Validation of the Barthel Index in Chinese nursing home residents: an item response theory analysis

Minyu Liang[1], Mei Yin[2], Bing Guo[2], Yichao Pan[3], Tong Zhong[4], Jieyi Wu[1] and Zengjie Ye[5]*

[1]School of Nursing, Guangzhou University of Chinese Medicine, Guangzhou, China, [2]Assisted Living Facility, Home For The Aged Guangzhou, Guangzhou, China, [3]Department of Vasculocardiology, Guangzhou First People's Hospital, Guangzhou, China, [4]Department of Tumor Radiotherapy, Zhuhai People's Hospital (Zhuhai Hospital Affiliated with Jinan University), Zhuhai, China, [5]School of Nursing, Guangzhou Medical University, Guangzhou, China

Background: The Barthel Index (BI) is used to standardize the grading of assessments for clinical needs, insurance support, and long-term care resource allocation in China. However, its psychometric properties among nursing home residents remain unclear. Therefore, this study aims to assess and modify the psychometric properties of BI in nursing home residents.

Methods: A total of 1,402 individuals undergoing evaluation in a nursing home facility in China were included in this study from November 2021 to November 2022. Correlations between items were examined to identify the potential multicollinearity concerns. The unidimensional item response theory (IRT) was used to validate and modify the single structure of BI. Furthermore, the logistic regression/IRT hybrid DIF detection method was conducted to assess differential item functioning (DIF) between the dementia group and the normal group.

Results: The pairing of items 5 ("bowl control") and 6 ("bladder control") revealed a local dependence issue, leading to their consolidation. Items 56 (bowel and bladder control) and 9 (mobility) both displayed poor fit indices and underwent category collapsing. Through the application of the generalized partial credit model, the adjusted scale displayed better fit indices, demonstrating a robust discriminative power (DC >1.5) and orderly thresholds. Furthermore, non-uniform DIF was identified in item 2 (bathing) between the dementia group and the normal group.

Conclusion: The modified BI demonstrated favorable psychometric properties and proved to be suitable for evaluating nursing home residents experiencing moderate functional impairment, which may provide a precise evaluation for long-term care resource allocation. Future studies could explore integrating supplementary measurements, such as objective indices, to assess a broader spectrum of functional statuses to potentially enhance the limited precision width observed in BI.

KEYWORDS

Barthel Index, psychometric properties, nursing home residents, item response theory, differential item functioning

Introduction

The Chinese population is on the brink of an aging era, and the percentage of individuals aged ≥65 years is projected to increase from 13% to 27.9% (180–380 million) between 2020 and 2050 (World Health Organization, 2018; China Development Research Foundation, 2020). The lifespan of individuals is correlated with low levels of physical activity and increased dependency. This eventually leads to a higher demand for long-term

care needs, which was estimated at 45.30 million in 2020 and is expected to increase by 39.0% to 59.32 million by 2030 (World Health Organization, 2015; Zapata-Lamana et al., 2021; Gong et al., 2022; Musa et al., 2022; Parra-Rizo et al., 2022). However, family-based non-professional care is becoming less feasible due to the China's historical one-child policy, the prevalence of empty nesters amid urbanization, and the widespread migration of the population (Qian et al., 2018). Consequently, care provided by nursing home staff is emerging as an indispensable alternative. Long-term care resource allocation (i.e., conditions of access to nursing home and clinical care needs) is identified by standardized grading assessments (Zhang et al., 2018), which typically centers on the assessment of activities of daily life (ADLs) (Fortinsky et al., 1999; Hébert et al., 1999; Holanda et al., 2022; Jeppestøl et al., 2022). The original 10-item Barthel Index (BI) has gained recognition and been integrated into the China's standardized grading assessment system for evaluating general ADL function in older adults due to its communicability, simplicity, and ease of scoring (Dickinson, 1992; Bouwstra et al., 2019; Medical Administration and Medical Authority, 2020). Studies have highlighted favorable psychometric properties of BI when used for patients with stroke, Parkinson's disease, and older adults (Pashmdarfard and Azad, 2020). However, one study indicated that the reliability of BI fluctuated when it was used for evaluating severe disability (Sainsbury et al., 2005). Given that nursing home residents, typically above 70 years of age, often exhibit serious disability or dependence (Muszalik et al., 2021; Kashiwagi and Morioka, 2022; Zhao et al., 2022), questions arise regarding the suitability of BI for their evaluation. A previous study using BI-10 to assess 644 patients with dementia across 19 long-term care facilities in Thailand, Japan, and South Korea revealed compromised item fit, including item bias, redundancy, and narrow threshold widths, casting doubt on the suitability of BI for assessing dementia and indicating a need for modification (Yi et al., 2020). However, to date, no study has been conducted to validate the use of the 10-item BI for nursing home residents in mainland China. Therefore, it is imperative to promptly validate the psychometric properties of BI for assessing nursing home residents in mainland China.

The classical test theory (CTT) ha been widely used to assess the psychometric properties of scales, focusing on assessing construct validity and internal consistency. However, it overlooks measurement precision, leading to an incomplete understanding of scale's psychometric properties (Hunsley and Mash, 2008). In contrast, the item response theory (IRT) has emerged as a more comprehensive framework, offering deeper insights into internal consistency, factor structure, and measurement precision. This theory provides a precise psychometric framework that delineates interactions between individuals and items to determine if items effectively measure the intended population. Additionally, precision is typically demonstrated through item information curves or test information, highlighting the optimal discrimination location of an item or a scale across individual latent traits (Reckase, 2009). Moreover, IRT enables the detection of item- and test-level biases via differential item functioning (DIF) analysis, allowing for adjustments to mitigate or eliminate these biases (Teresi et al., 2012). Moreover, IRT assessed item functions, such as the discrimination ability and difficulty of each item, without relying on the sample, a capability that CTT lacks (Reckase, 2009;

Teresi et al., 2012). Hence, compared to CTT, IRT offers more comprehensive information (e.g., item function, measurement precision) for instrument development and validation.

Therefore, the objective of this study is to validate the psychometric properties of the 10-item BI, tailoring their use to suit the specific context of nursing homes. The current study hypothesizes that the 10-item BI may have compromised psychometric properties for evaluating nursing home residents in mainland China and a modified version of the BI may serve as a suitable instrument.

Methods

Data and samples

The current study involved a secondary analysis of data obtained from an electronic evaluation system in a nursing home, encompassing 1568 residents enrolled consecutively from November 2021 to November 2022. The inclusion criteria were as follows: (1) individuals aged ≥60 years; (2) residents in a nursing home facility; (3) nursing home residents who had undergone evaluation. The exclusion criteria included participants with incomplete demographic or instrument data. As a result, 166 residents were excluded: 125 (8.4%) were not evaluated, and 41 (2.7%) had missing data. Finally, the study comprised 1,402 nursing home residents. Ethics approval was obtained from the ethics committee of the First Affiliated Hospital of Guangzhou University of Chinese Medicine (K-2023-046).

Measures

Demographic evaluation

The demographic evaluation included several factors, such as sex, age, education, comorbidities, marital status, cognitive status, and the BI level grading.

