0.06	0.03	0.03	0.96	0.06	0.337
Yes	Csc14	0.63	0.00	0.00	0.03	0.00	0.00	1.000
No	Csc14	0.07	0.00	0.00	0.03	0.00	0.00	1.000
Yes	Csc15	0.07	−0.01	0.00	0.03	0.00	0.00	1.000
No	Csc15	0.01	0.01	0.00	0.03	0.00	0.00	1.000
Yes	Csc16	0.01	−0.04	0.00	0.03	0.00	0.02	1.000
No	Csc16	−0.27	0.05	−0.02	0.03	−0.68	−0.02	0.498

(Continued)

TABLE 7 (Continued)

WPHACOTCCC	Items	DIF Measure	DIF score	DIF size	DIF SE	DIF t	DIF Contrast	*P*-value
Yes	Csc17	−0.29	−0.08	0.04	0.03	1.40	0.10	0.162
No	Csc17	0.59	0.10	−0.05	0.03	1.63	−0.10	0.104
Yes	Csc18	0.50	−0.03	0.00	0.05	0.00	−0.06	1.000
No	Csc18	−0.46	0.03	0.00	0.05	0.00	0.06	1.000
Yes	Pmoc19	−0.46	0.06	−0.03	0.03	−0.99	0.00	0.323
No	Pmoc19	0.01	−0.07	0.03	0.03	1.06	0.00	0.290
Yes	Pmoc20	0.07	0.03	0.00	0.03	0.00	0.00	1.000
No	Pmoc20	0.18	−0.03	0.00	0.04	0.00	0.00	1.000
Yes	Pmoc21	0.18	0.02	0.00	0.03	0.00	−0.05	1.000
No	Pmoc21	−0.59	−0.02	0.00	0.03	0.00	0.05	1.000
Yes	Pmoc22	−0.59	0.06	−0.03	0.03	−0.94	0.00	0.346
No	Pmoc22	0.07	−0.07	0.03	0.03	0.99	0.00	0.321
Yes	Pmoc23	0.13	0.02	0.00	0.03	0.00	0.00	1.000
No	Pmoc23	0.30	−0.02	0.00	0.03	0.00	0.00	1.000
Yes	Pmoc24	0.30	0.02	0.00	0.03	0.00	0.00	1.000
No	Pmoc24	−0.23	−0.02	0.00	0.03	0.00	0.00	1.000
Yes	Pmoc25	−0.23	0.02	0.00	0.03	0.00	0.00	1.000
No	Pmoc25	0.11	−0.02	0.00	0.03	0.00	0.00	1.000
Yes	Pmoc26	−0.49	0.06	−0.03	0.03	−1.01	−0.07	0.313
No	Pmoc26	−0.43	−0.08	0.04	0.03	1.19	0.07	0.234

WPHACOTCCC, whether participants had attended any courses or training related to cross-cultural care; DIF Score is the difference between the observed and the expected average observations. DIF Measure is the difficulty of the item for the category, with all else held constant. DIF Size is the difference between the DIF measure for a certain category and the baseline difficulty. DIF SE is the standard error of the second DIF measure. DIF t is the DIF contrast divided by the SE of the two DIF measures. It is equivalent to the Mantel–Haenszel significance test but has the advantage of allowing missing data.

we found that item 10, "I have concerns about culturally competent care," had a low factor loading (0.44) (49) and poor fit indices (infit MNSQ: 3.33, outfit MNSQ: 3.54) (50). This may be due to differences in the understanding of the term "concerns" between Chinese and Western cultures, which suggests that the item might not effectively convey its intended meaning within Chinese nursing culture. In addition, the item may have been conceptually ambiguous, leading to misinterpretation by the respondents, which affected the scale's overall performance. After consulting with the experts, the research team decided to remove this item. Following its deletion, the C-BENEFITS-CCCSAT was re-evaluated using the CTT and Rasch models.

