



Kent Academic Repository

Yu, Siyue, Xu, Richard Huan, Wang, Kailu, Sze, Judy Chu Dik, Cheung, Annie Wai Ling, Welch, Elizabeth, Smith, Nick and Wong, Eliza Lai Yi (2026) *Psychometric properties of the Chinese version of the ASCOT-SCT4 in Hong Kong: a validation study using classical test theory and Rasch analysis*. Health and Quality of Life Outcomes . ISSN 1477-7525.

Downloaded from

<https://kar.kent.ac.uk/116910/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1186/s12955-026-02640-7>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Psychometric properties of the Chinese version of the ASCOT-SCT4 in Hong Kong: a validation study using classical test theory and Rasch analysis

Siyue Yu · Richard Huan Xu · Kailu Wang · Judy Chu Dik Sze · Annie Wai Ling Cheung · Elizabeth Welch · Nick Smith · Eliza Lai Yi Wong

Received: 14 May 2026 / Accepted: 18 September 2026
© The Author(s) 2026

Abstract

Background Hong Kong's rapid population ageing increases demand for long-term care (LTC) and places pressure on public expenditure. The Adult Social Care Outcomes Toolkit Four-level Self-Completion Tool (ASCOT-SCT4) is a preference-based instrument specifically developed to measure social care-related quality of life (SCRQoL); however, evidence regarding the construct validity and reliability of its Chinese version in Hong Kong is lacking. This study evaluates these psychometric properties among older community care users using both Classical Test Theory (CTT) and Rasch analysis.

Methods A cross-sectional survey was conducted with 312 older adults aged ≥ 60 years receiving community care and support services; 119 participants completed a follow-up retest 1–4 weeks later. CTT analyses were conducted to assess internal consistency, factor structure, construct validity via hypothesis testing, and test-retest reliability. Rasch analysis was employed to evaluate item fit, response category ordering, person-item targeting, reliability (person and item), and differential item functioning (DIF) across age and gender groups.

Results The CTT analyses showed acceptable internal consistency (Cronbach's $\alpha=0.79$), good test-retest reliability (intraclass correlation coefficient=0.809), and sufficient construct validity ($>75\%$ of the hypotheses were confirmed). A one-factor structure was supported; however, the *Dignity* item had a low factor loading (0.269). The Rasch results indicated satisfactory item fit (Infit/Outfit statistics within 0.50–1.50), ordered response thresholds, and acceptable reliability (person: 0.721, item: 0.964), with separation indices of 1.606 (person) and 5.182 (item). However, a measurement gap was observed for respondents with higher (better) levels of social care-related quality of life (SCRQoL). No evidence of DIF was detected.

Conclusions The Chinese version of the ASCOT-SCT4 is a valid and reliable instrument for use among older community care users in Hong Kong. Nevertheless, the suboptimal performance of the *Dignity* item and limited measurement coverage at higher SCRQoL levels warrant further validation and potential refinement.

Keywords Long-term care · Quality of life · Preference-based measure · ASCOT · Psychometrics · Older population

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



Background

Hong Kong (HK) faces significant challenges from its rapidly ageing population, characterized by one of the highest life expectancies [1] and lowest fertility rates [2] in the world. By 2031, an estimated 28% of the population will be aged 65 or older [3], propelling the city into an “ultra-aged society” [4]. Ageing increases risks of chronic disease, functional limitations [5, 6], and psychosocial issues [7, 8], driving a surge in demand for long-term care (LTC) services. In HK, more than 80% of LTC service costs are covered by government subsidies [9], and public expenditure on LTC more than doubled between 2011 and 2022 [10]. Nevertheless, demand continues to outpace service provision, as evidenced by lengthy waiting lists for both community (approximately 5–15 months [11]) and residential care (approximately 6–22 months [12]). This disparity between limited capacity and vast unmet needs strains public finances and threatens the sustainability of the LTC system [9]. Consequently, evaluating the “value for money” of LTC services is increasingly critical to support evidence-based decision-making and optimize resource allocation.

In economic evaluations, both costs and outcomes are needed to estimate the value for money of services or interventions [13]. Outcomes are commonly measured using preference-based instruments, most often those measuring health-related quality of life (HRQoL), such as the EQ-5D [14, 15]. These instruments are health-focused and are suited to healthcare settings, where the primary aim is often to promote, restore, or maintain individuals’ health status [16]. However, LTC differs from healthcare in its primary focus: rather than directly targeting health gains, it aims to compensate for functional limitations by providing support with personal care, household activities, and social engagement, such as bathing, home cleaning, shopping, and participation in social activities [17, 18]. Although health outcomes remain relevant in LTC [19], many of its benefits may be reflected in non-health outcomes, such as maintaining independence, dignity, and social interaction [18], all of which are important to older adults’ well-being [19]. It has therefore been argued that the use of health-related preference-based instruments may underestimate LTC benefits, leading to inefficient resource allocation [14, 20–22]. Preference-based measures of well-being, such as the ICEpop CAPability measure for Older people (ICECAP-O) [23], offer a broader alternative. However, well-being is not a goal unique to LTC: it is a common goal of most public services, which may make it difficult to attribute observed differences specifically to LTC.

In response, the Adult Social Care Outcomes Toolkit (ASCOT) was developed to measure the impact of LTC (also known as social care in the UK) on quality of life (QoL), referred to as social care-related QoL (SCRQoL) [17, 24]. The ASCOT-SCT4, its four-level self-completion instrument for care receivers, is preference-based and intended both for the routine monitoring of LTC service outcomes—for example, in England’s annual Adult Social Care Survey—and for use in economic evaluation, for which it is recommended by the National Institute for Health and Care Excellence (NICE) [24, 25]. The instrument captures outcomes across eight domains: *Control over daily life*, *Personal cleanliness*, *Food and drink*, *Personal safety*, *Social participation*, *Occupation*, *Home cleanliness*, and *Dignity*. Each domain comprises four ordered response levels. Drawing on Sen’s capability approach [26], the highest level (‘ideal state’) reflects individuals’ ability and choice to achieve their desired state, whereas the remaining three levels represent functioning states defined by the extent to which care needs are met, ranging from ‘no needs’ through ‘some needs’ to ‘high needs’ [17]. The *Dignity* domain is an exception: its ‘ideal state’ reflects functioning rather than capability, as this domain captures the psychological impact of the way care and support are provided on individuals’ sense of self [17]. The domains of the ASCOT-SCT4 focus on aspects of QoL that LTC services are intended to influence, and therefore the instrument may be better able to reflect the effects of LTC services. Consistent with this, an observational study in England found that the estimated impact of home care service use was statistically significant when assessed using the earlier three-level self-completion version of ASCOT, but not when assessed using the EQ-5D [20]. Similarly, factor analysis in a previous study comparing ASCOT with the ICECAP-O showed that although the two instruments overlapped in their measurement of well-being, ASCOT captured some LTC-specific aspects of QoL that were not reflected in the ICECAP-O [27].

Despite growing interest in outcome measurement for LTC in HK, existing generic preference-based instruments available locally, including the EQ-5D [28], Short-Form Six-Dimension (SF-6D) [29], and Health Utilities Index Mark 2/3 (HUI2/3) [30], are measures of HRQoL. A validated preference-based instrument specifically designed to capture outcomes of LTC remains lacking. Our team previously translated the ASCOT-SCT4 into Traditional Chinese¹ (hereafter referred to as the ‘Chinese ASCOT-SCT4’) and established its content validity with local LTC users and experts [31]. However, other psychometric properties, particularly construct validity and reliability, remain unknown in the HK context. Consequently, it remains unclear whether the Chinese ASCOT-SCT4 captures the underlying theoretical construct of SCRQoL as intended and yields consistent and reproducible responses among local LTC users. Although the ASCOT-SCT4 has been validated in several countries including the UK [32, 33], Finland [34], Austria [35], and Japan [36], its psychometric performance can vary across cultural settings due to differences in language, norms, or construct understanding [37–39], making local validation essential. Therefore, the objective of this study is to evaluate the psychometric properties, including construct validity and reliability, of the descriptive system of the Chinese ASCOT-SCT4 among a sample of community care and support users in HK, using both CTT and Rasch analysis.

Methods

Study design and data collection

A cross-sectional survey was conducted between June 2023 and April 2024 among older adults receiving publicly funded community care and support services in HK. Eligibility criteria included (1) being aged 60 years or older²; (2) currently receiving publicly funded community care and/or support services³; (3) speaking Cantonese⁴; and (4) being able to provide written informed consent. Exclusion criteria included (1) being unable to read Chinese or understand Cantonese; (2) being unable to provide responses by self-report; and (3) having significant cognitive impairment that could affect the validity of the data.

A convenience sample of participants was recruited through collaborating elderly community centres and non-governmental organizations across the three main regions (i.e., HK Island, New Territories, and Kowloon) in HK. Participants who consented to follow-up contact were randomly selected for a second survey to assess test-retest reliability, conducted 1–4 weeks after the initial assessment. Prior to the retest survey, participants were asked about any changes in their living conditions (e.g., health status, QoL, or diseases) since the initial assessment. The retest was conducted only with those who reported no changes.

Trained interviewers administered the computer-assisted questionnaire in person, via Zoom or by telephone. Interviews took place at community centres, participants’ homes, or the research office. Participants completed a structured questionnaire that included the Chinese ASCOT-SCT4, eight validated measures, including the EQ-5D 5-level (EQ-5D-5 L) [28], EQ-Visual Analogue Scale (EQ-VAS) [28], 9-item Patient Health Questionnaire (PHQ-9) [40], the Control, Autonomy, Self-realization, and Pleasure-12 (CASP-12) [41], self-rated life satisfaction scale [42], Barthel Index [43], Lawton instrumental ADL (IADL) scale [44], and UCLA 3-item loneliness scale [45] (see details in Table S1), socio-demographic questions (e.g., age, gender, education, living arrangements), information on care service utilization (types of services used, self-rated satisfaction with care services

¹ Traditional Chinese is the predominant written form of Chinese in HK. The translated version was phrased in standard written Chinese.

² This age criterion was chosen to align with the eligibility requirements for publicly funded LTC services in HK, which are targeted at individuals aged 60 and above.

³ Users of publicly funded services were selected because LTC provision in HK is predominantly publicly funded [9] and recruitment was conducted through collaborating centres and organizations that mainly provided such services.

⁴ Cantonese is the most commonly used spoken language in HK.

on a 5-point Likert scale, ranging from 1 = “very dissatisfied” to 5 = “very satisfied”), and an item assessing how often they felt respected by care staff, with responses provided on a 3-point scale (“none of the time”, “most of the time”, and “all of the time”).

Measurement

The ASCOT-SCT4 comprises eight domains, each with four response levels, as described above. Table S2 presents the descriptive system, comprising the wording of each item and its corresponding response options, as well as definitions of each domain and response level. Each domain is represented by a single item. Although there are nine items in the instrument, the *Dignity filter* item (Q8) is used to help respondents express their feelings about needing help before answering the *Dignity* item (Q9); therefore, the filter item was not scored or included in psychometric evaluations [17].