Barthel Index

Mahoney and Barthel introduced the BI in 1965, comprising 10 items designed to assess activities feeding, bathing, grooming, dressing, using toilet, transferring (moving from the bed to the chair and back), mobility (on level surfaces), climbing stairs, and controlling bowel and bladder functions (Honey and Barthel, 1965). Among these, two items (items 2 and 3) feature dichotomous response categories, while six items (items 1, 4, 5, 6, 7, and 10) and two items (items 8 and 9) employ three- and four-graded response options, respectively. Total scores range from 0 to 100, with higher scores indicating greater levels of functional independence. In the current study, Cronbach's alpha was recorded at 0.95.

Statistical analysis

First, a descriptive statistic method was employed to summarize the demographic characteristics of nursing home residents and to delineate the distribution patterns within item response

TABLE 1 Demographic characteristics of nursing home residents (n = 1,402).

Characteristics		N (%)
Sex	Female	881 (62.8)
	Male	521 (37.2)
Age (years)	<80	544 (38.8)
	>=80	858 (61.2)
Education level	Middle school	916 (65.3)
	High school or technical secondary school	293 (20.9)
	College degree or above	193 (13.8)
Comorbidities	0	86 (6.1)
	1	708 (50.5)
	≥2	608 (43.4)
Cognitive status	Dementia disease	227 (16.2)
	No-dementia disease	1,175 (83.8)
Marital status	Having spouse	580 (41.4)
	No spouse	822 (58.6)
BI grading	0	197 (14.1)
	1	639 (45.6)
	2	130 (9.3)
	3	436 (31.0)

categories. Second, the item correlation analysis and item residual correlation matrix were performed for potential item consolidation. Specifically, pairs of items exhibiting correlation coefficients (r1) > 0.95 indicated multicollinearity concerns (Tleyjeh et al., 2008). The residual correlation values exceeding 0.1 were defined as contravening the assumption of local independence (Reeve et al., 2007; Kline, 2011). Items showing signs of multicollinearity or failing to meet the criteria for local independence were considered for item consolidation. Third, IRT was performed to analyze the unidimensional structure of the scale based on previous research (Wang et al., 2020). A unidimensional generalized partial credit model (GPCM) was selected to analyze polytomous item responses. Its equation is as follows (Muraki, 1992):

$$P(X_{ikj} = k|\theta_j) = \frac{\exp \sum_{h=0}^{k} [a_i (\theta_j - b_{ih})]}{\sum_{c=0}^{m_i} \exp \sum_{h=0}^{c} [a_i(\theta - b_{ih})]}$$

X_{ikj} refers to the person (j)'s response in category k of item i. The probability (P) of X_{ikj} is defined as a logistic probability function and is determined by the slope or discrimination parameter vector (a_i), item-category threshold parameter vector (b_{ih}), and person latent trait parameter vector (θ_j).

Adequate fit was defined by root mean square error of approximation (RMSEA) < 0.1, Tucker–Lewis index (TLI) >0.90, comparative fit index (CFI) >0.90, standardized root mean square residual (SRMSR) <0.08, and low values of Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and sample-adjusted BIC (SABIC) (Hu and Bentler, 1999; Byrne, 2006;

Hooper et al., 2008). Additionally, Pearson's χ2 (S-X^2) was used to identify misfit items and to determine the need for collapsing categories (Chalmers and Ng, 2017; Robinson et al., 2019).

The two crucial item parameters, namely discrimination and threshold/difficulty, were assessed based on the following criteria: discrimination was considered poor if it was <0.5, moderate if it was between 0.5 and 1.0, good if it was within the range of 1.0–1.5, and excellent if it was >1.5. Threshold/difficulty was deemed good if it satisfied the requirement of monotonicity of item response (Reckase and McKinley, 1991). Subsequently, measurement accuracy was determined by the information function concerning BI. Finally, item- and test-level biases between the dementia group and the normal group were identified by DIFs using logistic regression/the IRT hybrid DIF method (Choi et al., 2011). The statistical process involved the use of R, Mplus, and SPSS.

Results

Demographic characteristics

Most residents (83.8%) did not have a diagnosis of dementia. Among them, 62.8% were female individuals, and the mean age was 81.32 years (standard deviation = 9.0). A vast majority (93.9%) of them had multiple comorbidities. Further demographic details are presented in Table 1.

Item distributions, correlations, and local independence

Figure 1A summarizes the item distributions. The correlation analysis revealed that the correlation coefficients (r1) among each item ranged from 0.514 to 0.928 (Figure 1B). No item exhibited multicollinearity. However, it is noteworthy that the residual correlation coefficient between item 5 (bowel control) and item 6 (bladder control) surpassed 0.1, indicating a violation of local independence (Figure 1C). Consequently, items 5 and 6 were combined into a new item (bowel and bladder control). The modified BI comprised nine items including feeding, bathing, grooming, dressing, bowel and bladder control, toilet usage, transferring, mobility, and stair climbing. The sample was randomly divided into two equal groups, namely the developmental sample (N = 701) and the validation sample (N = 701).

