
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202308.00009

Development of a Short-form Health Literacy Scale Based on Classical Test Theory and Item Response Theory (Postprint)

Authors: Sun Xiaonan, Chen Ke, Military Operations Research, Tang Jingqi, Wang Fei, Sun Xinying, He Miao, Wu Yibo, He Miao, Wu Yibo

Date: 2023-08-02T00:00:00+00:00

Abstract

Background Health literacy is closely associated with health status. Currently, domestic research tends to utilize multi-dimensional, multi-item health literacy assessment instruments, and there remains a lack of concise yet effective health literacy assessment tools.

Objective To simplify the health literacy scale and conduct psychometric testing within the Chinese population.

Methods Individuals aged 18 years and older were selected from the data of the “China Family Health Index Survey (2021)” as study participants. Based on inclusion and exclusion criteria, a total of 7449 data points were ultimately included and randomly divided into two sample sets, with Sample Set 1 comprising 3680 data points and Sample Set 2 comprising 3769 data points. Participants were surveyed using a general information questionnaire, the 12-item Health Literacy Scale (HLS-SF12), the Perceived Social Support Scale (PSSS), and the Family Health Scale-Short Form (FHS-SF). Classical Test Theory and the Mokken model from Item Response Theory were applied to screen the original items of the HLS-SF12, and confirmatory analyses including reliability and validity were conducted on the simplified scales.

Results Using Classical Test Theory and the Mokken model, one 9-item version of the scale (HLS-SF9) and one 4-item version (HLS-SF4) were simplified. Neither HLS-SF9 nor HLS-SF4 exhibited ceiling or floor effects. Their Cronbach’s α coefficients were 0.913 and 0.842, respectively, and their split-half reliabilities were 0.871 and 0.815, respectively. Exploratory factor analysis of HLS-SF4 extracted one common factor with a cumulative variance contribution rate of 67.813%, and the factor loadings of all items were >0.81 . Confirmatory factor analysis results for HLS-SF9 showed $\chi^2/df=10.844$, Goodness-of-Fit

Index (GFI)=0.985, Adjusted Goodness-of-Fit Index (AGFI)=0.971, Normed Fit Index (NFI)=0.986, Comparative Fit Index (CFI)=0.987, and Root Mean Square Error of Approximation (RMSEA)=0.051. Correlation analysis results indicated that both HLS-SF9 and HLS-SF4 were positively correlated with PSSS ($r=0.367, 0.292; P<0.001$) and positively correlated with FHS-SF ($r=0.340, 0.237; P<0.001$). The criterion validity of HLS-SF9 against HLS-SF12, expressed as Intraclass Correlation Coefficient (ICC) (95%CI), was 0.989 (0.988, 0.999), while the criterion validity ICC (95%CI) for HLS-SF4 was 0.892 (0.886, 0.899).

Conclusion The simplified health literacy scales demonstrate good reliability and validity and serve as effective tools for rapidly assessing health literacy in the Chinese population. Researchers may make targeted selections based on requirements for research precision (HLS-SF9) or assessment time (HLS-SF4).

Full Text

Development of a Short Version of the Health Literacy Scale Based on Classical Test Theory and Item Response Theory

SUN Xiaonan^{1,2}, CHEN Ke³, WU Yunchou⁴, TANG Jingqi⁵, WANG Fei⁶, SUN Xinying⁷, HE Miao¹, WU Yibo⁷

¹Department of Social Science and Humanities, Harbin Medical University, Harbin 150081, China

²Public Health College of Harbin Medical University, Harbin 150081, China

³School of Philosophy, Anhui University, Hefei 230039, China

⁴School of Psychology, Southwest University, Chongqing 400715, China

⁵School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

⁶State Key Laboratory of Cognition and Learning, Department of Psychology, Beijing Normal University, Beijing 100875, China

⁷School of Public Health, Peking University, Beijing 100875, China

Corresponding authors: HE Miao, Professor; E-mail: hemiao767@163.com; WU Yibo; E-mail: bjmuwuyibo@outlook.com

Abstract

Background Health literacy is closely associated with health status. Most domestic studies tend to use multidimensional and multi-item tools for assessing health literacy, lacking simple and effective assessment methods. **Objective** To simplify the health literacy scale and conduct psychometric testing within the Chinese population. **Methods** Adults aged 18 years and above were selected from the “China Family Health Index Survey (2021).” Based on inclusion and exclusion criteria, a total of 7,449 participants were selected and randomly divided into two sample sets, including 3,680 cases in sample set 1 and 3,769 cases in