In the absence of a HK-specific utility index, a level summary score was calculated by assigning numeric values to each response level (i.e., 0 for ‘high needs’ and 3 for ‘ideal state’) and summing these values across all eight scored items, producing a total score ranging from 0 to 24. Higher scores represented better SCRQoL [17].

Data analyses

Descriptive statistics were computed as appropriate to summarize participants’ socio-demographic characteristics, their responses to the Chinese ASCOT-SCT4, and the retest interval. Each analysis was based on available cases for the variables involved.

CTT analysis

- (a) Internal consistency: Cronbach’s alpha was computed for the overall scale and for the scale with each item removed. Values ≥ 0.70 were considered acceptable, and ≥ 0.90 indicated good internal consistency. Corrected item-total correlations were used to assess how well each item aligned with the overall construct, with values < 0.20 suggesting potential removal [46–48].
- (b) Factor structure: Given potential differences in factor structure across cultural contexts, exploratory factor analysis (EFA) has been recommended for examining the structural characteristics of translated and cross-culturally adapted versions of ASCOT [49]. EFA was conducted on a polychoric correlation matrix using principal axis factoring with oblimin (oblique) rotation to evaluate the dimensionality of the instrument [50, 51]. Data suitability was confirmed by Bartlett’s test of sphericity ($p < 0.05$) [51] and the Kaiser-Meyer-Olkin (KMO) test (≥ 0.7 was considered acceptable) [52]. Factors were extracted based on the Kaiser criterion (eigenvalues > 1), scree plot inspection, and parallel analysis [51]. A factor loading of ≥ 0.30 was used as the threshold to determine meaningful item contributions to a factor, supporting a clean factor structure [50].
- (c) Construct validity: In the absence of a gold standard, construct validity was assessed via hypothesis testing against validated instruments measuring related constructs.
 - Convergent validity: Spearman rank correlation coefficients (r) were used to assess the strength and direction of the associations between the Chinese ASCOT-SCT4 summary score and the life satisfaction scale, CASP-12, PHQ-9, EQ-5D-5 L, and EQ-VAS. The life satisfaction scale and CASP-12 were used as measures of well-being, a construct closely related to SCRQoL [32, 53]. PHQ-9 assesses depressive symptoms, which are indicative of poorer mental health and lower QoL [27, 54, 55]. EQ-5D-5 L and EQ-VAS were used as health measures to assess HRQoL and health status, respectively. These health-focused constructs are conceptually distinct from, but theoretically related to, SCRQoL because health status can influence care needs, while unmet care needs can in turn affect health status [17, 33, 35, 56]. Moderate-to-strong positive

correlations were expected with the life satisfaction scale, CASP-12, EQ-5D-5 L, and EQ-VAS, whereas a moderate-to-strong negative correlation was expected with PHQ-9. Associations with well-being measures were expected to be stronger than those with health measures [56]. Correlations were interpreted as negligible (< 0.10), weak ($0.10-0.29$), moderate ($0.30-0.49$), strong ($0.50-0.70$), and very strong (> 0.70) [57].

- **Known-group validity:** The Kruskal-Wallis test examined the Chinese ASCOT-SCT4's ability (using the summary score) to differentiate between groups expected to differ based on depressive symptoms (PHQ-9), functional dependency ((I)ADL limitations), and satisfaction with care services [33, 35]. Depression was defined as a PHQ-9 score > 9 , based on a cut-off validated in HK [58, 59]. (I)ADL groups were defined as those without limitations versus those with one or more limitations [33]. Satisfaction with care services was grouped into satisfied ('very satisfied' or 'somewhat satisfied'), neutral, or dissatisfied ('somewhat dissatisfied' or 'very dissatisfied') categories [33]. A p -value < 0.05 was considered statistically significant; epsilon-squared (ϵ^2) effect sizes were calculated and interpreted as negligible (< 0.01), small ($0.01 \leq \epsilon^2 < 0.06$), moderate ($0.06 \leq \epsilon^2 < 0.14$), or large (≥ 0.14) [60].
 - **Item-level construct validity:** Higher (better) levels on individual ASCOT-SCT4 items were expected to correspond with higher scores of life satisfaction, CASP-12, EQ-5D-5 L, and EQ-VAS, as well as lower scores of depressive symptoms (PHQ-9). To examine whether the 'ideal state' reflected the concept of capability, significant differences in CASP-12 Control & Autonomy subscale (CASP-12-control) scores were expected between the 'ideal state' and 'no needs' response options [32]. Additionally, significant differences were expected for specific items based on categorical variables including perceived loneliness (measured by UCLA 3-item Loneliness scale), (I)ADL limitations (measured by Barthel index and Lawton IADL scale), problems with self-care, or usual activity (measured by EQ-5D-5 L), living arrangements, depression (measured by PHQ-9), satisfaction with care services, and perception of being respected by care staff, with the latter two variables serving as indicators of perceived quality of care [32, 34, 35, 61]. The Kruskal-Wallis test was employed for comparisons between the continuous scores of comparator instruments and the categorical response levels of individual ASCOT-SCT4 items [62]. Post-hoc analyses for multiple group comparisons were conducted using the Dunn's test [63]. Fisher's exact test was used for categorical comparisons where ASCOT-SCT4 item levels were compared against other categorical grouping variables, as the 'high needs' response option in most items contained low frequencies [64]. A p -value < 0.05 was considered statistically significant for all the tests.
 - **Hypotheses confirmation:** The hypotheses were formulated based on evidence from previous studies [17, 27, 32–35, 56, 61] and are detailed in Table S3. Overall construct validity at the scale and item levels were considered supported if at least 75% of the hypotheses were confirmed [65, 66].
- (d) **Test-retest reliability:** Stability over time was assessed for the summary score of the Chinese ASCOT-SCT4 using the intraclass correlation coefficient (ICC) based on a two-way mixed-effects, absolute-agreement model (interpretations: < 0.5 poor, $0.5-0.75$ moderate, $0.75-0.9$ good, > 0.9 excellent) [67]. Item-level reliability was assessed using quadratic weighted kappa coefficients (< 0.4 poor, $0.4-0.75$ moderate, > 0.75 excellent) [68, 69]. As a sensitivity analysis, Spearman's rank correlation was used to examine the associations between the retest interval in days and the absolute test-retest differences in the ASCOT-SCT4 summary score and individual item response categories [70, 71].

Rasch analysis

Given the polytomous and ordinal nature of the Chinese ASCOT-SCT4 responses, where response options have distinct wordings potentially reflecting uneven intervals, the Partial Credit Model (PCM) was employed, which is also in line with an earlier ASCOT study [72]. Item parameters were estimated using conditional maximum likelihood with the R package "eRm" [73].

- (a) Assumptions: Confirmatory Factor Analysis (CFA) was conducted to examine the unidimensionality (using the R package “lavaan” [74]). Model fit was considered acceptable if Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) > 0.95 , Root Mean Square Error of Approximation (RMSEA) < 0.06 , and Standardized Root Mean Square Residual (SRMR) < 0.08 [75]. Local Independence was examined by assessing the residual correlation matrix derived from a one-factor EFA using the principal axis factoring method [76, 77], where absolute residual correlations of item pairs > 0.30 suggested potential local dependence [78, 79].
- (b) Item thresholds: The PCM yielded three thresholds per item, representing values on the latent trait continuum where respondents are equally likely to select between adjacent categories [80, 81]. The Item Category Characteristic Curves (ICCCs) were visually inspected to assess whether each category possessed a specific (unique) range of latent trait levels, and whether the thresholds between adjacent categories were ordered correctly. Overlapping or disordered curves indicated potential redundancy of response options or that the response categories did not function as intended [82].
- (c) Item fit: Infit and Outfit mean square statistics were used to indicate item fit, with values between 0.5 and 1.5 as acceptable [83]. Values close to 1.0 indicate good item fit; values below 0.5 suggest that responses are overpredicted, while values above 1.5 suggest that responses are underpredicted [84, 85].
- (d) Targeting: A person-item map was used to visualize the alignment between the distribution of participant abilities and item threshold locations on the same latent trait continuum in logit units [80]. A well-targeted scale should include items that span the full range of person abilities, and the mean person ability closely aligns with the average scale difficulty (conventionally set to zero logits) [86–90]. A difference between these means larger than 1 logit suggests poor targeting [91–93].
- (e) Reliability: Person and item separation reliability coefficients > 0.70 were considered acceptable, indicating the reproducibility of person ordering and the stability of the item-difficulty hierarchy, respectively [94, 95]. Person and item separation indices > 1.5 were considered indicative of the instrument’s ability to distinguish at least two distinct strata (e.g., low vs. high) of respondent ability or item difficulty, respectively [90, 95].
- (f) Differential item functioning (DIF): DIF analysis was conducted using ordinal logistic regression to assess whether items functioned differently across subgroups defined by age (< 75 vs. ≥ 75 years) and gender (males vs. females) using the “lordif” R package [96]. Items were flagged for potential DIF if they showed a statistically significant likelihood ratio test ($\chi^2 p < 0.006$, Bonferroni-adjusted threshold of 0.05/8 [97]) and a moderate effect size, defined as a change in McFadden’s pseudo- R^2 (ΔR^2) > 0.13 [98, 99]. Given the smaller male subgroup, a sensitivity analysis was also conducted for gender-related DIF using a more sensitive threshold of $\Delta R^2 \geq 0.02$ [100, 101].

Sample size

For EFA, it is recommended that at least 5 or 10 participants per variable be included, with an absolute minimum sample size of 100 or 200 participants [102]. A sample of at least 300 participants is generally considered adequate for CFA [103]. A minimum sample size of 250 participants was required to fit the PCM [104, 105]. Based on these considerations, the target sample size for the initial survey was set at 300 participants. For the test-retest reliability assessment, a power calculation indicated that at least 57 participants were needed for the retest survey to distinguish good ($ICC \geq 0.75$) from poor ($ICC < 0.50$) reliability, assuming $\alpha = 0.05$ and 80% power and allowing for the potential exclusion of 20% of participants whose living conditions changed between assessments [106].

Results

Sample characteristics

A total of 312 participants were included in the data analyses and their characteristics are summarized in Table 1. The summary score and responses for each Chinese ASCOT-SCT4 item are shown in Table 2.

CTT results

Internal consistency The corrected item-total correlations for all items were above 0.20 (Table S4). Cronbach's alphas for the overall scale and for the scale excluding specific items were all above 0.70. The alpha decreased if any item was deleted, except for the *Dignity* item, where it improved slightly from 0.79 to 0.80.

Factor structure The eigenvalue criterion, scree plot (Figure S1), and parallel analysis (Figure S2) supported the retention of one factor. The data were suitable for factor analysis, as indicated by a KMO value greater than 0.80 and a statistically significant Bartlett's test of sphericity ($p < 0.001$).

The one-factor EFA solution accounted for 39.4% of the total variance, with all item loadings exceeding 0.30, except for the *Dignity* item, which had a loading of 0.269 (Table S5). A two-factor solution did not improve the factor loading of this item. Despite the weak loading, and although its removal increased the explained variance by 4.5% points in the one-factor structure, the *Dignity* item was retained due to its conceptual importance in measuring SCRQoL.