Confirmation of the unidimensional structure

In the developmental sample, the unidimensional structure GPCM met the good model fit criteria, with CFI > 0.95, TLI > 0.95, RMSEA < 0.1, and SRMSR < 0.08 (Figure 2). All items exhibited discrimination values >1.5, indicating robust discrimination among individuals across varying functional levels (Figure 2). Additionally, the thresholds displayed an orderly trend, escalating as the categories increased. Notably, item 56 (bowel and
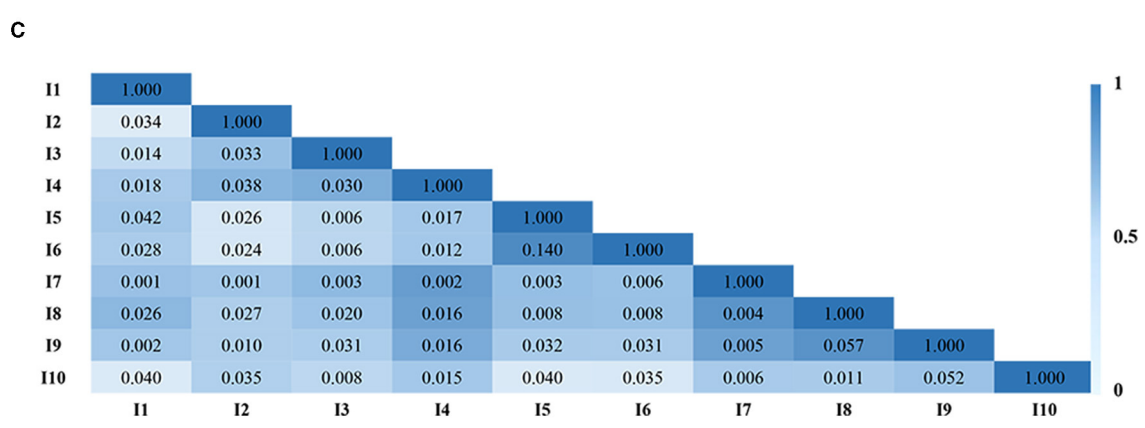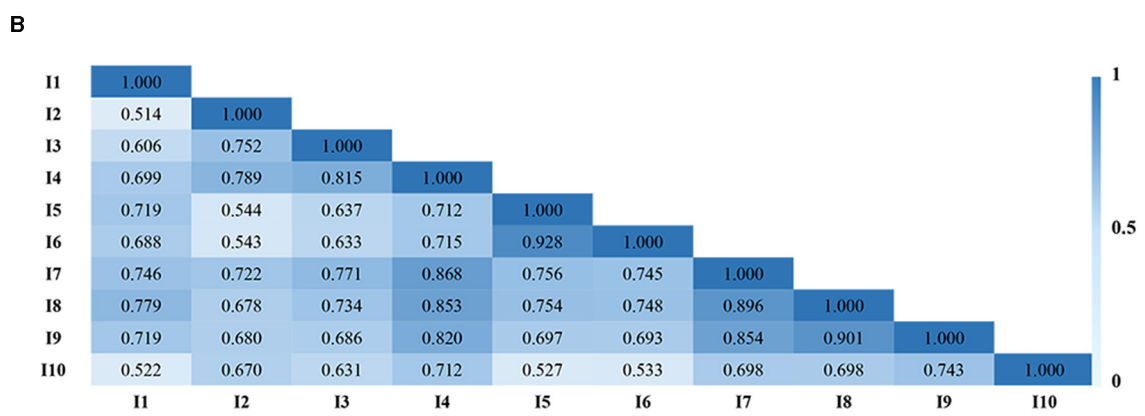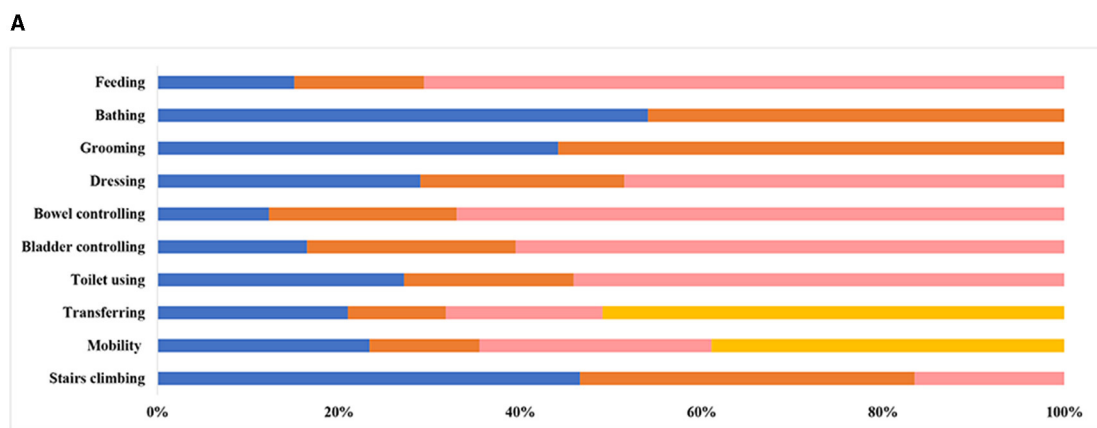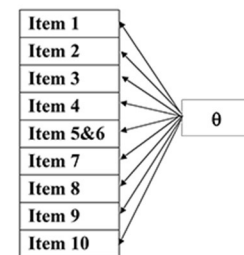
FIGURE 1
Item distribution **(A)**, item correlation **(B)**, and local independence **(C)**. **(B, C)** A deeper color suggested a stronger association. Its value < 0.10 indicated local independence satisfaction.

bladder control) and item 9 (mobility) were considered misfit (p < 0.05), suggesting the need for collapsing categories. Thus, the adjected categories with similar meanings were combined. For item 56, the adjacent rating categories of "occasional incontinence" and "complete incontinence" were integrated into a single category termed "incontinence". Consequently, item 56 was restructured into two categories, namely "incontinence" and "voluntary". As for item 9, the adjacent rating categories "wheelchair independent" and "walking with help" were combined into a single category named "some help from others or devices". Thus, item 9 encompassed categories of "dependent", "some help from others or devices", and "independent".

FIGURE 2
IRT analysis (development sample, *n* = 701).

Validation of the Barthel Index

In the validation sample, the unidimensional structure GPCM was employed to validate the modified BI. The residual correlation coefficients for item pairs were below 0.10, indicating compliance with the local independence hypothesis (Figure 3A). While most model indices indicated a good fit (CFI=0.98, TLI=0.97, and SRMSR=0.03), the RMSEA index exceeded 0.1. Additionally, the model displayed a robust discriminative power (DC >1.5), order thresholds, and no instances of item misfit (Figure 3B). Threshold widths for the items ranged from 0.42 to 1.16 logits. Item 10 (stair climbing) emerged as most challenging, followed by item 2 (bathing), while item 1 (feeding) was the easiest, followed by item 8 (transferring). Items 3 (grooming), 4 (dressing), 56 (bowel and bladder control), 7 (toilet usage), and 9 (mobility) demonstrated closely located difficulties (DF = −0.35–0.06). Illustrations of item characteristic curves (ICCs), test information, and test standard errors are presented in Figure 4. The item information covered latent trait (θ_j) levels from −3 to 3, with the majority concentrated between −1 and 1. This indicates that the modified BI offered optimal measurement accuracy for respondents with moderate functional impairment, rather than those at minimum or maximum function trait levels.

Differential item functioning

The dementia group and the normal group revealed a minimal distinction in the total expected score as depicted by the test characteristic curves (TCCs) of all items (both items with and without DIF) and DIF items (Figure 5A). Individual functional levels displayed no statistically significant difference between the dementia and normal groups (Figure 5B). Monte Carlo simulation-based thresholds (Figure 5C) identified non-uniform DIF in items

1 (feeding), 2 (bathing), 3 (grooming), and 4 (dressing) based on a chi-squared value ($p < 0.0011$). However, the differences in items 1, 3, and 4 were deemed negligible, given the proportionate β1 change (<5%) and R2 change (<0.02). Only item 2 (bathing) exhibited non-uniform DIF ($p < 0.0001$, Δ%B1 = 11.90%, R2 change > 0.02).

Discussion

The present study employed the IRT to validate the psychometric properties of the BI for nursing home residents, aiming to gather more detailed item information compared to the CTT (Reckase, 2009). This study represents the first large-sample examination of BI's psychometric properties among nursing home residents in mainland China. The findings confirmed our hypothesis that the 10-item BI showed unsatisfactory psychometric properties (violation of local independence and item misfit). While the modified BI demonstrated optimal psychometric properties: the unidimensional structure indicated a good model fit, the items possessed robust discriminative power, the thresholds displayed an orderly trend, and no item was misfit. Thus, the modified BI may present as the better instrument for evaluating nursing home residents. It retained nine items but still assessed ten aspects of ADLs, as some items and categories were consolidated, instead of being discarded, to achieve the optimal psychometric properties and preserve the maximum information from the original scale. Moreover, the modified BI showed high measurement accuracy for residents with moderate functional impairment rather than those with minimum or maximum function levels.