sample set 2. The general information questionnaire, short-form health literacy questionnaire (HLS-SF12), Perceived Social Support Scale (PSSS), and Family Health Scale Short-Form (FHS-SF) were administered to the respondents. Classical test theory (CTT) and the Mokken model in item response theory (IRT) were used to screen the original items, and validation analyses such as reliability and validity of the simplified scales were also conducted. **Results** A 9-item version (HLS-SF9) and a 4-item version (HLS-SF4) were simplified by using CTT and the Mokken model, respectively. Both the HLS-SF9 and the HLS-SF4 had no ceiling effect or floor effect, and the Cronbach' s α coefficients for both were 0.913 and 0.842, with split-half reliabilities of 0.871 and 0.815, respectively. The exploratory factor analysis of HLS-SF4 revealed one common factor, accounting for 67.813% of the cumulative variance, with factor loadings exceeding 0.81 for each item. The confirmatory factor analysis of HLS-SF9 showed that χ^2/df was 10.844, goodness of fit index (GFI) was 0.985, adjusted goodness of fit index (AGFI) was 0.971, normative fit index (NFI) was 0.986, comparative fit index (CFI) was 0.987, and root mean squared error of approximation (RMSEA) was 0.051. The correlation analysis demonstrated positive correlation of HLS-SF9 and HLS-SF4 with PSSS ($r=0.367, 0.292, P<0.001$), as well as FHS-SF ($r=0.340, 0.237, P<0.001$), respectively. The intraclass correlation coefficients (ICC) (95%CI) for the criterion validity of HLS-SF9 against HLS-SF12 was 0.989 (0.988, 0.999), while for HLS-SF4 against HLS-SF12 was 0.892 (0.886, 0.899). **Conclusion** The simplified health literacy scales have good reliability and validity, which are reliable and effective tools for assessing the health literacy of the Chinese population. Researchers can choose the scale according to the needs of research precision (HLS-SF9) or assessment time (HLS-SF4).

Keywords Health literacy; Item selection; Classical test theory; Mokken model; Short-form health literacy survey questionnaire

Health literacy refers to an individual' s ability to obtain, understand, and use health information to maintain and promote their own health [1]. The World Health Organization emphasizes that health literacy, as an indicator of cognitive ability and social skills [2], is an important determinant of health [3]. Limited health literacy can make it difficult for the public to fully recognize and understand diseases, hindering efficient and adequate utilization of medical resources. This particularly affects self-management of chronic diseases, leading to poor medication adherence and even increased hospitalization and mortality rates [4-6]. China' s "Healthy China 2030" Planning Outline emphasizes that residents' health literacy status is a core component of strategic goals, and improving health literacy should become an important element of health education, patient management, and health promotion [7]. This requires individuals to fully understand and use health information to adequately manage their own health issues and reduce social costs. Therefore, an objective, comprehensive, and accurate tool for assessing health literacy is a prerequisite for conducting

such research.

The 12-item short-form health literacy survey questionnaire (HLS-SF12) supports the original structure of HLS-EU-Q47 with good reliability and validity. Chinese scholars mostly use the “National Resident Health Literacy Monitoring Questionnaire” formulated by the National Health Commission. With the deepening of health literacy research, scholars have developed health literacy assessment tools for special populations [11-13]. Currently, research projects tend to use multidimensional and multi-item assessment tools, striving for comprehensive evaluation of respondents’ clinical or psychological characteristics. However, lengthy instruments have certain drawbacks: too many questions lead to longer response times, reduced patience, decreased response seriousness, and compromised authenticity and reliability of questionnaires. Simultaneously, they may cause respondents to feel their privacy is invaded, increasing psychological burden [14]. Short-form scales can largely avoid these disadvantages while reducing questionnaire completion time, facilitating broader respondent populations and application fields. Additionally, as health literacy has gained attention from scholars worldwide, the development of tools for measuring public health literacy has become active. Commonly used instruments include the Test of Functional Health Literacy in Adults (TOFHLA) and the European Health Literacy Survey Questionnaire (HLS-EU-Q) [8-9]. Building on these studies, Duong et al. [10] developed the HLS-SF12, suitable for assessing public health literacy in Asian countries, based on HLS-EU-Q47. Short health literacy assessment tools can be incorporated into patient visit evaluations and other assessment questionnaires to quickly screen groups with limited health literacy, facilitating targeted health education and evaluating intervention effects [8,15].

Considering the lack of simple assessment tools in China’s health literacy field, this study attempts to simplify HLS-SF12, which will help enable simple and accurate assessment of health literacy in larger populations or clinical environments, providing a reference basis for future intervention research.

1.1 Study Subjects

The “China Family Health Index Survey (2021)” was conducted from July 10 to September 15, 2021, using a multi-stage sampling method. The capitals of 23 provinces and 5 autonomous regions, plus 4 municipalities directly under the central government, were directly included in the survey. Using a random number table method, 2-6 cities were selected from non-capital prefecture-level administrative regions in each province and autonomous region, totaling 120 cities. Each city recruited at least one investigator or survey team. Investigators were required to use data from the “2021 Seventh National Population Census” to ensure that the obtained sample’s gender, age, and urban-rural distribution basically matched population characteristics. Inclusion criteria for the “China Family Health Index Survey (2021)” were: (1) age ≥ 12 years; (2) Chinese nationality; (3) permanent Chinese resident (annual out-of-town time ≤ 1 month); (4) voluntary participation with informed consent; (5) ability to

complete the online questionnaire independently or with investigator assistance; (6) understanding the meaning expressed in each questionnaire item. Exclusion criteria were: (1) unconsciousness or mental abnormality; (2) participation in other similar research projects; (3) unwillingness to cooperate. After questionnaire collection, two persons conducted logic checks and data screening back-to-back. This study was approved by the Ethics Committee of Jinan University (JNUKY-2021-018). The “China Family Health Index Survey (2021)” surveyed 11,668 residents, with 11,031 valid questionnaires recovered, yielding an effective recovery rate of 94.54%. From the data, individuals aged ≥ 18 years were selected as subjects for this study, with 7,449 cases ultimately included and randomly divided into two sample sets: sample set 1 with 3,680 cases and sample set 2 with 3,769 cases.

1.2 Methods

1.2.1 General Information Questionnaire

Developed by the researchers, the questionnaire included content on respondents' gender, age, ethnicity, household registration type, residence type, highest education level, marital status, and monthly household income per capita.