Construct validity – scale level

(a) Convergent validity

As hypothesized, the Chinese ASCOT-SCT4 summary score showed strong positive correlations with CASP-12 ($r = 0.635$, $p < 0.001$, H2) and EQ-5D-5 L ($r = 0.513$, $p < 0.001$, H5), moderate positive correlations with the life satisfaction scale ($r = 0.488$, $p < 0.001$, H1) and EQ-VAS ($r = 0.415$, $p < 0.001$, H6), and a strong negative correlation with PHQ-9 ($r = -0.585$, $p < 0.001$, H3). In line with expectations, weaker correlations were observed with the EQ-5D-5 L than with the CASP-12 (H5), and with the EQ-VAS than with the life satisfaction scale (H6).

(b) Known-group validity

All known-group comparisons were statistically significant and followed the expected directions (Fig. 1). Specifically, higher Chinese ASCOT-SCT4 summary scores were observed among respondents without depression (H4), those without (I)ADL limitations (H7, H8), and those satisfied with care services (H9). The magnitude of these differences was large for the depression groups, moderate for the ADL and care satisfaction groups, and small for the IADL groups.

Construct validity – item level As presented in Table 3, all individual items of the Chinese ASCOT-SCT4 showed correlations with the comparator instruments including EQ-5D-5 L (h5), EQ-VAS (h6), CASP-12 (h2), PHQ-9 (h4), and life satisfaction (h1) in the expected direction. However, a non-monotonic pattern was observed for the *Dignity* item, where the second response level ('no needs') was associated with higher comparator instrument scores than the first response level ('ideal state') across all measures except the life satisfaction scale (h2 and h4 not met). As expected, post-hoc comparisons showed significant differences in mean CASP-12-control scores between the 'ideal state' and 'no needs' response options for all items except *Dignity* (h3).

Relationships between individual items and the corresponding categorical variables are summarized in Table 4. For *Control over daily life*, participants without (I)ADL limitations were more likely to select the 'ideal state' (h-control1, h-control2). For *Personal cleanliness*, selection of the 'ideal state' was more common among participants without self-care-related ADL limitations and those reporting no problems with EQ-5D-5 L self-care

Table 1 Sample characteristics ($N=312$)

Variables	<i>N</i> (%)
Age group	
60–69	84 (26.92)
70–79	108 (34.62)
80+	120 (38.46)
Age (Mean $\pm$ SD)	76.00 $\pm$ 8.64
Median age (IQR)	75.50 (69.00, 83.00)
Age range	60–95
Gender	
Male	93 (29.81)
Female	219 (70.19)
Living arrangement	
Living alone	111 (35.58)
Living with family or others	201 (64.42)
Education	
Primary or below	136 (43.59)
Secondary	130 (41.67)
Post-secondary	45 (14.42)
Missing	1 (0.32)
Geographical locations	
HK Island	125 (40.06)
Kowloon	58 (18.59)
New Territories	129 (41.35)
Chronic conditions	
None	29 (9.29)
1–2	94 (30.13)
3+	189 (60.58)
ADL with difficulty	
None	157 (50.32)
1–2	92 (29.49)
3+	63 (20.19)
IADL with difficulty	
None	69 (22.12)
1–2	127 (40.71)
3+	116 (37.18)
Types of services received	
Meals	56 (17.95)
Home cleaning	69 (22.12)
Transportation	42 (13.50)
Personal care (Bathing)	7 (2.24)
Leisure activities	204 (65.38)
Physical rehabilitation/exercise	89 (28.53)
Basic nursing care/health management	107 (34.29)
Emergency services	82 (26.28)
Palliative care	5 (1.60)
Day-care centre	10 (3.21)
Volunteer activities	12 (3.85)
Survey mode	
Face-to-face or Zoom	235 (75.32)
Phone	77 (24.68)

SD: standard deviation

IQR: interquartile range

Table 2 Distribution of responses to the Chinese ASCOT-SCT4 items (N=312)

Items/summary score	Ideal state Freq. (%)	No needs Freq. (%)	Some needs Freq. (%)	High needs Freq. (%)
Control over daily life	117 (37.50%)	128 (41.03%)	55 (17.63%)	12 (3.85%)
Personal cleanliness	148 (47.44%)	146 (46.79%)	18 (5.77%)	0 (0.00%)
Food and drink	141 (45.19%)	137 (43.91%)	26 (8.33%)	8 (2.56%)
Personal safety	101 (32.37%)	148 (47.44%)	54 (17.31%)	9 (2.88%)
Social participation	103 (33.01%)	103 (33.01%)	84 (26.92%)	22 (7.05%)
Occupation	111 (35.58%)	108 (34.62%)	69 (22.12%)	24 (7.69%)
Home cleanliness	116 (37.18%)	135 (43.27%)	53 (16.99%)	8 (2.56%)
Dignity	142 (45.51%)	138 (44.23%)	26 (8.33%)	6 (1.92%)
	Mean ± SD	Median	Q1-Q3	Min; Max
Summary score (0–24)	17.33 ± 4.05	18	15.00–20.00	5.00; 24.00

SD: standard deviation

Q1 refers to the first quartile (25th percentile); Q3 refers to the third quartile (75th percentile)

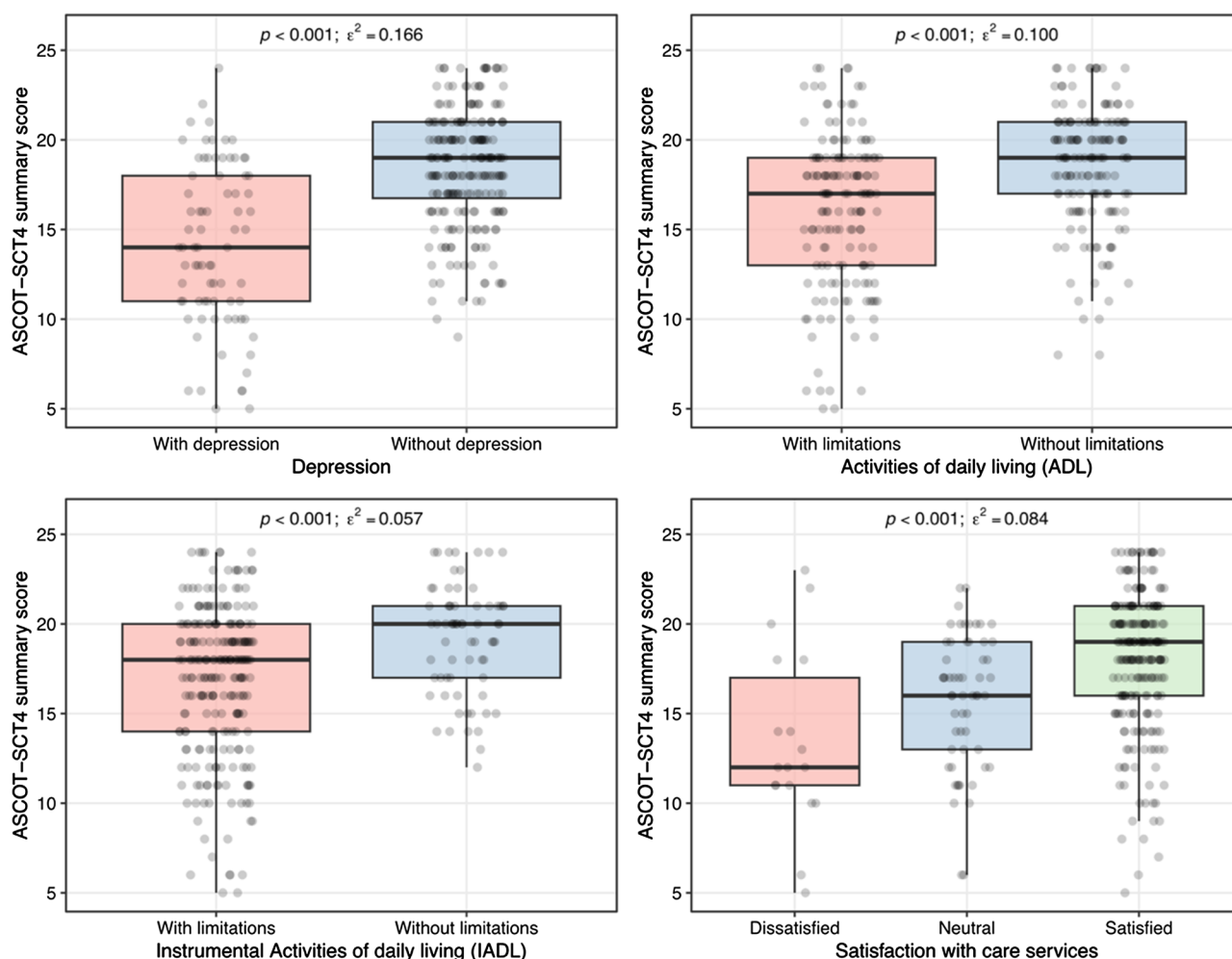


Fig. 1 Differences in Chinese ASCOT-SCT4 summary scores across groups

(h-personal1, h-personal2). Participants without food-related (I)ADL limitations tended to select the ‘ideal state’ for *Food and drink* (h-food1). Negative associations were observed between the *Personal safety* item and depression (h-safety1), as well as between the *Social participation* item and loneliness (h-social1). For *Occupation*, participants reporting no problems with the usual activities on the EQ-5D-5 L were more likely to select higher response levels (h-occu1). Positive associations were observed between the *Dignity* item and both satisfaction

Table 3 Means comparator-instrument scores across Chinese ASCOT-SCT4 response levels