While the fit indices in the validation samples displayed inconsistencies regarding model fitting (CFI and TLI >0.95 indicated a good model fit, while RMSEA >0.1 suggested a poor fit), the one-factor GPCM remained the optimal model. A study conducted by Keke et al. suggested that strong correlations among
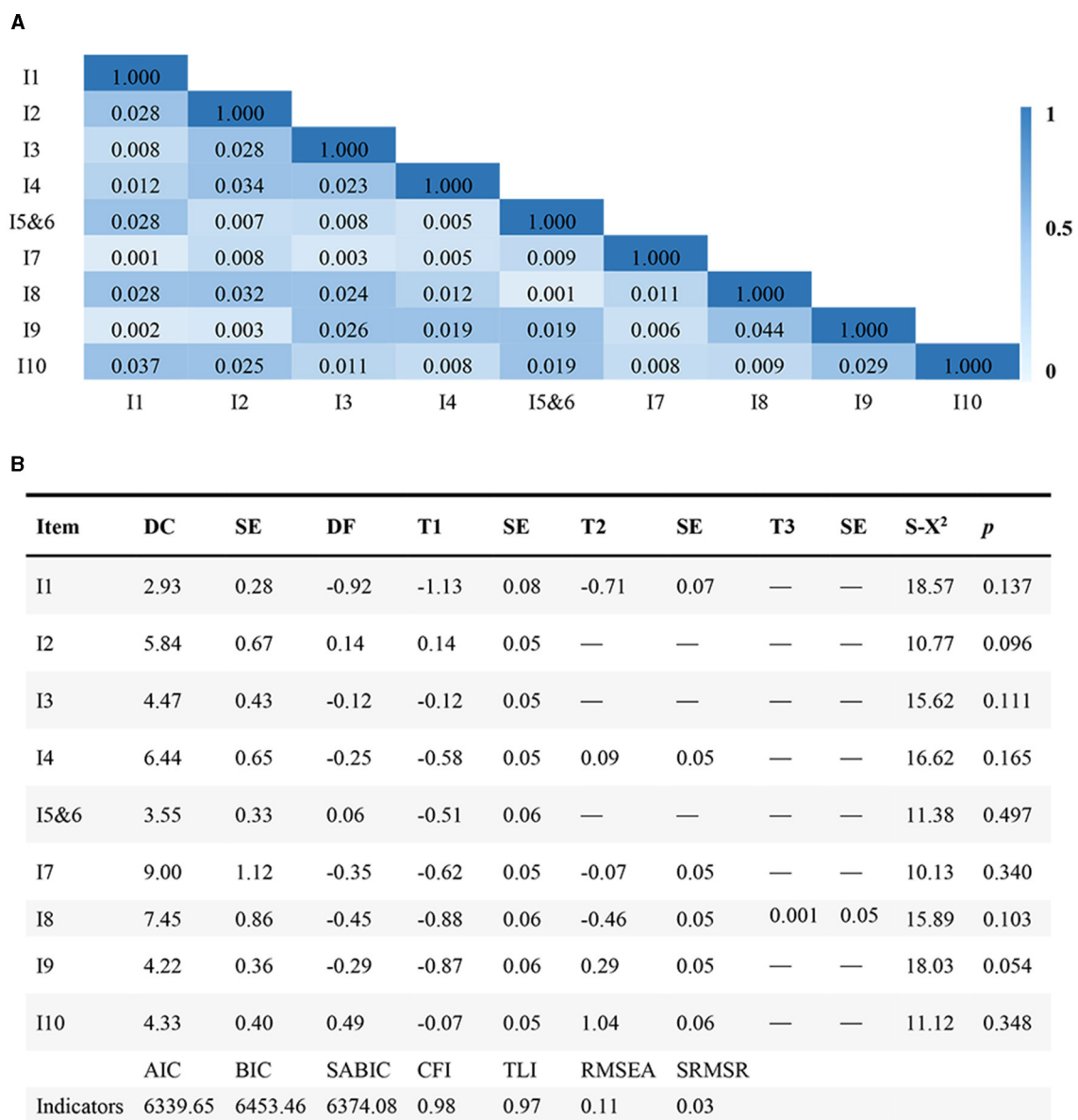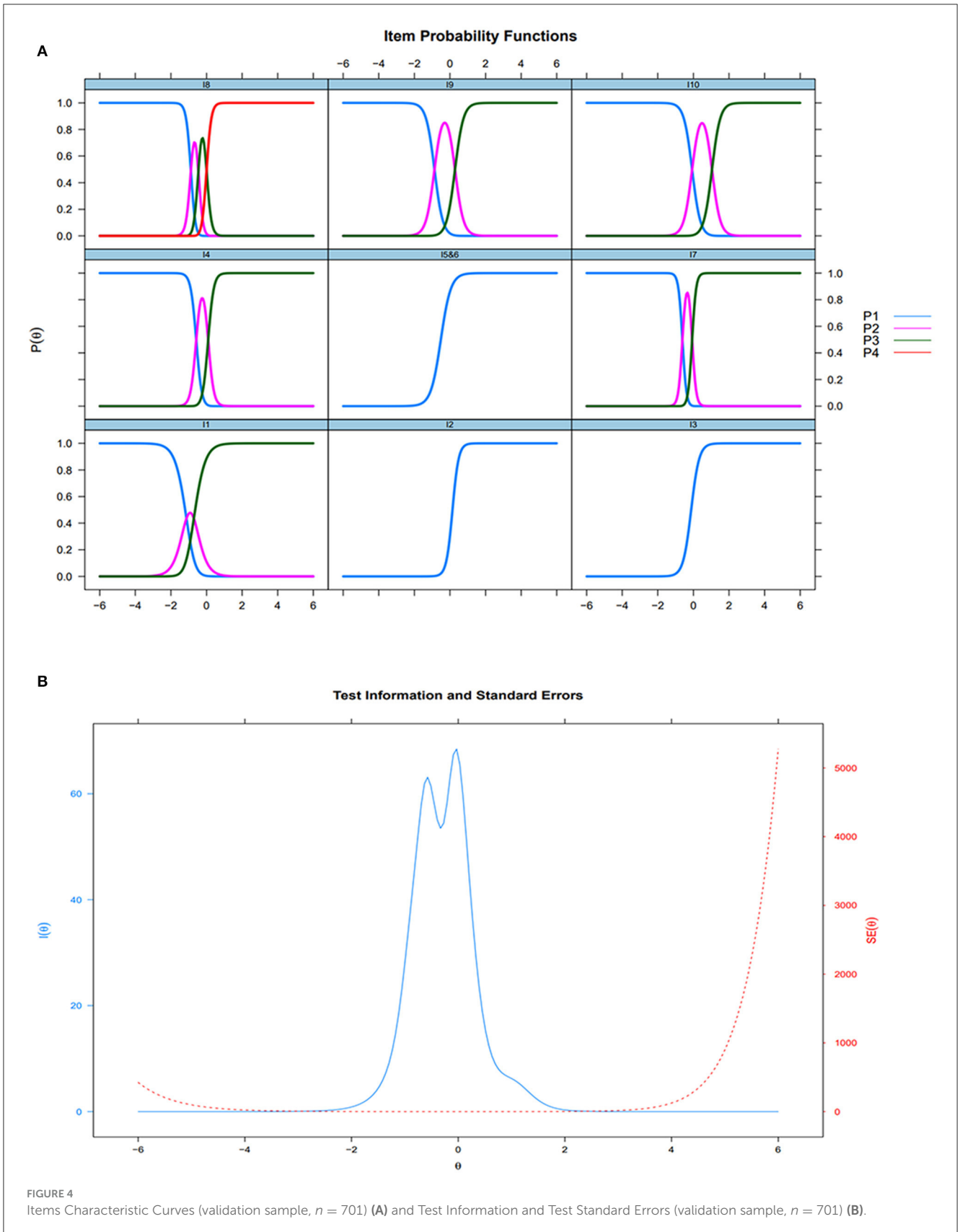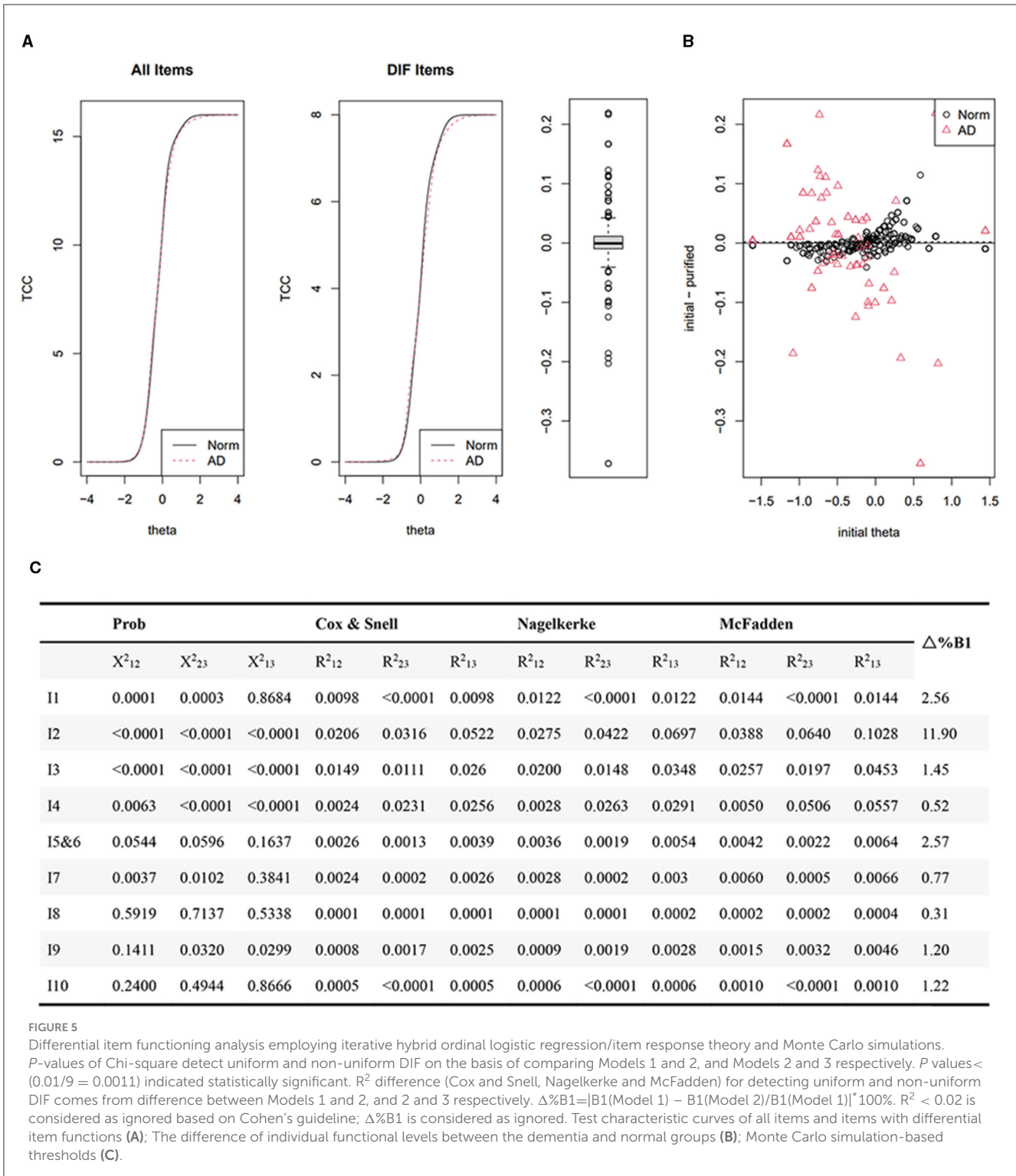
FIGURE 3
Local independence of the validation sample ($n = 701$) **(A)** and IRT analysis of the validation sample **(B)**. **(B)** A deeper color suggested a stronger item-pair residuals association. Its value < 0.10 indicated local independence satisfaction.