The CTT analysis demonstrated that the C-BENEFITS-CCCSAT exhibited strong reliability and validity overall. Cronbach's α was 0.80, with the dimension reliability values ranging from 0.700 to 0.905 and a split-half reliability value of 0.837. The test–retest reliability value was 0.881, and the S-CVI value was 0.928, with the I-CVI values ranging from 0.83 to 1.00. The CFA revealed a good model fit, with a CMIN/DF ratio of 1.071, a RMSEA value of 0.08, and the CFI, NFI, IFI, and TLI all exceeding 0.9. The factor loadings for all items were $\geq$ 0.50. The CR values ranged from 0.700 to 0.88, and the AVE values ranged from 0.420 to 0.55, meeting the criteria for convergent and discriminant validity. It is worth noting that the validity of the calibration correlations between the C-BENEFITS-CCCSAT and the C-SFCCSN was moderate. This may be attributed to differences in the dimensionality and theoretical frameworks of the two scales, despite both assessing intercultural caregiving competence. These differences likely contributed to the weaker correlations between certain dimensions. In addition, variations in the working environments of the sample may have influenced the correlation, thus limiting the comprehensiveness of the validity test. The Rasch analysis showed that the C-BENEFITS-CCCSAT had an item reliability of 1.00 and a person reliability of 0.76–0.89, both exceeding the 0.7 threshold, indicating strong reliability. The item separation index values ranged from 17.37 to 60.34, and the person separation index values ranged from 1.76 to 2.89, both exceeding the minimum criterion of 1.5, suggesting a good representation of both the items and persons. The infit MNSQ values for the items ranged from 0.87 to 1.08, and the outfit MNSQ values ranged from 0.86 to 1.27, meeting the model fit requirements. The PT-Measure Corr values ranged from 0.68 to 0.84, exceeding the minimum reference value of 0.5, showing that the items closely aligned with the measurement objectives and that the data fit the model well. The fit between individual ability and item difficulty was good, with the distribution of ability approximating a normal curve. The majority of the participants were clustered in the intermediate ability range, where most items were also located. The difficulty of the response categories followed the expected order, with neither the infit nor outfit statistics exceeding the acceptable fit range (0.6–1.4), indicating a good fit across the individuals, items, and categories. Furthermore, the DIF contrast value was below 0.5 logit in the subgroup with cross-cultural training experience, indicating no measurement bias across the groups. The combination of CTT and IRT not only validates the psychometric properties of the scale but also emphasizes the importance of assessing cross-cultural nursing competence in both nursing education and clinical practice.

The development of cross-cultural nursing competence is essential in nursing education (51). However, many nursing students still face gaps in cultural sensitivity and adaptation, which can impact their future clinical practice in multicultural settings. The

C-BENEFITS-CCCSAT serves as a standardized tool to help educators identify and address these gaps (7). Specifically, the scale can be used to assess students' cultural competence at different stages of the nursing curriculum. For example, at the start of the course, educators can use the scale to establish a baseline assessment of students' intercultural competence. If the results show low scores in dimensions such as "respect for cultural diversity," it indicates the need for additional instruction and training. Educators can then develop individualized teaching plans, such as case studies, role-playing, or scenarios (52). At the end of the course, the scale can be used again to assess student progress, allowing for data-driven adjustments to future course designs. This approach enhances the accuracy and effectiveness of nursing education, promoting the development of culturally sensitive nursing professionals.

In addition to identifying gaps in students' cultural competence, the scale can also serve as an assessment tool in nursing courses, particularly for formative and summative assessments. Educators can use the scale to evaluate students' cultural competence midway through a nursing program focused on intercultural nursing. If results show low scores in areas such as 'understanding the impact of cultural context on health beliefs,' educators can adjust the curriculum to include discussions on culturally relevant health beliefs or invite diverse cultural experts to share clinical experiences.

At the end of the program, educators can use the C-BENEFITS-CCCSAT to assess students' final performance in cultural competence. During graduation assessments, students can self-assess their cultural competence using the scale, while faculty members can evaluate students and incorporate the results into the overall teaching quality assessment. This ensures that nursing education focuses not only on theoretical knowledge but also on meeting the expected standards for cultural sensitivity and intercultural nursing practice (53). In addition, the scale can be integrated into clinical placement assessments. Clinical supervisors can use it to evaluate students' performance in communicating with culturally diverse patients and developing care plans during their placements. This allows for monitoring students' intercultural competence in real-world nursing environments and provides targeted feedback to help them improve their cultural nursing skills.