Items (number of participants)	EQ-5D-5 L (n=312)	EQ-VAS (n=307)	CASP-12 (n=306)	CASP-12 Control (n=307)	PHQ-9 (n=310)	Self-rated life satisfaction (n=303)
Control over daily life						
Ideal state (117)	0.829	75.761	38.569	18.991	4.466	7.701
No needs (128)	0.778	73.094	33.904	16.800	6.173	7.224
Some needs (55)	0.504	63.481	28.472	13.593	9.891	6.231
High needs (12)	-0.109	49.000	21.750	10.750	17.417	5.222
<i>H statistic and significance[#]</i>	72.427 (<i>p</i> <0.001)	32.456 (<i>p</i> <0.001)	72.432 (<i>p</i> <0.001)	73.178 (<i>p</i> <0.001)	52.939 (<i>p</i> <0.001)	39.984 (<i>p</i> <0.001)
Personal cleanliness						
Ideal state (148)	0.797	75.986	37.234	18.055	5.136	7.687
No needs (146)	0.674	69.507	32.091	16.049	7.255	6.800
Some needs (18)	0.370	52.941	27.444	13.167	13.778	5.812
<i>H statistic and significance[#]</i>	27.207 (<i>p</i> <0.001)	29.031 (<i>p</i> <0.001)	36.954 (<i>p</i> <0.001)	26.952 (<i>p</i> <0.001)	32.129 (<i>p</i> <0.001)	32.502 (<i>p</i> <0.001)
Food and drink						
Ideal state (141)	0.788	75.357	37.412	18.449	5.343	7.655
No needs (137)	0.701	70.467	32.846	16.102	6.632	6.903
Some needs (26)	0.550	64.923	28.846	13.769	10.231	6.231
High needs (8)	0.195	44.167	22.125	11.625	17.375	6.000
<i>H statistic and significance[#]</i>	21.353 (<i>p</i> <0.001)	18.470 (<i>p</i> <0.001)	45.766 (<i>p</i> <0.001)	43.605 (<i>p</i> <0.001)	24.417 (<i>p</i> <0.001)	34.191 (<i>p</i> <0.001)
Personal safety						
Ideal state (101)	0.873	77.525	37.870	18.860	4.020	7.842
No needs (148)	0.712	72.628	35.133	17.264	5.796	7.203
Some needs (54)	0.507	60.755	27.444	12.963	12.093	6.078
High needs (9)	0.227	54.375	21.000	10.444	16.444	5.375
<i>H statistic and significance[#]</i>	64.949 (<i>p</i> <0.001)	40.604 (<i>p</i> <0.001)	69.892 (<i>p</i> <0.001)	74.693 (<i>p</i> <0.001)	74.407 (<i>p</i> <0.001)	42.224 (<i>p</i> <0.001)
Social participation						
Ideal state (103)	0.795 ^{&}	75.569 ^{&}	37.931 ^{&}	19.069	4.206	7.765 [^]
No needs (103)	0.780	73.971	36.420	17.820	5.412	7.455
Some needs (84)	0.671	67.305	30.434	14.735	8.440	6.646
High needs (22)	0.196	59.238	21.955	9.818	16.591	4.722
<i>H statistic and significance</i>	53.722 (<i>p</i> <0.001)	24.420 (<i>p</i> =0.001)	77.859 (<i>p</i> <0.001)	96.156 (<i>p</i> <0.001)	72.638 (<i>p</i> <0.001)	51.176 (<i>p</i> <0.001)
Occupation						
Ideal state (111)	0.828	75.883	38.944	19.565	3.964	7.622 ^{&}
No needs (108)	0.762	72.407	34.406	16.868	6.019	7.393
Some needs (69)	0.625	69.045	31.088	14.507	8.420	6.615
High needs (24)	0.235	54.619	21.458	11.000	16.417	5.400
<i>H statistic and significance</i>	50.473 (<i>p</i> <0.001)	27.811 (<i>p</i> <0.001)	89.476 (<i>p</i> <0.001)	98.571 (<i>p</i> <0.001)	71.113 (<i>p</i> <0.001)	37.468 (<i>p</i> <0.001)
Home cleanliness						
Ideal state (116)	0.753	74.417 [^]	36.812	18.336	4.461	7.439 ^{&}
No needs (135)	0.728	72.647	34.353	16.835	6.799	7.351
Some needs (53)	0.631	64.725	29.906	14.321	9.755	6.280
High needs (8)	0.487	61.875	25.625	12.000	14.250	6.250
<i>H statistic and significance[#]</i>	19.567 (<i>p</i> =0.017)	16.753 (<i>p</i> <0.001)	33.443 (<i>p</i> <0.001)	37.956 (<i>p</i> <0.001)	47.077 (<i>p</i> <0.001)	21.943 (<i>p</i> <0.001)
Dignity						
Ideal state (142)	0.731 ^{&}	73.115 ^{&}	34.616 ^{&}	16.964 ^{&}	6.390 ^{&}	7.348 ^{&}
No needs (138)	0.757	72.139	35.426	17.493	5.511	7.348
Some needs (26)	0.489	64.000	29.885	14.115	11.308	6.375

Table 3 (continued)

Items (number of participants)	EQ-5D-5 L (<i>n</i> =312)	EQ-VAS (<i>n</i> =307)	CASP-12 (<i>n</i> =306)	CASP-12 Control (<i>n</i> =307)	PHQ-9 (<i>n</i> =310)	Self-rated life satisfaction (<i>n</i> =303)
High needs (6)	0.333	61.667	18.333	10.333	17.500	2.667
<i>H statistic and significance</i> [#]	16.749 (<i>p</i> <0.001)	6.595 (<i>p</i> <0.001)	20.148 (<i>p</i> <0.001)	21.661 (<i>p</i> <0.001)	25.679 (<i>p</i> <0.001)	22.672 (<i>p</i> <0.001)

[#]The ‘some needs’ and ‘high needs’ response categories were combined for the Kruskal–Wallis analyses because of the small number of responses in the ‘high needs’ category

EQ-5D-5 L: The 5-level EQ-5D

PHQ-9: The 9-item Patient Health Questionnaire

CASP-12: The Control, Autonomy, Self-realization, and Pleasure-12

& indicates that the difference between ‘ideal state’ and ‘no needs’ was not statistically significant at the 10% level ($p \geq 0.1$)

^ indicates that the difference between ‘ideal state’ and ‘no needs’ was not statistically significant at the 5% level but was significant at the 10% level ($0.05 \leq p < 0.1$)

Italicized values indicate mean scores that did not follow the expected monotonic ordering across the ASCOT-SCT4 response levels

with care services (h-dignity1) and perceived respect from care staff (h-dignity2). However, neither the association between living arrangement and *Personal safety* nor that between limitations in household cleaning and *Home cleanliness* reached statistical significance (h-safety2 and h-home1 not met).

Test-retest reliability Among the 119 participants (38.14% of the first survey sample) who completed the follow-up survey, the median retest interval was 19 days (interquartile range [IQR]: 14–22.5 days; range: 8–30 days). The ICC for the summary score was 0.809, indicating good test-retest reliability of the instrument. At the item level, weighted kappa coefficients ranged from 0.404 for *Dignity* to 0.665 for *Occupation*, indicating moderate test-retest agreement (Table 5). The sensitivity analysis showed no statistically significant association between the retest interval and the absolute test-retest difference in the summary score ($r=0.020$, $p=0.829$) or the absolute test-retest differences in individual item response categories (r ranged from -0.018 to 0.143 ; all $p>0.05$).

Rasch analysis results

Assumptions: unidimensionality and local independence The CFA supported unidimensionality of the instrument as indicated by fit statistics (CFI=0.982, TLI=0.975, RMSEA=0.054, and SRMR=0.046). Local independence was also supported; residual correlations for almost all item pairs were below 0.30. Only one pair (*Personal cleanliness* and *Occupation*) was marginally above this threshold with a correlation of -0.303 (Table S6).

Item fit and thresholds The estimated item thresholds and item fit are presented in Table 6. All items showed acceptable fit, with Infit and Outfit values within the range of 0.5–1.5. The ICCCs showed ordered thresholds for every item and a distinct, non-overlapping probability peak for each response category on the latent trait scale (Fig. 2).

Targeting As shown in the person-item map (Fig. 3), the items generally covered a broad portion of the person ability distribution; however, a measurement gap was evident at the lower end (below -1.444 logits). The distribution of person abilities was right-skewed, with an average of -0.822 logits, suggesting that, on the whole, the sample generally had good SCRQoL. The difference between the mean person ability and item difficulty (i.e., 0.0 logits) was within the acceptable range (difference < 1.0 logit).

Person and item reliability Both person and item separation reliabilities (0.721 and 0.964, respectively) and separation indices (1.606 and 5.182, respectively) met the acceptable thresholds (> 0.70 and > 1.50 , respectively).

DIF As shown in Table S7, although the likelihood-ratio tests for age-related DIF in *Home cleanliness* reached statistical significance ($p<0.006$), no items were flagged for potential DIF across age or gender groups because

Table 4 Relationship between the Chinese ASCOT-SCT4 items and related items of comparator instruments or demographic variables

Comparator variables	Ideal state Freq. (%)	No needs Freq. (%)	Some needs Freq. (%)	High needs Freq. (%)	<i>p</i> -value
Control over daily life					
Limitations in ADLs					< 0.001
No (<i>n</i> = 157)	81 (69.23%)	63 (49.22%)	13 (23.64%)	0 (0.00%)	
Yes (<i>n</i> = 155)	36 (30.77%)	65 (50.78%)	42 (76.36%)	12 (100.00%)	
Limitations in IADLs					
No (<i>n</i> = 69)	38 (32.48%)	28 (21.88%)	3 (5.45%)	0 (0.00%)	< 0.001
Yes (<i>n</i> = 243)	79 (67.52%)	100 (78.13%)	52 (94.55%)	12 (100.00%)	
Personal cleanliness					
Limitations in Self-care-related ADLs ^a					< 0.001
No (<i>n</i> = 204)	112 (75.68%)	86 (58.90%)	6 (33.33%)	0 (0.00%)	
Yes (<i>n</i> = 108)	36 (24.32%)	60 (41.10%)	12 (66.67%)	0 (0.00%)	
EQ-5D-5 L Self-care					< 0.001
No problem (<i>n</i> = 253)	134 (90.54%)	112 (76.71%)	7 (38.89%)	0 (0.00%)	
Had problems (<i>n</i> = 59)	14 (9.46%)	34 (23.29%)	11 (61.11%)	0 (0.00%)	
Food and drink					
Limitations in Food-related (I)ADLs items ^b					0.001
No (<i>n</i> = 197)	104 (74.47%)	77 (57.66%)	14 (53.85%)	2 (25.00%)	
Yes (<i>n</i> = 115)	37 (25.53%)	60 (42.34%)	12 (46.15%)	6 (75.00%)	
Personal safety					
Living arrangement					0.324
Living with others (<i>n</i> = 201)	69 (68.32%)	97 (65.54%)	29 (53.70%)	6 (66.67%)	
Living alone (<i>n</i> = 111)	32 (31.68%)	51 (34.46%)	25 (46.30%)	3 (33.33%)	
Depression					< 0.001
No (<i>n</i> = 232)	88 (88.00%)	122 (82.99%)	21 (38.89%)	1 (11.11%)	
Yes (<i>n</i> = 78)	12 (12.00%)	25 (17.01%)	33 (61.11%)	8 (88.89%)	
Social participation					
Feelings of loneliness					< 0.001
No (<i>n</i> = 183)	79 (77.45%)	67 (66.34%)	36 (42.86%)	1 (4.55%)	
Yes (<i>n</i> = 126)	23 (22.55%)	34 (33.66%)	48 (57.14%)	21 (95.45%)	
Occupation					
EQ-5D-5 L Usual activities					< 0.001
No problem (<i>n</i> = 195)	86 (77.48%)	73 (67.59%)	32 (46.38%)	4 (16.67%)	
Had problems (<i>n</i> = 117)	25 (22.52%)	35 (32.41%)	37 (53.62%)	20 (83.33%)	
Home cleanliness					
Limitations in household cleaning					0.112
No (<i>n</i> = 219)	84 (72.41%)	86 (63.70%)	42 (79.25%)	7 (87.50%)	
Yes (<i>n</i> = 93)	32 (27.59%)	49 (36.30%)	11 (20.75%)	1 (12.50%)	
Dignity					
Satisfaction with care services					< 0.001
Very satisfied or somewhat satisfied (<i>n</i> = 233)	119 (83.80%)	102 (76.12%)	12 (46.15%)	0 (0.00%)	
Neutral (<i>n</i> = 57)	18 (12.68%)	25 (18.66%)	10 (38.46%)	4 (66.67%)	
Very dissatisfied or somewhat dissatisfied (<i>n</i> = 18)	5 (3.52%)	7 (5.22%)	4 (15.38%)	2 (33.33%)	
Respected by care staff					< 0.001
All of the time (<i>n</i> = 225)	109 (76.76%)	101 (75.37%)	14 (53.85%)	1 (16.67%)	
Some of the time (<i>n</i> = 73)	30 (21.13%)	32 (23.88%)	9 (34.62%)	2 (33.33%)	
None of the time (<i>n</i> = 10)	3 (2.11%)	1 (0.75%)	3 (11.54%)	3 (50.00%)	