observed variables might yield a "good" CFI but a "bad" RMSEA, emphasizing the importance of balance among various fit indices (Lai and Green, 2016). In this study, item pairs exhibited strong correlations ($r1 = 0.51–0.93$), contributing to the unfavorable RMSEA performance. Although making potential modifications to the model, such as excluding highly correlated item pairs, might yield a "better" RMSEA, it would result in a loss of valuable item information. Therefore, despite the compromised RMSEA due to substantial item correlations, the one-factor GPCM remained the optimal model in this study.

Ensuring adherence to the local independence hypothesis is essential for accurate parameter calculation. In this study, item pairs (items 5 and 6) violated the local independence hypothesis (>0.1) and were combined into a new item (bowel and bladder control). This finding aligns with previous research indicating that the mean thresholds for "controlling bladder" and "controlling bowel" were nearly identical, suggesting potential redundancy between these items (Yi et al., 2020). Furthermore, physiological mechanisms suggest that bladder and bowel contraction are controlled by the pelvic musculature, potentially leading to concurrent urinary and bowel dysfunction (Hubscher et al., 2021; Kajbafzadeh et al., 2021). Among nursing home residents, who are typically aged over 81 years and have decreased contraction ability of the pelvic musculature, there exists a likelihood of concomitant urinary and bowel dysfunction. Hence, consolidating "bladder control" and "bowel control" into a single item is warranted.

FIGURE 4
Items Characteristic Curves (validation sample, $n = 701$) (A) and Test Information and Test Standard Errors (validation sample, $n = 701$) (B).

FIGURE 5
Differential item functioning analysis employing iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *P*-values of Chi-square detect uniform and non-uniform DIF on the basis of comparing Models 1 and 2, and Models 2 and 3 respectively. *P* values < (0.01/9 = 0.0011) indicated statistically significant. R^2 difference (Cox and Snell, Nagelkerke and McFadden) for detecting uniform and non-uniform DIF comes from difference between Models 1 and 2, and 2 and 3 respectively. △%B1=|B1(Model 1) − B1(Model 2)/B1(Model 1)|*100%. $R^2 < 0.02$ is considered as ignored based on Cohen's guideline; △%B1 is considered as ignored. Test characteristic curves of all items and items with differential item functions **(A)**; The difference of individual functional levels between the dementia and normal groups **(B)**; Monte Carlo simulation-based thresholds **(C)**.

