International Journal of Population Studies

RESEARCH ARTICLE

Validity and reliability of Mini-Mental State Examination in Older Adults in China: Inline Mini-Mental State Examination with cognitive functions

Teck Kiang Tan[1]*, Qiushi Feng[2]

[1]Institute for Applied Learning Sciences and Educational Technology, National University of Singapore, Singapore
[2]Department of Sociology, Centre for Family and Population Research, National University of Singapore, Singapore

Abstract: The main aim of the study is to validate the factor structure of the Mini-Mental State Examination (MMSE) of China's older population using the Chinese Longitudinal Healthy Longevity Survey. The validation process used the exploratory factor analysis (EFA) to determine the number of dimensions of MMSE, the confirmatory factor analysis (CFA) to confirm the factorial structure of MMSE, and the factorial invariance to conclude the factor structure does not differ between the young-old (aged 65 – 79) and old-old (aged 80 or older). The results of the EFAs suggested two possible factor structures: A six-factor and a seven-factor solution. The seven-factor confirmatory factor model turned out as the best fit by comparison to the four competing confirmatory models. Strict factorial invariance was attained for the two age groups, indicating a high level of measurement equality, a property of invariance was seldom achieved in the literature of factorial invariance studies. In comparison to the MMSE literature that focused solely on EFA that aims to establish a single summated score, the present study suggests using EFA, CFA, and factorial invariance that takes into consideration of measurement errors as the preferred procedure since it establishes the appropriate MMSE dimensionality that is in line with their respective cognitive functions.

Keywords: Mini-Mental State Examination; Validation; Exploratory factor analysis; Confirmatory factor analysis; Factorial invariance

ARTICLE INFO

Received: February 23, 2022
Accepted: May 17, 2022
Published: June 8, 2022

***CORRESPONDING AUTHOR**

Teck Kiang Tan,
Institute for Applied Learning Sciences and Educational Technology, National University of Singapore, Singapore.
alsttk@nus.edu.sg

CITATION

Tan TK, Feng Q. (2022). Validity and reliability of Mini-Mental State Examination in Older Adults in China: Inline Mini-Mental State Examination with cognitive functions. *International Journal of Population Studies*, 8(1):1-16. doi: 10.18063/ijps.v8i1.1285

1. Introduction

The cognitive function represents an important health dimension of older adults, which significantly affects their physical health, well-being, and mortality (Hsieh, Wu, Wang, *et al*., 2021; Langa, Llewellyn, Lang, *et al*., 2009; Park, Kwon, Jung, *et al*., 2012; van der Meulen, Irven, Bakunina, *et al*., 2021). The Mini-Mental State Examination (MMSE) was first introduced by Folstein *et al*. (1975) more than 40 years ago to measure cognitive performance. Since then, MMSE was widely used and has been accepted as an overall measure of cognitive impairment in clinical, research, and community settings. The MMSE is nowadays a widely used test of cognitive function among older adults. Its usability and applicability were well noted in social and medical research as a short screening tool, especially for older adults at risk of mild cognitive impairment (Folstein, Folstein, and McHugh, 1975; Lezak, Howieson, and Loring, 2004), screening the risk of dementia and Alzheimer's disease (Burke, Grudzien, Burgess, *et al*., 2021; Mitchell, 2009; Arevalo-Rodriguez, Smailagic, Roqué-Figuls, *et al*., 2021; Shigemori, Ohgi, Okuyama, *et al*., 2010).

The MMSE specifies a list of cognitive domain functions as an inventory that contains items to measure them which include orientation, registration, attention, calculation, spelling, recall, delayed recall, naming, repetition, verbal and written comprehension, and visuospatial capability (Davey and Jamieson, 2004; Folstein, Folstein, and McHugh, 1975; Reilly, Challis, Burns *et al*., 2004; Shulman, Herrmann, Brodaty, *et al*., 2006). The past studies that investigated the factorial structure of MMSE remained unsettled about its dimensionality. These studies did not show the alignment between the domain functions to the MMSE items about the dimensionality it is supposed to form and normally ended up with a lower dimension, not in line with the theoretically expected dimension (Folstein, Folstein, and McHugh, 1975; Park, Kwon, Jung, *et al*., 2012; Shigemori, Ohgi, Okuyama, *et al*., 2010; Tinklenberg, Brooks, Tanke, *et al*., 1990). The first study on MMSE factorial structure in the literature indicated a two-factor structure with the first factor including attention/concentration, language, and constructional praxis, and the second comprising time-space orientation and delayed recall (Fillenbaum, Heyman, Wilkinson, *et al*., 1987). Three years later, similar results of a two-factor solution were established with the first factor including writing, naming, immediate memory, reading a sentence, and verbal comprehension, and the second factor including constructional praxis, delayed recall, temporal orientation, attention/concentration, and spatial orientation (Tinklenberg, Brooks, Tanke, *et al*., 1990). The non-alignment of the MMSE items to the cognitive functions was clearly shown by these two studies that the MMSE items within the two factors differed. The number of dimensions increased in the later year publication to three factors (Shigemori, Ohgi, Okuyama, *et al*., 2010). While the number of factors was not a straightforward solution empirically, the number of dimensions also differed within a study between the control and experimental groups (Baek, Kim, Park, *et al*., 2016), and some ignored the determination of dimensionality procedure and directly used the domain MMSE summated subscores (e.g., Park, Kwon, Jung, *et al*., 2012).