1.2.2 HLS-SF12

The HLS-SF12, developed by Duong et al. [10] for public health literacy measurement, consists of 3 dimensions: health care, disease prevention, and health promotion, with 12 items total using a 4-point scale (1=very difficult, 2=difficult, 3=easy, 4=very easy). The standardized Health Literacy Index (HL index) is calculated using the formula: $HL\ index = (\text{mean} - 1) \times (50/3)$, where the mean is the average of all items for each individual, 1 is the minimum possible value of the mean (making the minimum index value 0), 3 is the range of the mean, and 50 is the maximum index value. Duong reported HLS-SF12' s Cronbach' s α coefficient >0.70 , with subscale coefficients of 0.49-0.72 for health care, 0.64-0.77 for disease prevention, and 0.59-0.81 for health promotion, indicating good internal consistency. With authorization from the original author, the Chinese version of HLS-SF12 [16] was used in this study. In this research, the scale' s Cronbach' s α coefficient was 0.932 in dataset 1, 0.933 in dataset 2, and 0.932 in the total dataset.

1.2.3 Perceived Social Support Scale (PSSS)

Developed by Zimet et al. [17] and revised by Jiang Qianjin, the PSSS consists of 3 dimensions: family support, friend support, and other support, with 4 items in each dimension (12 items total). Scale options range from “strongly disagree” to “strongly agree” scored 1-7, with higher scores indicating richer perceived social support. In this study, the scale' s Cronbach' s α coefficient was 0.947 in dataset 2 and 0.948 in the total dataset.

1.2.4 Family Health Scale Short-Form (FHS-SF)

The FHS-SF was developed by Crandall et al. [18] to assess family health function and was translated into Chinese by Wang et al. [19]. The scale has 4 dimensions with 10 items total. Items range from “strongly disagree” to “strongly agree” (1-5 points), with items 6, 9, and 10 reverse-scored. Higher total scores indicate better family health levels. In this study, the scale’s Cronbach’s α coefficient was 0.845 in dataset 2 and 0.846 in the total dataset.

1.3 Simplification Methods

1.3.1 Classical Test Theory (CTT) Simplification Methods

Four commonly used methods in CTT item analysis were applied to screen the original scale items: inter-item residual correlation method, correlation coefficient method, corrected item-total correlation (CITC) method, and independent samples t-test method.

- (1) The inter-item residual correlation method calculates the sum of residual correlations between each item and all other items, retaining items with the smallest inter-item residual correlations, expressed as the sum of MI values of item residuals [20].
- (2) The correlation coefficient method calculates Pearson correlation coefficients between each item and the total scale score, selecting items with larger absolute correlation values that are statistically significant [21].
- (3) The CITC method screens items based on internal consistency by calculating Cronbach’s α coefficients for the total scale or individual dimensions and comparing changes after deleting an item. If deleting an item substantially increases the Cronbach’s α coefficient, the item should be removed because its presence reduces internal consistency; otherwise, it should be retained [22].
- (4) The independent samples t-test method first arranges total scale scores from high to low, with the high-scoring group being the top 27% and the low-scoring group the bottom 27%. Independent samples t-tests are then conducted. If results show no significant difference in mean scores for an item between high and low groups, that item should be deleted [23].

1.3.2 Item Response Theory (IRT) Simplification Method—Mokken Model

Analysis was conducted using the Mokken model [24] in item response theory. The Mokken model belongs to nonparametric item response theory, which provides a more adaptable and flexible framework for real-world scenarios compared to parametric item response theory and is more suitable for short scales [25]. The R package “Mokken” was used to analyze the full health literacy scale, including

unidimensionality tests, local independence tests, and monotonicity tests for all items.

- (1) The Automated Item Selection Procedure (AISP) was used to test scale unidimensionality, implemented from $c=0$ to $c=0.55$ in steps of 0.05. When the c value increases, if the test is unidimensional, three stages may occur: most or all items merge into one scale; one smaller scale forms; or one or several small scales form while many items are deleted [26].
- (2) Homogeneity coefficients were used to check whether item settings were reasonable. Homogeneity coefficients are divided into three types: item pairs (H_{ij}), items (H_i), and scale (H_s). Higher H values indicate more accurate ranking of respondents' latent traits by total test scores. Based on experience, Mokken believed H_{ij} must be >0 , and H_i and H should not be <0.3 . More specific H value criteria are: when $H < 0.3$, the scale is unqualified; when $0.3 \leq H < 0.4$, the scale's accuracy is weak; when $0.4 \leq H < 0.5$, accuracy is moderate; when $H \geq 0.5$, accuracy is strong [27].
- (3) Local independence tests were completed using conditional association procedures, with three conditional association indices $W(1)$, $W(2)$, and $W(3)$ used for testing. Flagged items were deleted one by one according to the principle that items with the most W flags were removed. If items had the same number of flags, the item with the smaller homogeneity coefficient H_i was deleted [28].
- (4) Monotonicity of each item was tested. The minimum violation coefficient ($\#vi/\#ac$), significance ($\#zsig$), and Crit values were used as evaluation criteria. When all three indicators equal 0, monotonicity assumptions are met. However, in practical applications, a minimum violation coefficient <0.3 is acceptable [29], significance <1.96 is acceptable [30]. When $Crit > 80$, monotonicity assumptions are violated; when $40 \leq Crit \leq 80$, deletion should be considered based on item content and scale usage; when $Crit < 40$, the item basically satisfies monotonicity, and occasional violations can be considered sampling errors [31].