Regarding item fit, items 5, 6, and 9 displayed misfit. This misfit can stem from inappropriate item definitions, poor category definitions, or an excessive number of choices, leading to discordant participant responses (Andrich and Luo, 2003). Common strategies to address this issue involve eliminating misfit items or collapsing categories. Collapsing categories is the primary option to preserve more information. Following the collapsing of categories, item 56 comprises two categories "incontinence"

and "voluntary". This action was corroborated by prior research using the BI for assessing long-term care residents, which indicated the limited utility of the second category ("occasional incontinence") within items related to bowel and bladder control for discriminating latent traits among individuals (Liu et al., 2015). In nursing home settings, residents experiencing complete incontinence typically use incontinence products and are assessed under the "complete incontinence" category. Conversely, those

with intact sphincter control manage themselves and are assessed as having "voluntary" bowel and bladder control. Occasional incontinence poses a challenge as it often goes unnoticed until caregivers discover wet bed sheets (Sayabalian et al., 2019). Such instances might be classified as complete incontinence due to heavy caregiving burdens and workforce shortages in nursing homes, or they might go undetected, leading to categorization as voluntary bowel and bladder control. Consequently, the category of occasional incontinence becomes ambiguous. Thus, "incontinence" (consolidating categories of occasional and complete incontinence) and "voluntary" emerge as more suitable options for item 56 (bowel and bladder control) in nursing home assessments. Similarly, item 9 (mobility) exhibited misfit. The adjacent categories of "wheelchair independent" and "walking with help" might be ambiguous as many nursing home residents prefer wheelchairs despite being capable of walking short distances. Collapsing these two categories into a new category ("some help from others or device") rendered item 9 compatible with the GPCM. This aligns with previous studies employing the Rasch analysis, which also suggested the potential ambiguity of the "mobility" item's second (wheelchair independent) and third (walking with help) categories leading to model misfit (van Hartingsveld et al., 2006; Yi et al., 2020). Therefore, collapsing categories aligns with the GPCM analysis and practicality within nursing home settings.

In terms of item functions, all nine items exhibited excellent discrimination, implying their robust ability to distinguish between varying functional levels. No disorderly thresholds were observed, indicating the suitability of the categories within the BI. Moreover, the maximum information was concentrated within the latent trait (θ_j) ranging from −1 to 1, while minimal information function was detected at latent trait (θ_j) values near −3 or 3. This indicates that the BI possesses a narrow assessment precision width, effectively discriminating among nursing home residents with moderate functional impairments but lacking the capacity to differentiate residents with minimal functional limitations or severe functional impairments. Similarly, the 15-item BI also demonstrated a narrow assessment precise width ($-1 < \theta < 1$) (Zhao et al., 2022). Furthermore, reports of ceiling effects (>15% of the sample attaining the maximum score) and floor effects (>15% of the sample scoring the minimum) have been reported across different populations, such as patients with stroke (Balu, 2009), patients in the intensive care unit (Reis et al., 2021), and older hospitalized patients with cognitive spectrum disorders (Braun et al., 2021). These observations suggest an inherent limitation in the BI's measurement due to its focus on a limited range of ADLs and the absence of assessment beyond basic ADLs (Bouwstra et al., 2019). Consequently, this restricted range of individual locations within the BI might lead to a narrow precision width. The findings of the current study align with the observation that the BI comprises ADL items encompassing individual locations ranging from−1 to 1, offering precise measurement specifically for individuals experiencing moderate functional impairment. In our study, items such as "grooming", "dressing", "bowel and bladder control", "toilet usage", and "mobility" exhibited closely aligned difficulty levels, indicating potential redundancy among these items, prompting consideration for their removal. However, when a 5-item BI was developed from the original 10-item BI by eliminating redundant items, a significant loss of item information

occurred, leading to unsatisfactory recommended reliability criteria (Hobart and Thompson, 2001). Hence, retaining these items with similar difficulty levels is essential to maintain the integrity of the scale structure and to preserve item information.

Regarding measurement bias, prior reports have not addressed DIF between the dementia and normal groups concerning the BI. DIF highlights a potential bias in measurement for each item. In this study, only non-uniform DIF was identified in item 2 (bathing). This suggests a potential disparity in bathing abilities between the dementia and normal groups, a trend supported by previous studies indicating a higher dependency on bathing among individuals with dementia (Liu et al., 2015). Despite this, minimal differences surfaced between the dementia and normal groups in the total expected score across all items, including those with DIF. This implies that any distortion in the scale due to the DIF item is likely insignificant.

Implications for research and clinical practice

First, examining item difficulty through the GPCM could elucidate the sequence of ADL decline among nursing home residents, furnishing valuable insights into potential functional impairments. For instance, item 10 (stair climbing) emerged as most challenging, followed by item 2 (bathing), while item 1 (feeding) posed the least difficulty, followed by item 8 (transferring). This pattern suggests that stair climbing and bathing constitute the initial loss of ADLs while feeding and transferring represent the final stages of decline in ADLs. These findings are consistent with those of previous studies indicating that older individuals find "climbing stairs" and "bathing" most challenging due to higher oxygen consumption (Bauer et al., 2022), whereas "feeding" and "transferring" rank among the easiest ADLs (Liu et al., 2015). The inability to perform the simplest ADLs (such as feeding and transferring) indicates severe functional impairment, while difficulty with the most challenging tasks (such as stair climbing or bathing) suggests milder functional impairment, offering insights for tailored individual care based on the function level. Second, despite the widespread clinical use of the BI, there remains a lack of comprehensive information on its psychometric properties for nursing home residents. This study, benefitting from a large sample size, effectively used IRT, resulting in more precise and reliable outcomes. Using the GPCM analysis and drawing from nursing home practices, modifications were made to the BI, particularly combining items 5 and 6, and collapsing categories in items 56 and 9 to modify BI. These adaptations aim to align the BI with the nursing home environment while preserving maximum information. Moreover, the diverse pool of eligible participants in nursing homes, spanning a wide spectrum of functional statuses without specific diagnostic restrictions or limited comorbidities, ensures that the study outcomes are broadly applicable across populations. Third, while the BI provides precise measurements for individuals with moderate functional impairment, it also provides limited information for individuals with mild or severe functional impairments. This finding bears significance for long-term care insurance policies, which often rely on BI assessments.

Supplementing the BI with additional measurements, such as objective indices targeting mild or severe functional impairments, should be integrated into the long-term care assessment system. Finally, while the non-uniformity of item 2 did not significantly affect the overall scale performance, it highlights the necessity of further exploring minimal functional differences between the dementia and normal groups. A previous study demonstrated that employing a semi-structured interview focusing on ADL contents could facilitate the recognition of minimal impairments, enhancing precision in assessing daily functioning, including coordinated actions, proper execution, and completion levels (Cornelis et al., 2017). Crafting evaluation protocols tailored for assessing dementia-related groups could enhance the detection of minimal functional changes.