The exploratory factor analysis (EFA) was the main approach in the MMSE validation literature to determine the number of factors for the MMSE inventory and the main statistics used was the percentage of variance explained (e.g., Shigemori, Ohgi, Okuyama, *et al*., 2010). This statistic explains the level of variances that could be extracted from an MMSE inventory, however, it does not point out whether the number of dimensions is statistically and theoretically appropriate. Choosing a higher level of variance explained will end in a higher dimension and vice versa. This statistic easily allows an arbitrary decision in determining the dimension. Theoretically, a prior expectation about the dimensionality should be decided before starting a study such as recall and visuospatial capability which are most probably two separate factors. Using the appropriate and developed statistical procedures that have been established in the measurement literature to determine the dimensionality of MMSE were also absent from the MMSE validation literature. These developments include the Velicer's MAP criteria (Lim, Li, Xie, *et al*., 2019; Velicer and Jackson, 1990), the Horn's parallel analysis (Dinno, 2009; Garrodp, Abad, and Ponsoda, 2013; Hayton, Allen, and Scarpello, 2004; Horn, 1965), the very simple structure (VSS) criterion (Revelle and Rocklin, 1979), various information criteria, and approaches such as BIC and sample adjusted BIC (Schwartz, 1978; Sclove, 1987).

EFA to establish the number of dimensions is the preliminary step in a validation process. Further incorporating confirmatory factor analysis (CFA) and the factorial invariance test into the validation procedure were already established in the fields of measurement, psychology, and education (Bollen and Lennox, 1991; Brown, 2006; Liau, Chow, Tan, *et al*., 2010; Lane, Anderson, Ponce, *et al*., 2012) and acknowledged as part of the validation process. That is, after EFA, CFA should be carried out to confirm the structure of an inventory. Providing the fit of a hypothesized CFA model and comparing it with competing CFA models are part of the validation procedure to ensure that the best fitted CFA is chosen (Brown, 2006; Jackson, Gillaspy, and Purc-Steogebsib, 2009; Liau, Chow, Tan, *et al*., 2010). The factorial invariance further confirms if the chosen CFA model is applicable in various controlled conditions. The reliability and validity of measurements on the health of older adults were constantly a concern (Gu, 2005). Due to the major health disparity between young-old (aged 65 – 79 years old) and oldest-old (aged over 80 years old), it is desirable to check the factorial invariance of these two major age groups. The factorial invariance testing was carried out in this study to make sure the MMSE psychometric properties for both the young-old and old-old age groups were invariant. After the best fitted confirmatory factor model was chosen, four factorial invariance tests were carried out to ensure there is no inequality in their factor structure and form despite their declining health conditions (Putnick and Bornstein, 2016; Revuelta, Franco-Martinez, and Ximénez, 2021; Schürer, van Ophuysen, and Behrmann, 2021).

The main purpose of this study is to find out the MMSE factorial structure of China's older population using the Chinese Longitudinal Healthy Longevity Survey (CLHLS). While the reliability and validity of MMSE have been carried out for each wave of the CLHLS using EFA (Gu, 2005; CLHLS, 2020), the present paper is to reveal the factor structure of MMSE from a more refined methodological base. Using the most recent 2018 wave of the CLHLS, this study carried out EFA and CFA, further examined the factorial invariance for the young-old and old-old, and reported reliability levels with the consideration of the binary nature of the data, using an appropriate estimation method.

2. Data and Methods

2.1. Data Sources

The data were extracted from the CLHLS, which is a national longitudinal survey of Chinese older adults. The CLHLS started in the year 1998 and accomplished a total of eight waves in 1998, 2000, 2002, 2005, 2008/2009, 2011/2012, 2014, and 2018/2019. More details about the survey were published in the literature (Gu *et al*., 2021). The present study used the most recent wave in the year 2018/2019. The 2018/2019 wave had 15,896 older adults. For the purposes of this study, 7603 respondents were dropped for missing values in any of the MMSE items, which led to a sample size of 8293 older adults. As the sample size is large after deletion, it has little effect on the validation results. All participants involved in the survey consented to take part in the survey, including answering the MMSE items. The sample was further divided into the young-old aged 65 – 79 and the oldest-old aged 80+, with the former consisting of 4751 and the latter having 4118 persons.