Finally, the development of cultural competence is crucial to ensure that future nurses can provide high-quality care in multicultural settings. Culturally sensitive nursing involves not only effective verbal communication but also a deep respect for patients' cultural beliefs, lifestyles, and health perspectives (54). Nurses must recognize cultural differences, understand patients' needs, and adapt their care accordingly (55). For example, in clinical practice, nurses may encounter language barriers, such as when a diabetic patient from a non-native-speaking country lacks knowledge about diabetes management and struggles to understand hospital health education materials. A culturally competent nurse would recognize the challenges posed by the language barrier and use translation services or more accessible methods to help the patient understand their condition (56). In addition, nurses may use visual teaching tools tailored to the patient's cultural background, ensuring the patient actively engages in managing their diabetes. By incorporating the scale into clinical placements, educators can assess students' communication skills and cultural sensitivity with patients from diverse cultural backgrounds, providing timely feedback. This not only enhances students' intercultural nursing competence but also offers an effective tool for assessing cultural competence in future nursing practice.

5 Conclusion

This study successfully validated the C-BENEFITS-CCCSAT using a rigorous methodology that integrated CTT and IRT. The CTT analysis demonstrated strong reliability and validity, while the Rasch analysis confirmed the scale's good fit and measurement precision. By combining these two approaches, we ensured the scale's robustness in assessing cultural competence, sensitivity, and skills. The C-BENEFITS-CCCSAT is a valuable tool for nursing educators to identify and address gaps in students' cross-cultural competence. It should be noted that despite the deletion of item 10, the five-factor structure of the Chinese version of the scale is consistent with the original scale, which is perfect. It can be applied in both educational settings to monitor progress and in clinical settings to evaluate students' ability to deliver culturally competent care. This approach provides a reliable framework for enhancing nursing education and improving care in multicultural environments. Future studies should test the scale in diverse populations to further establish its applicability and generalizability.

6 Limitations

The sample in this study was limited to Chinese nursing students, which might have affected the generalizability of the findings to other cultural or geographic contexts. Due to differences in the cultural and social backgrounds, the nursing students from different regions or countries might have had significant variations in their understanding and practice of cultural competence. Therefore, the applicability of the results may be somewhat limited. Future research could validate the C-BENEFITS-CCCSAT in different cultural and geographic contexts to assess its applicability and reliability globally, thus improving its generalizability. Furthermore, while this study provides a valuable cultural competence assessment tool for nursing education in China, given the diversity in global nursing education, the scale may require some adjustments to suit the cultural characteristics of other countries and regions. For example, some items may need to be localized to better align with the realities of different cultural groups. Therefore, future research should consider expanding the C-BENEFITS-CCCSAT to nursing students from diverse cultural backgrounds and explore the challenges and necessary modifications when using this tool in international nursing education. In conclusion, while this study provides a useful tool for assessing cross-cultural nursing competence, its limitations should be thoroughly discussed and addressed in future international research to ensure that the tool can be widely applied in nursing education and clinical practice across different cultural and geographic settings.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Jinzhou Medical University's Research Ethics Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Cruz JP, Alquwez N, Cruz CP, Felicilda-Reynaldo RFD, Vitorino LM, Islam SMS. Cultural competence among nursing students in Saudi Arabia: a cross-sectional study. *Int Nurs Rev.* (2017) 64:215–23. doi: 10.1111/inr.12370

2. Xi J. The exchange and mutual learning of civilizations is an important driving force for the advancement of human civilization and the peaceful development of the world. *Ideol Polit Work Res.* (2019)

3. Lan L, Qisheng G, Chenglin Z. Influence mechanism analysis of the spatial evolution of inter-provincial population flow in China based on epidemic prevention and control. *Popul Res Policy Rev.* (2023) 42:37. doi: 10.1007/s11113-023-09780-4

4. De Almeida GMF, Nascimento TF, Da Silva RPL, Bello MP, Fontes CMB. Theoretical reflections of Leininger's cross-cultural care in the context of COVID-19. *Revista Gaucha Enfermagem.* (2021) 42:e20200209. doi: 10.1590/1983-1447.2021.20200209