^aSelf-care-related ADLs Included grooming, toileting, bathing, dressing, bowel control, and bladder control^bFood-related (I)ADLs included eating, food preparation, and shopping

Table 5 Test-retest reliability of the Chinese ASCOT-SCT4 ($n=119$)

CI: confidence interval
ICC: intraclass correlation coefficient
 $p < 0.001$ for all values

Items/Sum scores	Weighted kappa, 95% CI
Control over daily life	0.632 (0.514, 0.750)
Personal cleanliness	0.445 (0.289, 0.600)
Food and drink	0.550 (0.398, 0.701)
Personal safety	0.564 (0.441, 0.686)
Social participation	0.652 (0.515, 0.788)
Occupation	0.665 (0.550, 0.779)
Home cleanliness	0.556 (0.410, 0.701)
Dignity	0.404 (0.202, 0.605)
	ICC, 95% CI
Summary score	0.809 (0.736, 0.863)

Table 6 Estimated thresholds and item fit from partial credit model for the Chinese ASCOT-SCT4 ($N=312$)

Items	Mean threshold	Threshold 1 (SE)	Threshold 2 (SE)	Threshold 3 (SE)	Outfit MSQ	Infit MSQ
Control over daily life	0.449	-1.118 (0.139)	0.511 (0.173)	1.954 (0.337)	0.849	0.832
Personal cleanliness	0.640	-0.823 (0.129)	2.104 (0.262)	/	0.905	0.917
Food and drink	0.826	-0.872 (0.130)	1.524 (0.224)	1.827 (0.426)	0.791	0.838
Personal safety	0.493	-1.444 (0.141)	0.674 (0.171)	2.248 (0.376)	0.858	0.845
Social participation	0.036	-1.135 (0.151)	-0.309 (0.159)	1.551 (0.256)	0.715	0.739
Occupation	0.071	-1.070 (0.146)	-0.022 (0.166)	1.304 (0.255)	0.797	0.785
Home cleanliness	0.616	-1.168 (0.137)	0.638 (0.174)	2.379 (0.395)	1.045	1.039
Dignity	0.946	-0.863 (0.130)	1.555 (0.225)	2.147 (0.472)	1.336	1.247

MSQ: mean-square value

SE: standard error

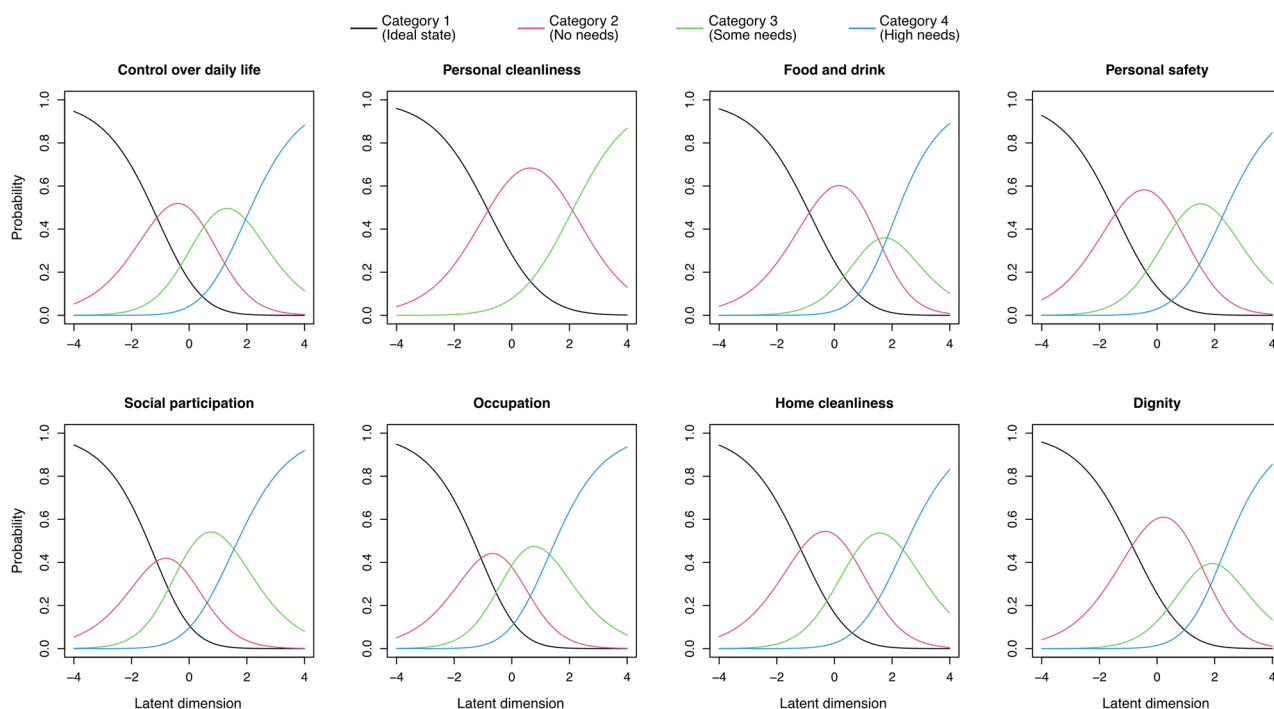
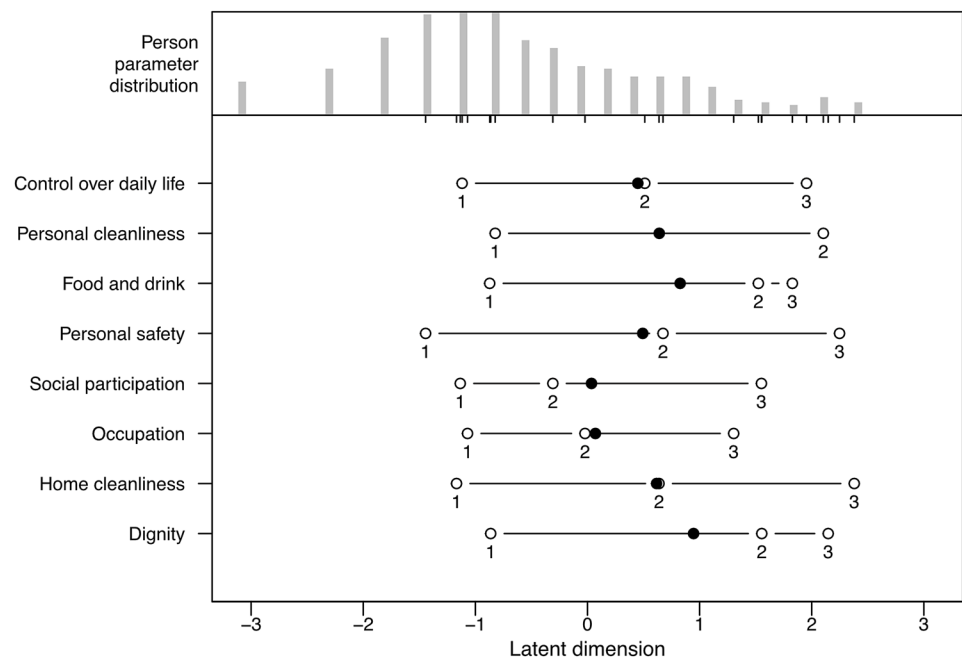

Fig. 2 Item Category Characteristic Curves (ICCCs) for each item of the Chinese ASCOT-SCT4

Fig. 3 Person-item map of the Chinese ASCOT-SCT4. Notes: For each of the eight ASCOT-SCT4 items, the open circles (○) labeled 1, 2, and 3 represent the item thresholds—the transition points between response categories 1–2, 2–3, and 3–4, respectively. The solid black dot (●) indicates the mean difficulty level for that specific item. In this model, lower (more negative) logit values represent higher (better) social care-related QoL (SCRQoL). The item difficulty levels (●) and category thresholds (○) are anchored to a mean of 0.0 logits



all observed effect sizes (ΔR^2) were well below the 0.13 threshold. In the sensitivity analysis, all gender-related ΔR^2 values remained below 0.02.

Discussion

This study is the first to investigate the ASCOT-SCT4 in a Chinese population and provides initial empirical evidence supporting the construct validity and reliability of its Chinese version among older adults receiving community care and support services in HK, based on both CTT and Rasch analysis. Nonetheless, the performance of the *Dignity* item and the instrument's precision for higher levels of SCRQoL warrant careful consideration.

Evidence from hypothesis testing largely supported construct validity at both scale and item levels and was generally consistent with findings from previous validation studies [27, 32–35, 56, 61]. At the item level, however, the expected association between *Personal safety* and living arrangement was not observed, consistent with findings from the Dutch [61] and Finnish [34] validation studies but not with those from the English study [32]. This inconsistency across studies may be attributable either to cultural or contextual differences in perceptions of safety or to the influence of factors other than living arrangement in HK. *Home cleanliness* was also not associated with limitations in household cleaning as hypothesized, a relationship for which evidence from previous ASCOT validation studies remains limited. A German study [35] reported such an association when respondents considered their expected situation in the hypothetical absence of care services. By contrast, the present study assessed participants' current situation while they were receiving care and support; such support may have compensated for limitations in household cleaning and attenuated the association between these limitations and the attained level of home cleanliness. Further hypothesis testing using other comparator variables is needed to provide more evidence on the construct validity of *Personal safety* and *Home cleanliness*.

The structural analysis via EFA supported a one-factor structure for the Chinese ASCOT-SCT4, consistent with the original English version [17]. However, the weak factor loading for *Dignity* suggests that this item may reflect the underlying SCRQoL construct less strongly than the other items in the HK context. This finding is consistent with previous studies in community care settings, which reported relatively low *Dignity* factor loadings of 0.20 [107], 0.28 [34], and 0.41 [36]. By comparison, generally higher *Dignity* loadings were reported for residential

care samples (0.48 [108] and 0.49 [72]) and samples comprising both community and residential care users (0.41 [49] and 0.54 [53]). Although the findings are mixed, the evidence points to weaker factor loadings for *Dignity* than for the other items, suggesting that it may contribute less to the SCRQoL construct. One possible explanation is that the relevance of *Dignity* may differ from that of the other items in that the intensity of care received may have a greater influence on it. Compared with residential care users, community care users are generally healthier and more independent and receive less intensive care [109]. *Dignity* may therefore be less central to their overall QoL, making the “no influence” response more applicable to them [17, 53]. Some support for this interpretation comes from cognitive interviews conducted during the development of the English ASCOT-SCT4, in which respondents who needed less help and received fewer services tended to select “no influence” [32, 53]. Similarly, some participants in our study who selected “no influence” mentioned that if receiving help made them feel worse about themselves, they could choose not to receive it, whereas those reporting such a negative impact indicated that they had little choice but to continue receiving care because they needed it to manage their daily lives. This explanation may also account for the non-monotonic associations of *Dignity* with external health and well-being instruments observed in both the current study and the English validation study [32], with comparable health and well-being reported by respondents selecting the second response option and those selecting the top response option.