2.2. Measures for MMSE Inventory

The MMSE inventory contained 23 items (Folstein, Folstein, and McHugh, 1975). The CLHLS developed a Chinese version of the MMSE, similar to that of Yi and Vaupel (2002), with revised items that take into account the Chinese cultural and socioeconomic context making the items easily understandable and practically answerable. These items comprise the components to measure orientation to time, place, and season, short registration, delayed registration, calculation, instruction, visualization, attention, and language. The content, construct, and concurrent validities were covered in the various waves of CLHLS which were not repeated in the present paper (e.g., Gu, 2005). The usual practice of CLHLS users tended to sum up these items for a score ranging from 0 to 30. In the present paper, all the 23 items were categorized as binary.

2.3. Procedure of Validation

The validation process is broadly breakdown into three major stages: First, EFA to determine the number of dimensions; second, CFA to obtain the best factor structure of MMSE; and last, factorial invariance testing to examine the degree of invariance concerning age. These three steps are elaborated on in the following three subsections. The various functions from the software package R were used for analyses.

2.3.1. EFA

EFAs were carried out to determine the number of dimensions and the factor structure of MMSE. First, factorability was carried out to assess the adequacy of carrying out factor analysis. Bartlett test of sphericity (Bartlett, 1951) and Kaiser–Mayer–Olkin (KMO) measures of sampling adequacy (MSA) (Kaiser, 1970; Kaiser and Rice, 1974) were generated to provide the evidence of factorability of factor analysis, indicating the adequacy for carrying out the factor analysis and providing evidence about the sufficient numbers of significant correlations among the 23 MMSE indicators to justify undertaking factor analysis (Pett, Lackey, and Sullivan, 2003). For individual MMSE items, individual MSA was reported. Function KMO and cortest.bartlett from the package psych were used to generate these three factorability indices (Revelle, 2021).

Next, the heterogeneous correlation matrix for the 23 binary coded MMSE items was generated using the function hector from the package polycor (Fox, 2019) as an input to carrying out EFA using the package psych, function fa (Revelle, 2021). The present study used several methods that were commonly recommended in the literature (Finch, 2020) to determine the dimension of the MMSE inventory. These methods included the Kaiser's greater than 1 rule (Kaiser, 1960), scree test (Cattell, 1966), Velicer's minimum average partial procedure criteria (map; Velicer, 1976; Velicer and Jackson, 1990), Horn's parallel analysis (Dinno, 2009; Glorfeld, 1995; Hayton, Allen, and Scarpello, 2004; Horn, 1965), VSS criterion (Revelle and Rocklin, 1979), BIC (Schwartz, 1978), sample size adjusted BIC (SABIC; Sclove, 1987), root mean square error of approximation (RMSEA; Browne and Cudeck, 1992), and standardized root means square of the residuals (RMSR; Bentler, 1995). The package psych, function VSS.scree, and VSS generated the scree plot, the VSS, RMSEA, RMSR, and the various information criteria indicators. The package paran and function paran produced the parallel analysis. The ultimate decision to determine the number of factors was based on the outcomes of the above-mentioned methods and the interpretability of the EFA results.

2.3.2. CFA

After determining the number of factors, CFAs were carried out using the package lavaan function cfa with the specification of ordered = TRUE to indicate binary CFA which was performed. The weighted least square mean and variance adjusted

estimator was used for the estimation. It used diagonally weighted least squares to estimate the model parameters. Model fit for CFAs was assessed using multiple indices, consisting of the χ^2 statistic (Bollen, 1989; Jöreskog, 1993), comparative fit index (CFI; Bentler, 1990), Tucker and Lewis index (TLI; Tucker and Lewis, 1973), (RMSEA; Browne and Cudeck, 1992), and standardized root mean square residual (SRMR; Brown, 2006). The χ^2, RMSEA, and SRMR assess how well the covariances predicted from the estimates reproduced the sample covariances (Pomplun and Omar, 2001). The CFI and TLI assess the degree of fit of the proposed model accounted for the sample covariances. RMSEA values approximating 0.06 demonstrate a close fit of the model (Browne and Cudeck, 1992; Hu and Bentler, 1999). A value <0.08 is generally considered a good fit for SRMR (Hu and Bentler, 1999). CFI and TLI values of 0.90 (Bentler, 1990) and 0.95 (Hu and Bentler, 1999), respectively, indicate an acceptable and good fit for the model.