1.4 Statistical Methods

Data processing was performed using SPSS 24.0, AMOS 24.0, and R 4.2.1, employing descriptive statistical analysis, confirmatory factor analysis, classical test theory item reduction, Mokken model item reduction, and validation analysis. To ensure the validity of the short-form scales and avoid reduced research power caused by insufficient sample size, the minimum required sample size was calculated [32]: assuming an intra-class correlation coefficient (ICC) of 0.90, type I error probability α of 0.05, achieving 95% statistical power required 223 participants. This study collected 7,449 valid data points (including both item screening and validation datasets), indicating sufficient sample size for subsequent data analysis.

2.2 Psychometric Testing of the Full Scale

Based on dataset 1 data, confirmatory factor analysis was conducted on HLS-SF12' s psychometric indicators, as shown in [Figure 1: see original paper]. Results showed normative fit index (NFI)=0.960, goodness of fit index (GFI)=0.957, adjusted goodness of fit index (AGFI)=0.934, comparative fit index (CFI)=0.962, and root mean squared error of approximation (RMSEA)=0.068. These fit indices indicated good model fit for the original scale. Internal consistency coefficient calculation yielded Cronbach' s $\alpha=0.932$, indicating good scale reliability.

2.3 Exploratory Analysis

Descriptive statistics were first performed on the sociodemographic information from dataset 1 and dataset 2, showing the number and percentage of each variable category. After using dataset 1 to verify HLS-SF12' s psychometric indicators, items were simplified using CTT and IRT methods respectively, yielding HLS-SF9 and HLS-SF4. Validation analyses were then conducted based on dataset 2 data.

2.3.1 Item Reduction Based on Classical Test Theory

First, the inter-item residual correlation method was used for item analysis, with MI value threshold using the default value. Results showed that items with the largest residual MI values in each dimension were items 3, 5, and 11, indicating these items had the weakest explanatory power for their dimensions among all items, and thus were considered for deletion.

Second, the correlation coefficient method was used for item analysis. Results showed that correlation coefficients between the full health literacy scale and each item were all >0.71 ($r=0.716-0.797$), indicating good consistency between these items and the scale, and all were considered for retention.

Third, CITC was used to analyze the scale. Results showed that Cronbach' s α coefficients after deleting each item ranged from 0.924-0.928, with internal consistency coefficients decreasing after item deletion. Therefore, no items needed to be deleted from the scale.

Finally, using the highest 27% ($\$ 37.5points$) and lowest 27% ($\$ 30.556$ points) of health literacy HL index scores to define high and low groups, independent samples t-tests were conducted. Results showed significant differences in scores for all items between high and low groups ($P<0.001$), indicating no items needed to be deleted.

In summary, using four classical test theory methods to streamline HLS-SF12, results indicated that items 3, 5, and 11 were considered for deletion in the inter-item residual correlation method. After deleting these three items, a 9-item three-dimensional short-form scale was formed. The health literacy scale

simplified based on classical test theory (HLS-SF9) includes 9 items: items 1, 2, 4, 6, 7, 8, 9, 10, and 12. Detailed item analysis results are shown in .

2.3.2 Item Reduction Based on Mokken Model

First, Mokken model analysis was performed on HLS-SF12. AISP was used to test scale unidimensionality, implemented from $c=0$ to $c=0.55$ with step size 0.05. Results showed that when c was set at 0-0.55, only one dimension could be obtained, with all items in that dimension.

Then, various homogeneity coefficients were calculated. Results showed that in this study, all H_{ij} values were >0.43 and all H_i values were >0.53 (details in), with $H=0.609$. This indicated that homogeneity coefficients could not be used to delete items.

Local independence testing was then conducted through conditional association analysis. In the first round, index $W(1)$ showed item 11 had 6 flags, items 2 and 8 each had 4 flags, and item 10 had 1 flag, so item 11 was deleted first. In subsequent rounds, based on results of indices $W(1)$, $W(2)$, and $W(3)$, items 8, 2, 12, 6, 9, 4, and 10 were sequentially deleted. After conditional association analysis, 4 items remained: items 1, 3, 5, and 7.

Monotonicity tests were then conducted on these 4 items. Items 1 and 7 did not violate monotonicity. Item 3 had a minimum violation coefficient of 0.02, significance of 1, and Crit value of 13, while item 5 had a minimum violation coefficient of 0.02, significance of 1, and Crit value of 19, all within acceptable ranges, and thus were retained.

In summary, the health literacy scale simplified based on Mokken model analysis (HLS-SF4) includes 4 items: items 1, 3, 5, and 7. See for details.

2.4 Validation

2.4.1 Ceiling and Floor Effect Testing

The reliability of HLS-SF9 and HLS-SF4 can be enhanced by minimal floor/ceiling effects. shows the ceiling and floor effects in dataset 2, neither exceeding 15%, indicating that respondents with the lowest or highest scores can be distinguished from each other, facilitating reliability measurement.

2.4.2 Reliability Validation

Using dataset 2 to test reliability of the two simplified health literacy scales, analysis showed Cronbach's α coefficients of 0.913 and 0.842, and split-half reliabilities of 0.871 and 0.815 for HLS-SF9 and HLS-SF4, respectively. Reliability after deleting each item was \$ \$0.910 and \$ \$0.810, indicating good reliability indices.