5. Gualda DM, Hoga LA. Leininger's transcultural theory. *Revista Escola Enfermagem U S P.* (1992) 26:75–86. doi: 10.1590/0080-6234199202600100075

6. Li Y, Wang Y, Luo Q, Wang A, Yu H. Application of the cross-cultural nursing theory in China. *Nurs Integrated Trad Chinese Western Med.* (2019) 5:222–24. doi: 10.11997/nitcwm.201902062

7. Yava A, Tosun B, Papp K, Tóthová V, Şahin E, Yılmaz EB, et al. Developing the better and effective nursing education for improving transcultural nursing skills cultural competence and cultural sensitivity assessment tool (BENEFITS-CCCSAT). *BMC Nurs.* (2023) 22:331. doi: 10.1186/s12912-023-01476-6

8. Chae D, Kim J, Kim S, Lee J, Park S. Effectiveness of cultural competence educational interventions on health professionals and patient outcomes: a systematic review. *Japan J Nurs Sci.* (2020) 17:e12326. doi: 10.1111/jjns.12326

9. Choi J-S, Kim J-S. Effects of cultural education and cultural experiences on the cultural competence among undergraduate nursing students. *Nurse Edu Prac.* (2018) 29:159–162. doi: 10.1016/j.nepr.2018.01.007

10. Ge Y. Study on cultural competence and cultural sensitivity of nursing undergraduate students [master's thesis]. Naval Medical University (2006).

11. Zhao Y, Zhong Q, Wang Q, Yu D. Research progress on intercultural sensitivity of nursing students. *Chinese Nurs Res.* (2021) 2:268–72. doi: 10.12102/j.issn.1009-6493.2021.02.013

12. Hu S-T, Chou P-C, Wu L-M, Feng J-Y, Yeh T-P. The public's perception of the nursing profession: validity and reliability of the Chinese version of the nursing image scale in Taiwan. *Nurs Health Sci.* (2024) 26:e13137. doi: 10.1111/nhs.13137

13. Ge Y, Zheng C, Wang X, Liu T. Psychometric properties of the Chinese version of the health behavior motivation scale: a translation and validation study. *Front Psychol.* (2024) 15:1279816. doi: 10.3389/fpsyg.2024.1279816

14. Chae D, Park Y. Development and cross-validation of the short form of the cultural competence scale for nurses. *Asian Nurs Res.* (2018) 12:69–76. doi: 10.1016/j.anr.2018.02.004

15. Zhu Y, Wang C, Dai M, Wang L. Translation of cultural competence scale for clinical nurses and its reliability and validity. *J Nursing (China).* (2024) 31:46-50. doi: 10.16460/j.issn1008-9969.2024.17.046

16. Li J, Yang Z, Qi R, Tan M, Ji X, Hou B, et al. Psychometric evaluation of the Chinese version of motivation for nursing student scale (MNSS): a quantitative and cross-sectional design. *Nurse Educ Pract.* (2023) 71:103690. doi: 10.1016/j.nepr.2023.103690

17. 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) 22:468. doi: 10.1186/s12912-023-01631-z

18. del Pozo-Herce P, Martínez-Sabater A, Chover-Sierra E, Gea-Caballero V, Satústegui-Dordá PJ, Saus-Ortega C, et al. Application of the Delphi method for content validity analysis of a questionnaire to determine the risk factors of the Chemsex. *Healthcare.* (2023) 11:2905. doi: 10.3390/healthcare11212905

19. Du X, Liu X, Zhao Y, Wang S. Psychometric testing of the 10-item perceived stress scale for Chinese nurses. *BMC Nurs.* (2023) 22:430. doi: 10.1186/s12912-023-01602-4

20. Verdú-Soriano J, González-de la Torre H. Rasch analysis implementation in nursing research: a methodological approach. *Enfermería Clínica.* (2024) 34:493–506. doi: 10.1016/j.enfcle.2024.11.009

21. Planinic M, Boone WJ, Susac A, Ivanjek L. Rasch analysis in physics education research: why measurement matters. *Physical Rev Physics Educ Res.* (2019) 15:020111. doi: 10.1103/PhysRevPhysEducRes.15.020111

22. George D, Mallery P. IBM SPSS statistics 26 step by step: A simple guide and reference. *16th* ed. New York: Routledge (2019).