Limitations

Certain limitations warrant consideration in this study. First, although the nurse administering the BI underwent prior training, the specific characteristics of these nurses were neither recorded nor accounted for, potentially introducing rater bias. Second, the narrow precision width of the BI is another limitation, resulting in less precise evaluations at the minimum and maximum levels of ADLs. Future research endeavors could integrate the BI with objective assessment tools to complement these limitations. Third, the sample size of individuals with dementia comprised <20% of the total sample. Further studies with larger dementia cohorts are necessary to validate DIF between the dementia and normal groups. Finally, among the 1,402 residents included, 166 were excluded for various reasons. This exclusion could potentially introduce selection bias due to unknown differences in their functional levels.

Conclusion

The modified BI demonstrated favorable psychometric properties and proved to be suitable for evaluating nursing home residents experiencing moderate functional impairment, which may provide precise evaluation for long-term care resource allocation. Future studies could explore integrating supplementary measurements, such as objective indices, to assess a broader spectrum of functional statuses to potentially enhance the limited precision width observed in the BI.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving humans were approved by the First Affiliated Hospital of Guangzhou University of Chinese Medicine Ethics Committee. 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

ML: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Writing – original draft. MY: Conceptualization, Data curation, Investigation, Project administration, Resources, Writing – review & editing. BG: Conceptualization, Data curation, Investigation, Project administration, Resources, Writing – review & editing. YP: Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. TZ: Conceptualization, Data curation, Writing – review & editing. JW: Conceptualization, Data curation, Writing – review & editing. ZY: Conceptualization, Supervision, Validation, Writing – review & editing.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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

Andrich, D., and Luo, G. (2003). Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. *J. Appl. Meas.* 4, 205–221.

Balu, S. (2009). Differences in psychometric properties, cut-off scores, and outcomes between the Barthel Index and Modified Rankin Scale in pharmacotherapy-based stroke trials: systematic literature review. *Curr. Med. Res. Opin.* 25, 1329–1341. doi: 10.1185/03007990902875877

Bauer, S. R., Cawthon, P. M., Ensrud, K. E., Suskind, A. M., Newman, J. C., Fink, H. A., et al. (2022). Osteoporotic fractures in men (MrOS) research group. Lower urinary tract symptoms and incident functional limitations among older community-dwelling men. *J. Am. Geriatr. Soc.* 70, 1082–1094. doi: 10.1111/jgs.17633

Bouwstra, H., Smit, E. B., Wattel, E. M., van der Wouden, J. C., Hertogh, C. M. P. M., Terluin, B., et al. (2019). Measurement properties of the barthel index in geriatric rehabilitation. *J. Am. Med. Dir. Assoc.* 20, 420–425. doi: 10.1016/j.jamda.2018.09.033

Braun, T., Thiel, C., Schulz, R. J., and Grüneberg, C. (2021). Responsiveness and interpretability of commonly used outcome assessments of mobility capacity in older hospital patients with cognitive spectrum disorders. *Health Qual. Life Outcomes.* 19, 68. doi: 10.1186/s12955-021-01690-3

Byrne, B. M. (2006). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming.* Mahwah: Lawrence Erlbaum Associates.

Chalmers, R. P., and Ng, V. (2017). Plausible-value imputation statistics for detecting item misfit. *Appl. Psychol. Meas.* 41, 372–387. doi: 10.1177/0146621617692079

China Development Research Foundation (2020). *China Development Report 2020: Trends and Policies in China's Aging Population.* Available online at: https://cdrf-en.cdrf.org.cn/jjhdt/5478.htm (accessed June 11, 2020).

Choi, S. W., Gibbons, L. E., and Crane, P. K. (2011). lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo Simulations. *J. Stat. Softw.* 39, 1–30. doi: 10.18637/jss.v039.i08

Cornelis, E., Gorus, E., Beyer, I., Bautmans, I., and De Vriendt, P. (2017). Early diagnosis of mild cognitive impairment and mild dementia through basic and instrumental activities of daily living: Development of a new evaluation tool. *PLoS Med.* 14, e1002250. doi: 10.1371/journal.pmed.1002250

Dickinson, E. J. (1992). Standard assessment scales for elderly people. Recommendations of the Royal College of Physicians of London and the British Geriatrics Society. *J. Epidemiol. Community Health.* 46, 628–629. doi: 10.1136/jech.46.6.628

Fortinsky, R. H., Covinsky, K. E., Palmer, R. M., and Landefeld, C. S. (1999). Effects of functional status changes before and during hospitalization on nursing home admission of older adults. *J. Gerontol. A Biol. Sci. Med. Sci.* 54, M521–526. doi: 10.1093/gerona/54.10.M521

Gong, J., Wang, G., Wang, Y., Chen, X., Chen, Y., Meng, Q., et al. (2022). Nowcasting and forecasting the care needs of the older population in China: analysis of data from the China Health and Retirement Longitudinal Study (CHARLS). *Lancet Public Health.* 7, e1005–e1013. doi: 10.1016/S2468-2667(22)00203-1

Hébert, R., Brayne, C., and Spiegelhalter, D. (1999). Factors associated with functional decline and improvement in a very elderly community-dwelling population. *Am. J. Epidemiol.* 150, 501–510. doi: 10.1093/oxfordjournals.aje.a010039

Hobart, J. C., and Thompson, A. J. (2001). The five item Barthel index. *J. Neurol. Neurosurg. Psychiatr.* 71, 225–230. doi: 10.1136/jnnp.71.2.225

Holanda, C. M. A., Nóbrega, P. V. N., and Maciel, Á. C. C. (2022). Physical performance as a predictor of mortality in nursing home residents: A five-year survival analysis. *Geriatr. Nurs.* 47,151–158. doi: 10.1016/j.gerinurse.2022.07.002

Honey, F. I., and Barthel, D. W. (1965). Functional evaluation: the barthel index. *Md 415 State Med J.* 14, 61–65.

Hooper, D., Coughlan, J., and Mullen, M. R. (2008). Structural equation modeling: guidelines for determining model fit. *Electron J. Bus Res. Methods.* 6, 53–60. doi: 10.21427/D7CF7R

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modeling.* 6, 1–55. doi: 10.1080/10705519909540118

Hubscher, C. H., Wyles, J., Gallahar, A., Johnson, K., Willhite, A., Harkema, S. J., et al. (2021). Effect of different forms of activity-based recovery training on bladder, bowel, and sexual function after spinal cord injury. *Arch. Phys. Med. Rehabil.* 102, 865–873. doi: 10.1016/j.apmr.2020.11.002

Hunsley, J., and Mash, E. J. (2008). *A Guide to Assessments that Work.* Oxford: Oxford University.

Jeppestøl, K., Kirkevold, M., and Bragstad, L. K. (2022). Assessing acute functional decline in older patients in home nursing care settings using the Modified Early

Warning Score: a qualitative study of nurses' and general practitioners' experiences. *Int. J. Older People Nurs.* 17, e12416. doi: 10.1111/opn.12416

Kajbafzadeh, A. M., Ahmadi, H., Montaser-Kouhsari, L., Sabetkish, S., Ladi-Seyedian, S., and Sotoudeh, M. (2021). Intravesical electromotive administration of botulinum toxin type A in improving the bladder and bowel functions: Evidence for novel mechanism of action. *J. Spinal Cord Med.* 44, 89–95. doi: 10.1080/10790268.2019.1603490

Kashiwagi, M., and Morioka, N. (2022). Determinants associated with the incidence of occupational accidents among visiting nurses from home-visit nursing agencies: secondary analysis of cross-national survey data in Japan. *Geriatr. Gerontol. Int.* 22, 588–596. doi: 10.1111/ggi.14420

Kline, R. (2011). *Principles and Practice of Structural Equation Modeling.* Guilford: Guilford press.