2.4.3 Validity Testing

2.4.3.1 Structural Validity Bartlett' s sphericity test and KMO measure were conducted on HLS-SF4 in dataset 2. HLS-SF4' s Bartlett sphericity test value was 5915.883 ($P<0.01$) and KMO measure was 0.807, indicating suitability for factor analysis. Exploratory factor analysis extracted one factor with eigenvalue >1 , verifying its unidimensionality from a CTT perspective, with cumulative variance contribution rate of 67.813% and factor loadings >0.81 for all items.

Confirmatory factor analysis was performed on HLS-SF9' s 9 items ([Figure 2: see original paper]). HLS-SF9' s confirmatory factor analysis model fit indices showed $\chi^2/df=10.844$, GFI=0.985, AGFI=0.971, NFI=0.986, CFI=0.987, and RMSEA=0.051. Except for χ^2/df , all reached ideal criteria. It should be noted that χ^2/df eliminates the influence of degrees of freedom but not sample size. With a sample size of 3,769 (large sample), research indicates that when sample size is large, the overall model fit chi-square value increases significantly with sample size, so other important indices can be considered while this index can be ignored. Therefore, the three-dimensional 9-item HLS-SF9 showed good model fit.

2.4.3.2 Empirical Validity Previous studies have shown that health literacy is significantly correlated with perceived social support and family health [36-38]. This study used dataset 2 to conduct correlation analysis simultaneously among PSSS, FHS-SF, HLS-SF12, HLS-SF9, and HLS-SF4. Results showed HLS-SF12 was positively correlated with PSSS ($r=0.361$, $P<0.001$) and FHS-SF ($r=0.329$, $P<0.001$). HLS-SF9 and HLS-SF4 were positively correlated with PSSS ($r=0.367$, $P<0.001$; $r=0.292$, $P<0.001$) and FHS-SF ($r=0.340$, $P<0.001$; $r=0.237$, $P<0.001$), indicating good empirical validity for both HLS-SF9 and HLS-SF4.

2.4.4 Measurement Content Consistency

Based on dataset 2, this study analyzed criterion validity of the two simplified health literacy scales. The ICC (95%CI) for HLS-SF9 against HLS-SF12 was 0.989 (0.988, 0.999), while for HLS-SF4 against HLS-SF12 was 0.892 (0.886, 0.899), indicating good to excellent criterion validity and high consistency between the simplified scales and full version, though HLS-SF9 showed higher consistency than HLS-SF4.

3.1 Rationality of Scale Simplification Process and Results

HLS-SF12 was simplified using principal component analysis. Validation results from six Asian countries/regions [10] showed it adequately reflected the theoretical structure of health literacy and demonstrated valid differences among health literacy level subgroups. Criterion testing showed HLS-SF12' s explanatory effectiveness for the Asian health literacy scale HLS-EU-Q47 was higher than the

European scale HLS-EU-Q16. The HLS-SF12 scale has been applied by foreign scholars in studies of patients in general outpatient, orthopedic, and traditional Chinese medicine departments [39], and validated in Vietnamese farmers, health workers during COVID-19, and outpatients [40-42], all demonstrating good reliability and validity and cross-cultural, regional, and social group applicability as an effective measurement tool for multi-group health literacy.

Previous simplification studies commonly used classical test theory and Mokken model frameworks. On one hand, CTT is a historically long-standing, widely applied, and well-known psychometric theory [43]. It represents observed test scores as the sum of true scores and error scores, and based on its assumptions, has derived over a dozen parameter calculation formulas including reliability, validity, item difficulty, and discrimination through decades of practice, establishing a complete testing methodology system and standardized procedures that make the entire testing process more objective and scientific. Additionally, CTT has a relatively understandable mathematical model, parameter concepts, and estimation methods. Its advocated standardization techniques effectively control measurement errors, and more importantly, its theoretical and methodological system is relatively complete with weak prerequisite assumptions that are easily satisfied in practical work [44]. For example, Yu Binbin et al. [45] used extreme value method, correlation coefficient method, and CITC based on CTT to simplify the Critical Thinking Disposition Scale to 28 items, with reliability and validity tests showing strong usability of the simplified scale.

On the other hand, IRT models have been proven to have more advantages, using nonlinear models to establish nonlinear relationships between respondents' item responses and their latent traits, which better fits practical testing situations [46]. The Mokken model is a type of nonparametric item response theory model with nonparametric characteristics that also follows IRT's basic principles, compensating for shortcomings of parametric item response theory models [47]. After Mokken model analysis of all items, items not meeting theoretical assumptions can be deleted or modified to further improve scale quality [48]. Wang et al. [49] and Wu et al. [50] used the Mokken model to simplify the New General Self-Efficacy Scale and Perceived Social Support Scale respectively, showing good reliability and validity for practical application.

Before simplification, this study used confirmatory factor analysis to verify the structural validity of the original scale in dataset 1, showing good structural validity of HLS-SF12. In the scale simplification process based on classical test theory, this study also simplified the health literacy scale according to four commonly used CTT item analysis methods. One item in each dimension was considered for deletion in the inter-item residual correlation method; after deleting these items, a 9-item short-form scale (HLS-SF9) was formed.

In the scale simplification process based on nonparametric item response theory, this study conducted Mokken analysis on the full scale. First, unidimensionality testing of the full scale showed only one dimension could be obtained, with good homogeneity coefficients. However, in local independence testing, 8 items

with conditional associations were sequentially deleted, followed by monotonicity testing with no additional items deleted, ultimately yielding a 4-item short-form scale (HLS-SF4).