23. Kim Y-M. Validation of psychometric research instruments: the case of information science. *J Am Soc Inf Sci Technol.* (2009) 60:1178–91. doi: 10.1002/asi.21066

24. 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 cross-sectional study. *Front Public Health.* (2024) 12:1349514. doi: 10.3389/fpubh.2024.1349514

25. 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 Psych.* (2021) 12:692498. doi: 10.3389/fpsyt.2021.692498

26. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. *Implement Sci.* (2017) 12:108. doi: 10.1186/s13012-017-0635-3

27. Yue M, Chen Q, Liu Y, Cheng R, Zeng D. Psychometric properties of the Chinese version of the nurses' attitudes towards communication with the patient scale among Chinese nurses. *BMC Nurs.* (2024) 23:779. doi: 10.1186/s12912-024-02415-9

28. Li W, Zhang Y, Liang J, Yu H. Psychometric evaluation of the Chinese version of the media health literacy questionnaire: a validation study. *Digit Health* (2023) 9:20552076231203801. doi: 10.1177/20552076231203801

29. Song R, Zhou R, Lin K, Wang W, Qin X. Psychometric validation of social network sites use motivation (SNSUM) scale among Chinese adolescents and undergraduates. *Acta Psychol.* (2024) 248:104435. doi: 10.1016/j.actpsy.2024.104435

30. Tabri N, Elliott CM. Principles and practice of structural equation modeling. *Canad Graduate J Sociol Criminol.* (2012) 1:59–60. doi: 10.15353/cgjsc.v1i1.3787

31. Labrague LJ, Arteche DL, Rosales RA, Santos MCL, Calimbas NDL, Yboa BC, et al. Development and psychometric testing of the clinical adjustment scale for student nurses (CAS-SN): a scale for assessing student nurses' adaptation in clinical settings. *Nurse Educ Today.* (2024) 142:106350. doi: 10.1016/j.nedt.2024.106350

32. Dong A, Zhang H, Kong L, Lu T, Zheng C, Ai F, et al. Chinese version of the physical resilience scale (PRS): reliability and validity test based on Rasch analysis. *BMC Public Health.* (2024) 24:2541. doi: 10.1186/s12889-024-19978-6

33. De Souza AC, Alexandre NMC, Guirardello EDB. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia Serviços Saúde.* (2017) 26:649–59. doi: 10.5123/S1679-49742017000300022

34. Hanis N, Ismail N, Abu Kassim NL, Idrus F. Validation of egalitarian education questionnaire using Rasch measurement model. *J Appl Meas.* (2020) 21:91–100.

35. van der Linden WJ, Hambleton RK, (eds.). Item response theory: brief history, common models, and extensions. In: *Handbook of modern item response theory.* New York, NY: Springer (1997), p. 1–28.

36. Luo P, Wan J, Bian W. Chinesization and Rasch model analysis of the nutritional self-efficacy questionnaire in elderly patients with chronic eye disease. *J Nurs Sci.* (2023) 38:1001–4152. doi: 10.3870/j.issn.1001-4152.2023.14.106

37. Wan S, Yang D, Liu N. Reliability and validity of the Chinese version of Amyotrophic lateral sclerosis impairment multidomain scale. *J Chongqing Med Univ.* (2024). doi: 10.13406/j.cnki.cyxb.003703

38. Matheny LM, Clanton TO. Rasch analysis of reliability and validity of scores from the foot and ankle ability measure (FAAM). *Foot Ankle Int.* (2020) 41:229–36. doi: 10.1177/1071100719884554

39. Cordier R, Speyer R, Schindler A, Michou E, Heijnen BJ, Baijens L, et al. Using Rasch analysis to evaluate the reliability and validity of the swallowing quality of life questionnaire: an item response theory approach. *Dysphagia.* (2018) 33:441–56. doi: 10.1007/s00455-017-9873-4

40. Boone WJ. Rasch analysis for instrument development: why, when, and how? *CBE Life Sci Educ.* (2016) 15:rm4. doi: 10.1187/cbe.16-04-0148

41. McLaughlin JE, Angelo TA, White PJ. Validating criteria for identifying core concepts using many-facet rasch measurement. *Front Educ.* (2023) 8:1150781. doi: 10.3389/feduc.2023.1150781