Lai, K., and Green, S. B. (2016). The problem with having two watches: assessment of fit when RMSEA and CFI disagree. *Multivariate Behav. Res.* 51, 220–239. doi: 10.1080/00273171.2015.1134306

Liu, W., Unick, J., Galik, E., and Resnick, B. (2015). Barthel Index of activities of daily living: item response theory 417 analysis of ratings for long-term care residents. *Nurs. Res.* 64, 88–99. doi: 10.1097/NNR.0000000000000072

Medical Administration and Medical Authority (2020). *Notice on the Implementation of Needs Assessment and Standardization of Services for Aged Care.* Available online at: http://www.nhc.gov.cn/yzygj/s7653/201908/426ace6022b747ceba12fd7f0384e3e0.shtml (accessed Jan 20, 2020).

Muraki, E. A. (1992). generalized partial credit model: application of an EM algorithm. *Appl. Psychol. Meas.* 16, 159–176. doi: 10.1177/014662169201600206

Musa, M. K., Akdur, G., Brand, S., Killett, A., Spilsbury, K., Peryer, G., et al. (2022). The uptake and use of a minimum data set (MDS) for older people living and dying in care homes: a realist review. *BMC Geriatr.* 22, 33. doi: 10.1186/s12877-021-02705-w

Muszalik, M., Kotarba, A., Borowiak, E., Puto, G., Cybulski, M., and Kędziora-Kornatowska, K. (2021). Socio-demographic, clinical and psychological profile of frailty patients living in the home environment and nursing homes: a cross-sectional study. *Front. Psychiatry.* 12:736804. doi: 10.3389/fpsyt.2021.736804

Parra-Rizo, M. A., Vásquez-Gómez, J., Álvarez, C., Diaz-Martínez X, Troncoso, C., Leiva-Ordoñez, A. M., Zapata-Lamana, R., et al. (2022). Predictors of the level of physical activity in physically active older people. *Behav. Sci. (Basel).* 12, 331. doi: 10.3390/bs12090331

Pashmdarfard, M., and Azad, A. (2020). Assessment tools to evaluate activities of daily living (ADL) and instrumental activities of daily living (IADL) in older adults: a systematic review. *Med. J. Islam. Repub. Iran.* 34, 33. doi: 10.47176/mjiri.34.33

Qian, Y., Qin, W., Zhou, C., Ge, D., Zhang, L., and Sun, L. (2018). Utilisation willingness for institutional care by the elderly: a comparative study of empty nesters and non-empty nesters in Shandong, China. *BMJ Open.* 8, e022324. doi: 10.1136/bmjopen-2018-022324

Reckase, M. D. (2009). *Multidimensional Item Response Theory.* New York: Springer.

Reckase, M. D., and McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Appl. Psychol. Meas.*15. doi: 10.1177/014662169101500407

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med. Care.* 45, S22–S31. doi: 10.1097/01.mlr.0000250483.85507.04

Reis, N. F. D., Biscaro, R. R. M., Figueiredo, F. C. X. S., Lunardelli, E. C. B., and Silva, R. M. D. (2021). Early Rehabilitation Index: translation and cross-cultural adaptation to Brazilian Portuguese; and Early Rehabilitation Barthel Index: validation for use in the intensive care unit. *Rev Bras Ter Intensiva.* 33, 353–361. doi: 10.5935/0103-507X.20210051

Robinson, M., Johnson, A. M., Walton, D. M., and MacDermid, J. C. (2019). A comparison of the polytomous Rasch analysis output of RUMM2030 and R (ltm/eRm/TAM/lordif). *BMC Med. Res. Methodol.* 19, 36. doi: 10.1186/s12874-019-0680-5

Sainsbury, A., Seebass, G., Bansal, A., and Young, J. B. (2005). Reliability of the Barthel Index when used with older people. *Age Ageing.* 34, 228–232. doi: 10.1093/ageing/afi063

Sayabalian, A., Easton-Garrett, S., Kassabian, A., and Kunze, M. B. (2019). The impact of a resident's occasional incontinence: the under-recognized incontinence. *Geriatr. Nurs.* 40,648–650. doi: 10.1016/j.gerinurse.2019.11.006

Teresi, J. A., Ramirez, M., Jones, R. N., Choi, S., and Crane, P. K. (2012). Modifying measures based on differential item functioning (DIF) impact analyses. *J. Aging Health.* 24,1044–1076. doi: 10.1177/0898264312436877

Tleyjeh, I. M., Steckelberg, J. M., Georgescu, G., Ghomrawi, H. M., Hoskin, T. L., Enders, F. B., et al. (2008). The association between the timing of valve surgery and 6-month mortality in left-sided infective endocarditis. *Heart*. 94, 892–896. doi: 10.1136/hrt.2007.118968

van Hartingsveld, F., Lucas, C., Kwakkel, G., and Lindeboom, R. (2006). Improved interpretation of stroke trial results using empirical Barthel item weights. *Stroke*. 37, 162–166. doi: 10.1161/01.STR.0000195176.50830.b6

Wang, S. H., Shi, J. J., Sun, Y., Liu, J. M., Chen, Y. J., and Ma, Y. (2020). Reliability and validity of the simplified version Modified Barthel Index in convalescence period of stroke. *Chin. J. Rehabilitat*. 35, 179–182. doi: 10.3870/zgkf.2020.04.003

World Health Organization (2015). *World Report on Ageing and Health*. Available online at: https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf?sequence=1 (accessed April 1, 2024).

World Health Organization (2018). Ageing-and Health. Available online at: https://www.ncl.ac.uk/who-we-are/strengths/ageing-health/ (accessed April 1, 2024).

Yi, Y., Ding, L., Wen, H., Wu, J., Makimoto, K., and Liao, X. (2020). Is barthel index suitable for assessing activities of daily living in patients with dementia. *Front. Psychiatry*. 11:282. doi: 10.3389/fpsyt.2020.00282

Zapata-Lamana, R., Poblete-Valderrama, F., Cigarroa, I., and Parra-Rizo, M. A. (2021). The practice of vigorous physical activity is related to a higher educational level and income in older women. *Int. J. Environ. Res. Public Health*. 18, 10815. doi: 10.3390/ijerph182010815

Zhang, Y., Wang, Q., Jiang, T., and Wang, J. (2018). Equity and efficiency of primary health care resource allocation in mainland China. *Int. J. Equity Health*. 17, 140. doi: 10.1186/s12939-018-0851-8

Zhao, M., Wang, Y., Wang, S., Yang, Y., Li, M., and Wang, K. (2022). Association between depression severity and physical function among Chinese nursing home residents: the mediating role of different types of leisure activities. *Int. J. Environ. Res. Public Health*. 19, 3543. doi: 10.3390/ijerph19063543