Simplification work based on classical test theory and Mokken model yielded two streamlined versions, HLS-SF9 and HLS-SF4. Using dataset 2 data for validation analysis, ceiling and floor effect analysis showed both effects were low, allowing good discrimination between high and low scorers and facilitating further reliability analysis. Reliability tests showed both short-form scales had good reliability, though HLS-SF9's reliability indices were superior to HLS-SF4's. In structural validity testing, Bartlett's sphericity test and KMO measure were first conducted, confirming HLS-SF4's suitability for factor analysis. Exploratory factor analysis was then used to verify HLS-SF4's simplified structure, extracting one common factor with eigenvalue >1 through principal component analysis, consistent with the unidimensional structure premise tested by AISP in Mokken model analysis. Confirmatory factor analysis was performed on HLS-SF9, with results showing excellent fit indices for its three-dimensional model, indicating the 9-item short-form scale's division into three dimensions was reasonable. Notably, the two short-form scales used two different structural validity verification methods for two reasons: first, classical test theory simplification did not involve dimension reduction, with dimensions determined and consistent with the original scale; second, Mokken model simplification broke the original dimensions, merging them into a single dimension, with retained items only involving the first two dimensions.

Empirical validity test results showed both HLS-SF9 and HLS-SF4 had good empirical validity. ICC indices comparing the two simplified scales with the full version showed high consistency in measured content, though HLS-SF9 was higher than HLS-SF4. In summary, HLS-SF9 retains the original scale's factor structure with superior reliability and empirical validity and higher consistency with the original scale's measured content, while HLS-SF4 has fewer items and higher factor loadings. Therefore, retaining both simplified versions is reasonable. Researchers can choose more targeted assessment scales according to their research circumstances: if the primary purpose is to obtain more precise measurement results, HLS-SF9 can be used; if the primary purpose is to shorten overall questionnaire response time (such as in large cross-sectional survey projects where precision can be sacrificed for a more streamlined questionnaire), HLS-SF4 is more appropriate.

In infectious or non-infectious disease prevention and control, health literacy is an important factor that cannot be ignored. For example, in the COVID-19 infodemic, health literacy became a key tool for the public to identify "misinformation" [51]. This study used large-sample national research data to potentially reduce bias from regional differences and improve research generalizability [52], randomly generating two sample groups for mutual validation of results. Compared with commonly used domestic assessment questionnaires, the simplified health literacy scale has fewer items, shorter response time, and lower response

difficulty, making it more suitable for assessing all age groups or use in comprehensive questionnaires.

3.2 Limitations and Future Research

This study used two theoretical frameworks, both strictly adhering to scale simplification principles, but certain limitations remain. For example, this study selected a large-sample national cross-sectional study as the data source. Future longitudinal studies are needed to validate the effectiveness and stability of the short-form scales. In multi-field practical application, more adaptive validation and adjustment are also needed.

In summary, this study used classical test theory and the Mokken model to screen HLS-SF12 items. The simplified 9-item and 4-item health literacy scales demonstrate good reliability and validity in China's general population and can serve as reliable and streamlined tools for assessing health literacy. Researchers can make targeted choices based on research precision needs (HLS-SF9) or assessment time constraints (HLS-SF4).

Author Contributions: Sun Xiaonan and Chen Ke conceived and designed the study and wrote the manuscript. Sun Xiaonan, Chen Ke, Wu Yunchou, Wang Fei, and Sun Xinying implemented the study and conducted feasibility analysis. Sun Xiaonan and Tang Jingqi collected and organized data. Sun Xiaonan, Chen Ke, and Wu Yunchou performed statistical processing, results analysis, and interpretation. Sun Xiaonan, Chen Ke, and Tang Jingqi revised the manuscript. He Miao and Wu Yibo are responsible for the overall article and supervision.

Conflict of Interest: This article has no conflicts of interest.

References [1] Health literacy: report of the council on scientific affairs. ad hoc committee on health literacy for the council on scientific affairs, American medical association[J]. JAMA, 1999, 281(6): 552-557. [2] BIE R. Study on the Evaluation Index System of Residents' Health Literacy Based on Public Health Emergencies[D]. Nanjing: Nanjing University of Chinese Medicine, 2012. [3] SHEIHAM A. Closing the gap in a generation: health equity through action on the social determinants of health. A report of the WHO Commission on Social Determinants of Health (CSDH) 2008[J]. Community Dent Health, 2009, 26(1): 2-3. [4] MUIR K W, SANTIAGO-TURLA C, STINNETT S S, et al. Health literacy and adherence to glaucoma therapy[J]. Am J Ophthalmol, 2006, 142(2): 223-226. DOI: 10.1016/j.ajo.2006.03.018. [5] BAKER D W, GAZMARARIAN J A, WILLIAMS M V, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees[J]. Am J Public Health, 2002, 92(8): 1278-1283. DOI: 10.2105/ajph.92.8.1278. [6] SUDORE R L, YAFFE K, SATTERFIELD S, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study[J]. J Gen Intern Med, 2006, 21(8): 806-812. DOI: 10.1111/j.1525-1497.2006.00539.x. [7] NIE X Q, LI Y H, LI C N, et al. The association between health literacy and self-rated health among resi-