42. Qassem Alyami I. Psychometric analysis of childhood executive functioning inventory (CHEXI) in Saudi Arabian ADHD children: calibration with Rasch model. *Appl Neuropsychol Child.* (2024) 13:394–401. doi: 10.1080/21622965.2023.2208698

43. Lord FM. Applications of item response theory to practical testing problems. New York, NY: Routledge (2012).

44. Orlando Edelen MO, Thissen D, Teresi JA, Kleinman M, Ocepek-Welikson K. Identification of differential item functioning using item response theory and the likelihood-based model comparison approach. Application to the Mini-mental state examination. *Med Care.* (2006) 44:S134–42. doi: 10.1097/01.mlr.0000245251.83359.8c

45. van Roij J, Kieffer JM, van de Poll-Franse L, Husson O, Raijmakers NJH, Gelissen J. Assessing measurement invariance in the EORTC QLQ-C30. *Qual Life Res.* (2022) 31:889–901. doi: 10.1007/s11136-021-02961-8

46. Douglas MK, Pierce JU, Rosenkoetter M, Pacquiao D, Callister LC, Hattar-Pollara M, et al. Standards of practice for culturally competent nursing care: 2011 update. *J Transcult Nurs.* (2011) 22:317–33. doi: 10.1177/1043659611412965

47. McEvoy E, Henry S, Karkavandi MA, Donnelly J, Lyon M, Strobel N, et al. Culturally responsive, trauma-informed, continuity of care(r) toolkits: a scoping review. *Women Birth.* (2024) 37:101834. doi: 10.1016/j.wombi.2024.101834

48. Feng Y, Ou-Yang Z-Y, Lu J-Y, Yang Y-F, Zhang Q, Zhang M-M, et al. Cross-cultural adaptation and psychometric properties of the Mainland Chinese version of the manchester orofacial pain disability scale (MOPDS) among college students. *BMC Med Res Methodol.* (2023) 23:159. doi: 10.1186/s12874-023-01976-8

49. Ab Hamid MR, Sami W, Mohmad Sidek MH. Discriminant validity assessment: use of Fornell & Larcker criterion versus HTMT criterion. *J Phys Conf Ser*. (2017) 890:012163. doi: 10.1088/1742-6596/890/1/012163

50. Al-Qerem W, Jarab A, Al Bawab AQ, Eberhardt J, Alasmari F, Hammad A, et al. Validation of an Arabic tool for assessing vaccination literacy: a factor and Rasch analysis. *Hum Vaccin Immunother*. (2024) 20:2381297. doi: 10.1080/21645515.2024.2381297

51. Farokhzadian J, Nematollahi M, Dehghan Nayeri N, Faramarzpour M. Using a model to design, implement, and evaluate a training program for improving cultural competence among undergraduate nursing students: a mixed methods study. *BMC Nurs*. (2022) 21:85. doi: 10.1186/s12912-022-00849-7

52. Pires R, Marques M, Oliveira H, Goes M, Pedrosa M, Lopes M. Simulated practice in the development of clinical reasoning in nursing students: a systematic review protocol. *MethodsX*. (2025) 14:103144. doi: 10.1016/j.mex.2024.103144

53. Gehring DR, Titus SK, George R. The perceived concerns of nurse educators' use of GenAI in nursing education: protocol for a scoping review. *Health Sci Rep*. (2025) 8:e70411. doi: 10.1002/hsr2.70411

54. Brooks LA, Manias E, Bloomer MJ. Culturally sensitive communication in healthcare: a concept analysis. *Collegian*. (2019) 26:383–91. doi: 10.1016/j.colegn.2018.09.007

55. Claeys A, Berdai-Chaouni S, Tricas-Sauras S, De Donder L. Culturally sensitive care: definitions, perceptions, and practices of health care professionals. *J Transcult Nurs*. (2021) 32:484–92. doi: 10.1177/1043659620970625

56. Sommers CL, Bonnel WB. Nurse educators' perspectives on implementing culturally sensitive and inclusive nursing education. *J Nurs Educ*. (2020) 59:126–32. doi: 10.3928/01484834-20200220-02