Beyond care intensity, cultural factors should be considered when interpreting the weaker factor loading of *Dignity*. Our previous cross-cultural adaptation and content-validation study [31] supported the conceptual equivalence and relevance of *Dignity* in HK after the unfamiliar literal phrase “think and feel about myself” was adapted using “self-esteem” and “self-confidence”. Nevertheless, the use of different expressions to convey *Dignity* across language versions may lead to systematic differences in how respondents interpret and respond to the item. Whether such differences contributed to the weaker factor loading remains unclear. At this stage, removing the *Dignity* item would be premature, given its contribution to the content comprehensiveness of the instrument [31]. In addition, excluding *Dignity* only marginally improved internal consistency; the item also showed the expected associations with both indicators of perceived quality of care and demonstrated acceptable fit under the Rasch measurement framework. However, studies involving larger and more diverse samples, including residential care users, are needed to delve deeper into the interpretation and psychometric performance of the *Dignity* item and to clarify its role in the dominant factor structure. The measurement equivalence of the *Dignity* item across language versions should also be examined. Continued evidence of a low factor loading for *Dignity* across validation studies and care settings in HK would indicate a need to review the item, including whether it should be revised or removed.

The Rasch analysis validated that the response options functioned hierarchically and distinctly, suggesting that the response categories for all items were nonredundant and distinguishable by respondents. The response categories functioned more clearly in our study than in previous validation studies of the Japanese ASCOT-SCT4 (using a graded response model) [36] and the English ASCOT-SCT4 (using the PCM) [72], in which the ‘some needs’ category overlapped substantially with adjacent response categories for half of the items, particularly those concerning basic needs, such as *Food and drink* and *Personal cleanliness*. These divergent findings may be attributable to variations in survey administration methods. The Japanese study used self-completion [36], whereas the English validation study relied on proxy reports from care staff [72]. Responses in the present study were instead obtained through interviewer-assisted administration, potentially contributing to higher response quality. Future research should investigate whether response category functioning differs systematically across various administration methods. Despite the positive finding on response category structure, the Rasch analysis of targeting revealed room for improvement. Limited item coverage was evident at the lower end of the latent trait scale on the person-item map, indicating that the instrument may measure SCRQoL less precisely among those with higher SCRQoL. This ceiling effect could result in an underestimation of the incremental benefits provided by care interventions, potentially leading to inefficient resource allocation in economic evaluations. Future refinement could involve adding new items or response categories that are easier to endorse, thereby improving measurement among individuals with higher SCRQoL.

The findings suggest that the Chinese ASCOT-SCT4 has potential as an LTC-specific preference-based outcome measure in HK. By capturing service users' own perspectives on aspects of QoL that LTC services are intended to influence, it complements existing health-focused outcome measures and offers a person-centred approach to evaluating LTC services. At the individual level, domain profiles may facilitate communication between service users and care providers and inform care planning. At the service and system levels, the instrument could support routine outcome monitoring, service evaluation, and quality assessment. When HK-specific preference weights become available, it could also be used in local economic evaluations and to inform resource-allocation decisions in LTC.

This study has several limitations. Firstly, participants were recruited through convenience sampling among users receiving publicly funded community care and support services. This may have introduced selection bias and limited the generalizability of the findings to the broader LTC population, particularly residential care users and LTC users not receiving publicly funded services, as these groups may differ from the present sample in the care services received, levels of care intensity, and care environments, potentially influencing the instrument's psychometric performance. Future studies should validate the instrument in these groups and among populations receiving different levels of care intensity. Secondly, the male subgroup ($n = 93$) in the DIF analysis included fewer than the minimum recommended 100 participants [110], which warrants caution when interpreting the results. Additionally, the small number of responses in the 'high needs' categories may have reduced the precision of threshold estimates in the PCM. Studies with larger samples are needed to confirm these findings. Lastly, other important psychometric properties, such as responsiveness and interpretability, including minimally important differences [66], were not evaluated in this study, underscoring the need for further research.

Conclusion

This study provides evidence, based on both CTT and Rasch analyses, that the Chinese version of the ASCOT-SCT4 is valid and reliable among older adults receiving community care and support services in HK. The findings support its potential as a tool for outcome measurement, quality assessment, and economic evaluation in LTC. However, the suboptimal performance of the *Dignity* item and limited measurement coverage at higher SCRQoL levels warrant further investigation. Furthermore, its use in economic evaluations could be facilitated by developing local preference weights from a representative sample of a relevant population (e.g., the general population or LTC users) using valuation methods such as discrete choice experiments or best–worst scaling.

Abbreviations

HK	Hong Kong
LTC	Long-term care
QoL	Quality of life
HRQoL	Health-related quality of life
EQ-5D	EuroQol-5 Dimension
ICECAP-O	ICEpop CAPability measure for Older people
ASCOT	Adult Social Care Outcomes Toolkit
ASCOT-SCT4	Adult Social Care Outcomes Toolkit four-level self-completion questionnaire
SCRQoL	Social care-related quality of life
NICE	National Institute for Health and Care Excellence
SF-6D	Short-Form Six-Dimension
HUI2/3	Health Utilities Index Mark 2/3
CTT	Classical Test Theory
EQ-5D-5L	EuroQol-5 Dimension 5-Level

EQ-VAS	EuroQol Visual Analogue Scale
PHQ-9	9-item Patient Health Questionnaire
CASP-12	Control, Autonomy, Self-realization, and Pleasure-12
ADL	Activities of Daily Living
IADL	Instrumental Activities of Daily Living
UCLA	University of California, Los Angeles
EFA	Exploratory Factor Analysis
KMO	Kaiser-Meyer-Olkin
ICC	Intraclass Correlation Coefficient
PCM	Partial Credit Model
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
TLI	Tucker-Lewis Index
RMSEA	Root Mean Square Error of Approximation
SRMR	Standardized Root Mean Square Residual
ICCC	Item Category Characteristic Curve
DIF	Differential Item Functioning
IQR	Interquartile range

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1186/s12955-026-02640-7>.

Acknowledgements We sincerely thank all study participants for their time and contributions. We are also grateful to the collaborating non-governmental organizations, elderly community centers, and social workers for their essential assistance with participant recruitment and data collection. Furthermore, we extend our thanks to the ASCOT developers at the University of Kent for their support and to all research staff who contributed to this project.

Author contributions SY, RX, and ELW conceived and designed the study. SY and JS recruited participants and conducted the interviews. SY conducted the statistical analyses, interpreted the results, and drafted the initial manuscript. RX and ELW provided supervision for the data analysis, interpretation of the findings, and manuscript writing. KW, AC, EW, and NS critically reviewed the manuscript for important intellectual content and contributed to its revision. All authors have read and approved the final manuscript for submission.

Funding This study was funded by the Health and Medical Research Fund (HMRF) from the Health Bureau of Hong Kong (Ref. No. 18190641). The funder had no role in the study design, data collection, data analysis, data interpretation, or manuscript preparation.

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate This study received ethics approval from the Clinical Research Ethics Committee at the Chinese University of Hong Kong (Ref. No. NTEC-2021-281). Written informed consent was obtained from all participants.

Consent for publication Not applicable. The study does not report any identifiable individual data; therefore, consent for publication was not required.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The

images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. World Bank. Life expectancy at birth, total (years) [Internet]. World Bank Open Data. <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>. Accessed 6 Aug 2026.
2. World Bank. Fertility rate, total (births per woman) [Internet]. World Bank Open Data. <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN>. Accessed 6 Aug 2026.
3. Census and Statistics Department, Hong Kong Special Administrative Region. Hong Kong population projections 2022–2046 [Internet]. Hong Kong: Census and Statistics Department, Hong Kong Special Administrative Region; 2023. https://www.censtatd.gov.hk/en/data/stat_report/product/B1120015/att/B1120015092023XXXXB01.pdf. Accessed 17 Mar 2025.
4. Gu D, Andreev K, Dupre ME. Major trends in population growth around the world. *China CDC Wkly*. 2021;3:604–13. <https://doi.org/10.46234/ccdcw2021.160>.
5. Census and Statistics Department, Hong Kong Special Administrative Region. Special topics report no. 63: persons with disabilities and chronic diseases [Internet]. Hong Kong: Census and Statistics Department, Hong Kong Special Administrative Region; 2021. Report No.: 63. https://www.censtatd.gov.hk/en/data/stat_report/product/C0000055/att/B11301632021XXXXB0100.pdf. Accessed 17 Mar 2025.
6. Yu R, Leung G, Leung J, Cheng C, Kong S, Tam LY, et al. Prevalence and Distribution of Intrinsic Capacity and Its Associations with Health Outcomes in Older People: The Jockey Club Community eHealth Care Project in Hong Kong. *J Frailty Aging*. 2022;11:302–8. <https://doi.org/10.14283/jfa.2022.19>.
7. Nicholson NR. A review of social isolation: an important but underassessed condition in older adults. *J Prim Prev*. 2012;33:137–52. <https://doi.org/10.1007/s10935-012-0271-2>.
8. Fung AWT, Chan W-C, Wong CS-M, Chen EY-H, Ng RM-K, Lee EH-M, et al. Prevalence of anxiety disorders in community dwelling older adults in Hong Kong. *Int Psychogeriatr*. 2017;29:259–67. <https://doi.org/10.1017/S1041610216001617>.
9. Elderly Commission. Elderly services programme plan [Internet]. Hong Kong: Elderly Commission; 2017. <https://www.elderlycommission.gov.hk/wp-content/uploads/2024/02/Full-Report-of-the-ESPP.pdf>. Accessed 17 Mar 2025.
10. Research Office, Legislative Council Secretariat. Long-term care insurance systems in selected places [Internet]. Hong Kong: Legislative Council of the Hong Kong Special Administrative Region; 2023. https://app7.legco.gov.hk/rpdb/en/uploads/2023/IN/IN15_2023_20230728_en.pdf. Report No.: IN15/2023. Accessed 13 Aug 2026.
11. Social Welfare Department, Hong Kong Special Administrative Region. Waiting list for subsidised community care services for the elderly [Internet]. Social Welfare Department; 2026. [https://www.swd.gov.hk/storage/asset/section/1021/en/LTC%20Statistics%20CCS\(Eng\)\(Jan%202026\).pdf](https://www.swd.gov.hk/storage/asset/section/1021/en/LTC%20Statistics%20CCS(Eng)(Jan%202026).pdf). Accessed 8 June 2026.
12. Social Welfare Department, Hong Kong Special Administrative Region. Waiting list for subsidised residential care services for the elderly [Internet]. Social Welfare Department; 2026. [https://www.swd.gov.hk/storage/asset/section/1022/en/CWL/LTC%20statistics_HP-EN\(202604\).pdf](https://www.swd.gov.hk/storage/asset/section/1022/en/CWL/LTC%20statistics_HP-EN(202604).pdf). Accessed 6 Aug 2026.
13. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 4th ed. Oxford: Oxford University Press; 2015.
14. Makai P, Brouwer WB, Koopmanschap MA, Stolk EA, Nieboer AP. Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. *Soc Sci Med*. 2014;102:83–93. <https://doi.org/10.1016/j.socscimed.2013.11.050>.
15. Bulamu NB, Kaambwa B, Ratcliffe J. A systematic review of instruments for measuring outcomes in economic evaluation within aged care. *Health Qual Life Outcomes*. 2015;13:179. <https://doi.org/10.1186/s12955-015-0372-8>.
16. World Health Organization. *The world health report 2000: health systems: improving performance*. Geneva: World Health Organization; 2000.
17. Netten A, Burge P, Malley J, Potoglou D, Towers A-M, Brazier J, et al. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess*. 2012;16:1–166. <https://doi.org/10.3310/hta16160>.
18. Kane RA, Kane RL, Ladd RC. *The heart of long-term care*. New York: Oxford University Press; 1998. <https://doi.org/10.1093/oso/9780195122381.001.0001>.
19. Milte CM, Walker R, Luszcz MA, Lancsar E, Kaambwa B, Ratcliffe J. How important is health status in defining quality of life for older people? An exploratory study of the views of older South Australians. *Appl Health Econ Health Policy*. 2014;12:73–84. <https://doi.org/10.1007/s40258-013-0068-3>.