dents of China aged 15-69 years[J]. *Am J Prev Med*, 2021, 60(4): 569-578. DOI: 10.1016/j.amepre.2020.05.032. [8] LI T, YI Q Y, SUN M. Research progress on health literacy assessment tools[J]. *Nursing Journal of Chinese People' s Liberation Army*, 2015, 32(18): 29-32. DOI: 10.3969/j.issn.1008-9993.2015.18.009. [9] OUYANG Y, WANG X H, YANG C, et al. Research progress on health literacy assessment tools for the elderly[J]. *Nursing Journal of Chinese People' s Liberation Army*, 2018, 35(2): 39-43, 48. DOI: 10.3969/j.issn.1008-9993.2018.02.008. [10] DUONG T V, ARINGAZINA A, KAYUPOVA G, et al. Development and validation of a new short-form health literacy instrument (HLS-SF12) for the general public in Six Asian countries[J]. *Health Lit Res Pract*, 2019, 3(2): e91-e102. DOI: 10.3928/24748307-20190225-01. [11] OU G Z, ZHANG S Y, GUAN J H. Construction of health information literacy evaluation index system for diabetic patients using Delphi method[J]. *Chinese Journal of Health Education*, 2014, 30(2): 107-110. DOI: 10.3969/j.issn.1002-9982.2014.02.003. [12] WANG G L, JIANG X J, ZHANG Y Q. Construction of health literacy evaluation index system for chronic disease patients using Delphi method[J]. *Chinese Journal of Public Health*, 2018, 34(1): 71-74. DOI: 10.11847/zgggws1113588. [13] JIN F F. Development and Evaluation of Health Literacy Scale for Chinese Rural Residents[D]. Beijing: Chinese Center for Disease Control and Prevention, 2018. [14] LU T Y, WU D R. Overview of item screening methods in short-form scale research[C]// 7th Symposium on Evidence-Based Medicine Methods in Traditional Chinese Medicine/Integrated Traditional Chinese and Western Medicine, Urumqi, 2013. [15] CAWTHON C, MION L C, WILLENS D E, et al. Implementing routine health literacy assessment in hospital and primary care patients[J]. *Jt Comm J Qual Patient Saf*, 2014, 40(2): 68-76. DOI: 10.1016/s1553-7250(14)40008-4. [16] SUN X N, LV K, WANG F, et al. Validity and reliability of the Chinese version of the Health Literacy Scale Short-Form in the Chinese population[J]. *BMC Public Health*, 2023, 23(1): 385. DOI: 10.1186/s12889-023-15237-2. [17] ZIMET G D, POWELL S S, FARLEY G K, et al. Psychometric characteristics of the multidimensional scale of perceived social support[J]. *J Pers Assess*, 1990, 55(3/4): 610-617. DOI: 10.1080/00223891.1990.9674095. [18] CRANDALL A, WEISS-LAXER N S, BROADBENT E, et al. The family health scale: reliability and validity of a short- and long-form[J]. *Front Public Health*, 2020, 8: 587125. DOI: 10.3389/fpubh.2020.587125. [19] WANG F, WU Y C, SUN X N, et al. Reliability and validity of the Chinese version of a short form of the family health scale[J]. *BMC Prim Care*, 2022, 23(1): 108. DOI: 10.1186/s12875-022-01690-2. [20] CHAI X Y, LI X Y, CAO J, et al. Development of the Chinese Positive Youth Development Scale (Short Form): Based on a large sample data[J]. *Studies of Psychology and Behavior*, 2020, 18(5): 631-637. [21] OLATUNJI B O, WILLIAMS N L, TOLIN D F, et al. The Disgust Scale: item analysis, factor structure, and suggestions for refinement[J]. *Psychol Assess*, 2007, 19(3): 281-297. DOI: 10.1037/1040-3590.19.3.281. [22] JAARSMAT, ARESTEDT K F, MÅRTENSSON J, et al. The European Heart Failure Self-care Behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument[J]. *Eur J Heart Fail*, 2009, 11(1): 99-105. DOI: 10.1093/eurjhf/hfn007. [23] COSKUN Y D, DEMIREL M. Lifelong learning

tendency scale: the study of validity and reliability[J]. *Procedia Soc Behav Sci*, 2010, 5: 2343-2350. DOI: 10.1016/j.sbspro.2010.07.461. [24] MOKKEN R J. A theory and procedure of scale analysis: with applications in political research[M]. New York: De Gruyter Mouton, 1971. [25] JUNKER B W, SIJTSMA K. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory[J]. *Appl Psychol Meas*, 2001, 25(3): 258-272. DOI: 10.1177/01466210122032064. [26] HEMKER B T, SIJTSMA K, MOLENAAR I W. Selection of unidimensional scales from a multidimensional item bank in the polytomous mokken I RT model[J]. *Appl Psychol Meas*, 1995, 19(4): 337-352. DOI: 10.1177/014662169501900404. [27] ZHANG J. Application and evaluation of nonparametric item response theory in dimensionality analysis[J]. *Psychological Exploration*, 2010, 30(3): 80-83. DOI: 10.3969/j.issn.1003-5184.2010.03.016. [28] STRAAT J H, VAN DER ARK L A, SIJTSMA K. Using conditional association to identify locally independent item sets[J]. *Methodol*, 2016, 12(4): 117-123. [29] SIJTSMA K, MEIJER R R, ANDRIES VAN DER ARK L. Mokken scale analysis as time goes by: an update for scaling practitioners[J]. *Pers Individ Differ*, 2011, 50(1): 31-37. DOI: 10.1016/j.paid.2010.08.016. [30] PENG W. A Study on the Measurement, Current Status and Relationship with Procrastination of Self-Criticism Rumination among Higher Vocational Students[D]. Guiyang: Guizhou Normal University, 2020. [31] MOLENAAR I, SIJTSMA K. MPS5 for Windows. A program for Mokken scale analysis for polytomous items[Z]. 2000. <https://api.semanticscholar.org/CorpusID:142645170>. [32] STEVENS M L, CROWLEY P, GARDE A H, et al. Validation of a short-form version of the Danish need for recovery scale against the full scale[J]. *Int J Environ Res Public Health*, 2019, 16(13): 2334. DOI: 10.3390/ijerph16132334. [33] TERWEE C B, BOT S D, DE BOER M R, et al. Quality criteria were proposed for measurement properties of health status questionnaires[J]. *J Clin Epidemiol*, 2007, 60(1): 34-42. DOI: 10.1016/j.jclinepi.2006.03.012. [34] MESSNER E M, TERHORST Y, BARKE A, et al. The German version of the mobile app rating scale (MARS-G): development and validation study[J]. *JMIR Mhealth Uhealth*, 2020, 8(3): e14479. DOI: 10.2196/14479. [35] KOO T K, LI M Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research[J]. *J Chiropr Med*, 2016, 15(2): 155-163. DOI: 10.1016/j.jcm.2016.02.012. [36] ZHANG J, QI Y L, MA Y, et al. Relationship between parenting style, perceived social support and health literacy of college students[J]. *Journal of Bengbu Medical College*, 2017, 42(4): 483-486, 490. DOI: 10.13898/j.cnki.issn.1000-2200.2017.04.021. [37] ZHANG W, LIU Y, ZHAO Y N, et al. Study on the influence of family function and community health services on infectious disease health literacy of community residents[J]. *Journal of Nursing Management*, 2022, 22(5): 331-335. [38] LI H M, SUN J, YAO Q, et al. Family health: concept, measurement and current status in China[J]. *Chinese Journal of Health Policy*, 2022, 15(4): 1-7. DOI: 10.3969/j.issn.1674-2982.2022.04.001. [39] DUONG T V, CHANG P W, YANG S H, et al. A new comprehensive short-form health literacy survey tool for patients in general[J]. *Asian Nurs Res*, 2017, 11(1): 30-35. DOI: 10.1016/j.anr.2017.02.001. [40] DUONG T V, NGUYEN T T P, PHAM K M, et al. Validation of the short-form health lit-

eracy questionnaire (HLS-SF12) and its determinants among people living in rural areas in Vietnam[J]. *Int J Environ Res Public Health*, 2019, 16(18): 3346. DOI: 10.3390/ijerph16183346. [41] TRAN T V, NGUYEN H C, PHAM L V, et al. Impacts and interactions of COVID-19 response involvement, health-related behaviours, health literacy on anxiety, depression and health-related quality of life among healthcare workers: a cross-sectional study[J]. *BMJ Open*, 2020, 10(12): e041394. DOI: 10.1136/bmjopen-2020-041394. [42] PHAM K M, PHAM L V, PHAN D T, et al. Healthy dietary intake behavior potentially modifies the negative effect of COVID-19 lockdown on depression: a hospital and health center survey[J]. *Front Nutr*, 2020, 7: 581043. DOI: 10.3389/fnut.2020.581043. [43] SUN X M, GUAN D D. A comparative study of classical test theory and item response theory[J]. *China Examinations: Research Edition*, 2009(9): 10-17. DOI: 10.19360/j.cnki.11-3303/g4.2009.09.003. [44] YANG Z M, ZHANG L. *Generalizability Theory of Measurement and Its Application*[M]. Beijing: Educational Science Publishing House, 2003. [45] YU B B, YU L Z. Critical thinking disposition scale for LIS research and practice: feasibility test of simplifying CTDI-CV[J]. *Library and Information*, 2020(4): 21-27. DOI: 10.11968/tsyqb.1003-6938.2020058. [46] EMBRETSON S E, REISE S P. *Item Response Theory*[M]. Oxford: Psychology Press, 2013. [47] VAN DER ARK L A. Mokken scale analysis in R[J]. *J Stat Soft*, 2007, 20(11): 1-19. DOI: 10.18637/jss.v020.i11. [48] YUAN S L, HE Z. Nonparametric item response theory model—Mokken model[J]. *Journal of Guiyang University: Natural Sciences*, 2020, 15(4): 101-106. DOI: 10.16856/j.cnki.52-1142/n.2020.04.024. [49] WANG F, CHEN K, DU Z, et al. Reliability and validity analysis and Mokken model of New General Self-Efficacy Scale-Short Form (NGSES-SF)[J]. 2022. DOI: 10.31234/osf.io/r7aj3. [50] WU Y, TANG J, DU Z, et al. Development of a Short Version of the Perceived Social Support Scale: Based on Classical Test Theory and Item Response Theory[J]. 2022. DOI: 10.31234/osf.io/d95zg. [51] PAAKKARI L, OKAN O. COVID-19: health literacy is an underestimated problem[J]. *Lancet Public Health*, 2020, 5(5): e249-250. DOI: 10.1016/S2468-2667(20)30086-4. [52] SHIN H R, CHOI E Y, KIM S K, et al. Health literacy and frailty in community-dwelling older adults: evidence from a nationwide cohort study in South Korea[J]. *Int J Environ Res Public Health*, 2021, 18(15): 7918. DOI: 10.3390/ijerph18157918.

(Received: January 21, 2023; Revised: June 20, 2023)

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.