20. Forder JE, Caiels J. Measuring the outcomes of long-term care. *Soc Sci Med.* 2011;73:1766–74. <https://doi.org/10.1016/j.socscimed.2011.09.023>.
21. Bulamu NB, Kaambwa B, Ratcliffe J. Economic evaluations in community aged care: a systematic review. *BMC Health Serv Res.* 2018;18:967. <https://doi.org/10.1186/s12913-018-3785-3>.
22. Brazier J, Tsuchiya A. Improving cross-sector comparisons: going beyond the health-related QALY. *Appl Health Econ Health Policy.* 2015;13:557–65. <https://doi.org/10.1007/s40258-015-0194-1>.
23. Proud L, McLoughlin C, Kinghorn P. ICECAP-O, the current state of play: A systematic review of studies reporting the psychometric properties and use of the instrument over the decade since its publication. *Qual Life Res.* 2019;28:1429–39. <https://doi.org/10.1007/s11136-019-02114-y>.
24. Rand S, Smith N, Welch E, Allan S, Caiels J, Towers A-M. Use of the adult social care outcomes toolkit (ASCOT) in research studies: an international scoping review. *Qual Life Res.* 2025;34:2437–50. <https://doi.org/10.1007/s11136-025-03958-3>.
25. National Institute for Health and Care Excellence. The Social Care Guidance Manual [Internet]. London: National Institute for Health and Care Excellence (NICE); 2016. <http://www.ncbi.nlm.nih.gov/books/NBK395863/>. Accessed 30 Apr 2023.
26. Nussbaum MC, Sen A, editors. The quality of life. Oxford: Clarendon; 1993. <https://doi.org/10.1093/0198287976.001.0001>.
27. Hackert MQN, van Exel J, Brouwer WBF. Valid Outcome Measures in Care for Older People: Comparing the ASCOT and the ICECAP-O. *Value Health.* 2017;20:936–44. <https://doi.org/10.1016/j.jval.2017.03.012>.
28. Wong EL, Ramos-Goni JM, Cheung AW, Wong AY, Rivero-Arias O. Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. *Patient.* 2018;11:235–47. <https://doi.org/10.1007/s40271-017-0278-0>.
29. Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. *Value Health.* 2008;11:295–303. <https://doi.org/10.1111/j.1524-4733.2007.00233.x>.
30. Kehyayan V, Yasin YM, Al-Hamad A. Toward a Clearer Understanding of Value-Based Healthcare: A Concept Analysis. *J Nurs Manag.* 2025;2025:8186530. <https://doi.org/10.1155/jonm/8186530>.
31. Yu S, Sze JCD, Cheung AWL, Wang K, Welch E, Smith N, et al. Translation, cultural adaptation, and content validation of the Hong Kong Chinese version of Self-completion Adult Social Care Outcomes Toolkit (ASCOT-SCT4) for care service users. *Health Qual Life Outcomes.* 2025;23:57. <https://doi.org/10.1186/s12955-025-02389-5>.
32. Malley JN, Towers A-M, Netten AP, Brazier JE, Forder JE, Flynn T. An assessment of the construct validity of the ASCOT measure of social care-related quality of life with older people. *Health Qual Life Outcomes.* 2012;10:21. <https://doi.org/10.1186/1477-7525-10-21>.
33. Rand S, Malley J, Towers A-M, Netten A, Forder J. Validity and test-retest reliability of the self-completion adult social care outcomes toolkit (ASCOT-SCT4) with adults with long-term physical, sensory and mental health conditions in England. *Health Qual Life Outcomes.* 2017;15:163. <https://doi.org/10.1186/s12955-017-0739-0>.
34. Nguyen L, Linnosmaa I, Jokimäki H, Rand S, Malley J, Razik K, et al. Social care-related outcomes in Finland. Construct validity and structural characteristics of the Finnish ASCOT measure with older home care users. *Health Soc Care Community.* 2021;29:712–28. <https://doi.org/10.1111/hsc.13328>.
35. Trukeschitz B, Litschauer J, Hajji A, Kieninger J, Schoch A, Malley J, et al. Cross-cultural adaptation and construct validity of the German version of the Adult Social Care Outcomes Toolkit for service users (German ASCOT). *Health Qual Life Outcomes.* 2020;18:326. <https://doi.org/10.1186/s12955-020-01533-7>.
36. Nakamura-Thomas H, Morikawa M, Moriyama Y, Shirowa T, Kyougoku M, Razik K, et al. Japanese translation and cross-cultural validation of the Adult Social Care Outcomes Toolkit (ASCOT) in Japanese social service users. *Health Qual Life Outcomes.* 2019;17:59. <https://doi.org/10.1186/s12955-019-1128-7>.
37. Peasgood T, Mukuria C, Brazier J, Marten O, Kreimeier S, Luo N, et al. Developing a New Generic Health and Wellbeing Measure: Psychometric Survey Results for the EQ-HWB. *Value Health.* 2022;25:525–33. <https://doi.org/10.1016/j.jval.2021.11.1361>.
38. Raczek AE, Ware JE, Bjorner JB, Gandek B, Haley SM, Aaronson NK, et al. Comparison of Rasch and Summated Rating Scales Constructed from SF-36 Physical Functioning Items in Seven Countries: Results from the IQOLA Project. *J Clin Epidemiol.* 1998;51:1203–14. [https://doi.org/10.1016/S0895-4356\(98\)00112-7](https://doi.org/10.1016/S0895-4356(98)00112-7).
39. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25:3186–91. <https://doi.org/10.1097/00007632-200012150-00014>.
40. Yu X, Tam WWS, Wong PTK, Lam TH, Stewart SM. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Compr Psychiatry.* 2012;53:95–102. <https://doi.org/10.1016/j.comppsy.2010.11.002>.
41. Wu T-Y, Chie W-C, Kuo K-L, Wong W-K, Liu J-P, Chiu S-T, et al. Quality of life (QOL) among community dwelling older people in Taiwan measured by the CASP-19, an index to capture QOL in old age. *Arch Gerontol Geriatr.* 2013;57:143–50. <https://doi.org/10.1016/j.archger.2013.03.010>.

42. Wong EL, Cheung AW, Wong AY, Xu RH, Ramos-Goñi JM, Rivero-Arias O. Normative Profile of Health-Related Quality of Life for Hong Kong General Population Using Preference-Based Instrument EQ-5D-5L. *Value Health*. 2019;22:916–24. <https://doi.org/10.1016/j.jval.2019.02.014>.
43. Mahoney FI, Barthel DW. FUNCTIONAL EVALUATION: THE BARTHEL INDEX. *Md State Med J*. 1965;14:61–5.
44. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9:179–86. https://doi.org/10.1093/geront/9.3_part_1.179.
45. Liu T, Lu S, Leung DKY, Sze LCY, Kwok WW, Tang JYM, et al. Adapting the UCLA 3-item loneliness scale for community-based depressive symptoms screening interview among older Chinese: a cross-sectional study. *BMJ Open*. 2020;10:e041921. <https://doi.org/10.1136/bmjopen-2020-041921>.
46. Hutchinson C, Ratcliffe J, Cleland J, Milte R, Muller A, Ly M, et al. Further Evidence of Feasibility, Validity, and Reliability of Quality of Life-Aged Care Consumers: Evidence From Home-Based Care Settings. *Value Health*. 2024;27:1206–14. <https://doi.org/10.1016/j.jval.2024.05.003>.
47. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. 6th ed. Oxford: Oxford University Press; 2024. <https://doi.org/10.1093/med/9780192869487.001.0001>.
48. Kline P. Handbook of psychological testing. 2nd ed. London: Routledge; 2000. <https://doi.org/10.4324/9781315812274>.
49. Rand S, Towers A-M, Malley J, Silarova B. Exploring the structural characteristics of the adult social care outcomes toolkit (ASCOT) and ASCOT-Carer. *NIHR Open Res*. 2023;2:21. <https://doi.org/10.3310/nihropenres.13259.2>.
50. Hair JF, Babin BJ, Anderson RE, Black WC. Multivariate data analysis [Internet]. 8th ed. Andover, UK: Cengage; 2018. <http://ebookcentral.proquest.com/lib/cuhk-ebooks/detail.action?docID=6351360>. Accessed 24 Mar 2025.
51. Watkins MW. Exploratory factor analysis: A guide to best practice. *J Black Psychol*. 2018;44:219–46. <https://doi.org/10.1177/0095798418771807>.
52. Ashing-Giwa KT, Kim J, Tejero JS. Measuring quality of life among cervical cancer survivors: preliminary assessment of instrumentation validity in a cross-cultural study. *Qual Life Res*. 2008;17:147–57. <https://doi.org/10.1007/s11136-007-9276-3>.
53. Malley J. A Comparison of Options for Indicators from the Adult Social Care Survey (ASCS) [Internet]. Canterbury: Personal Social Services Research Unit, University of Kent; 2010. <https://www.pssru.ac.uk/pub/dp2736.pdf>. Report No.: 2736. Accessed 13 Aug 2026.
54. Chan SW, Chiu HF, Chien W, Thompson DR, Lam L. Quality of life in Chinese elderly people with depression. *Int J Geriatr Psychiatry*. 2006;21:312–8. <https://doi.org/10.1002/gps.1461>.
55. Nakamura CA, Mitchell PM, Peters TJ, Moreno-Agostino D, Araya R, Scazufca M, et al. A validation study of the EQ-5D-5L and ICEpop capability measure for older people among older individuals with depressive symptoms in Brazil. *Value Health Reg Issues*. 2022;30:91–9. <https://doi.org/10.1016/j.vhri.2021.11.005>.
56. van Leeuwen KM, Bosmans JE, Jansen APD, Hoogendijk EO, van Tulder MW, van der Horst HE, et al. Comparing Measurement Properties of the EQ-5D-3L, ICECAP-O, and ASCOT in Frail Older Adults. *Value Health*. 2015;18:35–43. <https://doi.org/10.1016/j.jval.2014.09.006>.
57. Hopkins WG. A scale of magnitudes for effect statistics [Internet]. A New View of Statistics. Will G Hopkins; 2002. <https://sportsci.org/resource/stats/effectmag.html>.
58. Cheng C, Cheng M. To validate the Chinese version of the 2Q and PHQ-9 questionnaires in Hong Kong Chinese patients. *Hong Kong Pract*. 2007;29:381–90.
59. Chin WY, Chan KT, Lam CL, Wong SY, Fong DY, Lo YY, et al. Detection and management of depression in adult primary care patients in Hong Kong: a cross-sectional survey conducted by a primary care practice-based research network. *BMC Fam Pract*. 2014;15:30. <https://doi.org/10.1186/1471-2296-15-30>.
60. Fiel Peres F. Effect sizes for nonparametric tests. *Biochem Med (Zagreb)*. 2026;36:010101. <https://doi.org/10.11613/bm.2026.010101>.
61. van Leeuwen KM, Bosmans JE, Jansen AP, Rand SE, Towers A-M, Smith N, et al. Dutch translation and cross-cultural validation of the Adult Social Care Outcomes Toolkit (ASCOT). *Health Qual Life Outcomes*. 2015;13:56. <https://doi.org/10.1186/s12955-015-0249-x>.
62. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc*. 1952;47:583–621. <https://doi.org/10.1080/01621459.1952.10483441>.
63. Dunn OJ. Multiple comparisons using rank sums. *Technometrics*. 1964;6:241–52. <https://doi.org/10.1080/00401706.1964.10490181>.
64. Fisher RA. Statistical methods for research workers. 5th ed. Edinburgh: Oliver and Boyd; 1934.
65. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>.
66. Mokkink LB, Elsmann EB, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Qual Life Res*. 2024;33:2929–39. <https://doi.org/10.1007/s11136-024-03761-6>.
67. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15:155–63. <https://doi.org/10.1016/j.jcm.2016.02.012>.

68. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. 3rd ed. Hoboken, NJ: Wiley; 2003. <https://doi.org/10.1002/0471445428>.
69. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. *BMC Med Res Methodol*. 2010;10:82. <https://doi.org/10.1186/1471-2288-10-82>.
70. Kuijlaars IA, van Emst M, van der Net J, Timmer MA, Fischer K. Assessing the test–retest reliability and smallest detectable change of the Haemophilia Activities List. *Haemophilia*. 2021;27:108–12. <https://doi.org/10.1111/hae.14226>.
71. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg*. 2018;126:1763–8. <https://doi.org/10.1213/ANE.0000000000002864>.
72. Rand S, Towers A-M, Allan S, Webster L, Palmer S, Carroll R, et al. Exploratory factor analysis and Rasch analysis to assess the structural validity of the Adult Social Care Outcomes Toolkit Proxy version (ASCOT-Proxy) completed by care home staff. *Qual Life Res*. 2024;33:1555–67. <https://doi.org/10.1007/s11136-024-03631-1>.
73. Mair P, Hatzinger R. Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *J Stat Softw*. 2007;20:1–20. <https://doi.org/10.18637/jss.v020.i09>.
74. Rosseel Y, lavaan. An R package for structural equation modeling. *J Stat Softw*. 2012;48:1–36. <https://doi.org/10.18637/jss.v048.i02>.
75. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1–55. <https://doi.org/10.1080/10705519909540118>.
76. Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes*. 2003;1:27. <https://doi.org/10.1186/1477-7525-1-27>.
77. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007;45:S22–31. <https://doi.org/10.1097/01.MLR.0000250483.85507.04>.
78. Christensen KB, Makransky G, Horton M. Critical Values for Yen’s Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Appl Psychol Meas*. 2017;41:178–94. <https://doi.org/10.1177/0146621616677520>.
79. Robinson M, Johnson AM, Walton DM, MacDermid JC. A comparison of the polytomous Rasch analysis output of RUMM2030 and R (ltm/eRm/TAM/lordif). *BMC Med Res Methodol*. 2019;19:36. <https://doi.org/10.1186/s12874-019-0680-5>.
80. Debelak R, Strobl C, Zeigenfuss MD. An Introduction to the Rasch Model with Examples. New York: Chapman and Hall/CRC; 2022. <https://doi.org/10.1201/9781315200620>.
81. Andrich D. Application of a psychometric rating model to ordered categories which are scored with successive integers. *Appl Psychol Meas*. 1978;2:581–94. <https://doi.org/10.1177/014662167800200413>.
82. Masters GN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47:149–74. <https://doi.org/10.1007/BF02296272>.
83. de Ayala RJ. The theory and practice of item response theory. New York, NY, US: Guilford Press; 2009. pp. pxx–448.
84. Duncan PW, Bode RK, Lai SM, Perera S. Glycine Antagonist in Neuroprotection Americas Investigators. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Arch Phys Med Rehabil*. 2003;84:950–63. [https://doi.org/10.1016/S0003-9993\(03\)00035-2](https://doi.org/10.1016/S0003-9993(03)00035-2).
85. Chang K-C, Wang J-D, Tang H-P, Cheng C-M, Lin C-Y. Psychometric evaluation, using Rasch analysis, of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: further item validation. *Health Qual Life Outcomes*. 2014;12:148. <https://doi.org/10.1186/s12955-014-0148-6>.
86. Pallant JF, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*. 2007;46:1–18. <https://doi.org/10.1348/014466506X96931>.
87. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: An introduction and illustrative example. *Int J Nurs Stud*. 2009;46:380–93. <https://doi.org/10.1016/j.ijnurstu.2008.10.007>.
88. Wright BD. Solving Measurement Problems with the Rasch Model. *J Educ Meas*. 1977;14:97–116. <https://doi.org/10.1111/j.1745-3984.1977.tb00031.x>.
89. Franchignoni F, Giordano A, Ferriero G. Rasch analysis of the short form 8-item Parkinson’s Disease Questionnaire (PDQ-8). *Qual Life Res*. 2008;17:541–8. <https://doi.org/10.1007/s11136-008-9341-6>.
90. Wright BD, Masters GN. Rating scale analysis. Chicago: MESA; 1982.
91. Gothwal VK, Bagga DK, Sumalini R. Rasch validation of the PHQ-9 in people with visual impairment in South India. *J Affect Disord*. 2014;167:171–7. <https://doi.org/10.1016/j.jad.2014.06.019>.
92. Chang F-H, Chen C-Y, Liang W-M, Lin M-R. Validating the quality of life after brain injury through rasch analysis. *J Head Trauma Rehabil*. 2017;32:E11–8. <https://doi.org/10.1097/HTR.0000000000000258>.
93. Linacre JM. Sample size and item calibration stability. *Rasch Meas Trans*. 1994;7(4):328.
94. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value Health*. 2015;18:25–34. <https://doi.org/10.1016/j.jval.2014.10.005>.

95. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch. paper? *Arthritis Rheum.* 2007;57:1358–62. <https://doi.org/10.1002/art.23108>.
96. Choi SW, Gibbons LE, Crane PK. Iordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *J Stat Softw.* 2011;39:1–30. <https://doi.org/10.18637/jss.v039.i08>.
97. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ.* 1995;310:170. <https://doi.org/10.1136/bmj.310.6973.170>.
98. Fayers PM, Machin D. Quality of life: the assessment, analysis and reporting of patient-reported outcomes [Internet]. 3rd ed. Chichester, UK: Wiley-Blackwell; 2016. <https://doi.org/10.1002/9781118758991>.
99. Zumbo BD. A Handbook on the Theory and Methods of Differential Item Functioning (DIF). Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defence; 1999.
100. Condon DM, Chapman R, Shaunfield S, Kallen MA, Beaumont JL, Eek D, et al. Does recall period matter? Comparing PROMIS® physical function with no recall, 24-hr recall, and 7-day recall. *Qual Life Res.* 2020;29:745–53. <https://doi.org/10.1007/s11136-019-02344-0>.
101. Martin M, Blaisdell B, Kwong JW, Bjorner JB. The Short-Form Headache Impact Test (HIT-6) was psychometrically equivalent in nine languages. *J Clin Epidemiol.* 2004;57:1271–8. <https://doi.org/10.1016/j.jclinepi.2004.05.004>.
102. Hair JF Jr, Black WC, Babin BJ, Anderson RE. Multivariate data analysis: A global perspective. 7th ed. Upper Saddle River, NJ: Pearson; 2010.
103. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol Assess.* 1995;7:286–99. <https://doi.org/10.1037/1040-3590.7.3.286>.
104. DeMars C. Item response theory. New York: Oxford University Press; 2010. <https://doi.org/10.1093/acprof:oso/9780195377033.001.0001>.
105. Mair P. Modern psychometrics with R. 1st ed. Cham: Springer; 2018. <https://doi.org/10.1007/978-3-319-93177-7>.
106. Rathbone A, Shaw S, Kumbhare D. ICC.Sample.Size: calculation of sample size and power for ICC [Internet]. Version 1.0. Comprehensive R Archive Network (CRAN); 2015. <https://doi.org/10.32614/CRAN.package.ICC.Sample.Size>.
107. Silarova B, Rand S, Towers A-M, Jones K. Feasibility, validity and reliability of the ASCOT-Proxy and ASCOT-Carer among unpaid carers of people living with dementia in England. *Health Qual Life Outcomes.* 2023;21:54. <https://doi.org/10.1186/s12955-023-02122-0>.
108. Towers A-M, Rand S, Collins G, Smith N, Palmer S, Cassell J. Measuring quality of life in care homes when self-report is challenging: the construct validity, structural characteristics and internal consistency of the mixed-methods adult social care outcomes toolkit. *Age Ageing.* 2023;52:afad168. <https://doi.org/10.1093/ageing/afad168>.
109. Khadka J, Hutchinson C, Milte R, Cleland J, Muller A, Bowes N, et al. Assessing feasibility, construct validity, and reliability of a new aged care-specific preference-based quality of life instrument: evidence from older Australians in residential aged care. *Health Qual Life Outcomes.* 2022;20:159. <https://doi.org/10.1186/s12955-022-02065-y>.
110. Lai J-S, Teresi J, Gershon R. Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Eval Health Prof.* 2005;28:283–94. <https://doi.org/10.1177/0163278705278276>.

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

Authors and Affiliations

Siyue Yu¹ · Richard Huan Xu² · Kailu Wang^{1,3} · Judy Chu Dik Sze^{1,3} ·
Annie Wai Ling Cheung^{1,3} · Elizabeth Welch⁴ · Nick Smith⁵ · Eliza Lai Yi Wong^{1,3}

✉ Eliza Lai Yi Wong
lywong@cuhk.edu.hk

- ¹ JC School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China
- ² Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong SAR, China
- ³ Center for Health Systems and Policy Research, JC School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

⁴ Personal Social Services Research Unit, University of Kent, Canterbury, UK

⁵ Centre for Health Service Studies, University of Kent, Canterbury, UK