2.3.3. Factorial invariance

After determining the CFA model that best described the structure of the MMSE, factorial invariance testing was carried out to examine the degree of invariance concerning age. Factorial invariance is a concept applied in the context of psychometric analysis of an inventory (Revuelta, Franco-Martinez, and Ximénez, 2021; Schürer, van Ophuysen, and Behrmann, 2021; Putnick and Bornstein, 2016) to measure the degree of invariance assurance for a categorical variable about its applicability level. In the current context, this concept postulates the psychometric properties of MMSE, and its applicability for the two age groups, young-old and old-old. The first invariance testing is configural invariance, commonly referred to as pattern invariance. This testing procedure aims to examine whether both the young-old and old-old have the same MMSE factor structure (Cheung and Rensvold, 2002; Horn and McArdle, 1992; Vandenberg and Lance, 2000). While the configural invariance is satisfied, the next step is to set the factor loadings to be equal across the two age groups. This is generally referred to as metric invariance. Metric invariance is built on configural invariance by requiring that in addition to the equality of the structural form, the factor loadings are also equivalent across the two age groups. The next level of constraint that adds to the invariance is the intercepts. It was also referred to as scalar equivalence (Mullen, 1995), indicating the existence of strong factorial invariance (Meredith, 1993). While the mean is further constrained across the groups, it is referred to as strict factorial invariance, indicating systematic group differences in means matrices are due to group differences in common factor score distributions (Yoon and Millsap, 2007). The factorial invariance testing was carried out using function cfa from the package lavaan (Rosseel, 2012) with the specification of group.equal to "loadings," "intercepts," and "means" for metric, scalar, and strict invariance, respectively.

3. Results

3.1. Descriptive Statistics

Descriptive statistics were generated using the function dfSummary from the package summarytools (Comtois, 2021). Table 1 shows the mean and standard deviation of the 23 MMSE items for the study and the two age groups of young-old and old-old.

3.2. EFA

3.2.1. Factorability

The KMO measure of sampling adequacy was 0.9, indicating the adequacy of undertaking factor analysis. Similarly, the Bartlett test of sphericity also indicated the sufficiency numbers of significant correlations that it was unlikely the population correlation matrix was an identity matrix ($\chi^2 = 69{,}555$; $P < 0.001$). The individual MSA that ranged from 0.77 to 0.97 also gave the same conclusion.

3.2.2. Determining number of factors

The various methods of determining the number of factors gave diverse outcomes that ranged from 1 to 7. However, the majority of the fit indicators showed either a six-factor or a seven-factor solution. The heterogeneous correlation correlogram with the correlation coefficients shown in Figure A1 and without correlation coefficients shown in Figure A2 indicated either a six-factor or a seven-factor solution. The scree plot and Horn's parallel analysis showed a six-factor solution (Figure A3). Out of the six fit indices from Table A1, RMSEA, BIC, and SABIC suggested seven factors. While the Velicer's minimum average partial procedure criteria (map) showed a five-factor model, the two vss1 and vss2 indicated a one-factor and a two-factor solution, respectively. The Kaiser's greater than 1 rule indicated

Table 1. MMSE items and older group descriptive statistics.

MMSE item	Mean	Standard deviation
Orientation		
Morning, noon, afternoon, and night	0.99	0.12
Month	0.95	0.22
Mid-autumn festival (month and date)	0.94	0.25
Four seasons	0.97	0.18
Home – region/village	0.98	0.15
Short/delayed recall/registration		
Registration of desk	0.96	0.20
Registration of apple	0.95	0.22
Registration of dress	0.94	0.24
Delayed recall		
Recall of desk – repetition	0.84	0.37
Recall of apple – repetition	0.84	0.37
Recall of dress – repetition	0.80	0.40
Calculation		
Calculation – 20-3	0.95	0.21
Calculation – 20-3-3	0.89	0.32
Calculation – 20-3-3-3	0.88	0.32
Calculation – 20-3-3-3-3	0.85	0.35
Calculation – 20-3-3-3-3-3	0.86	0.35
Language		
Verbal repetition a sentence	0.95	0.21
Name pen	0.99	0.08
Name watch	0.99	0.08
Comprehend instruction		
Instruction 1 – Pick paper with right hand	0.97	0.16
Instruction 2 – Fold the paper	0.97	0.16
Instruction 3 – Place paper on the floor	0.98	0.13
Visuospatial/drawing		
Draw intersecting pentagons	0.64	0.48
Age		
All (65 and above)	80.5	10.11
Young-old (65–79)	72.2	4.27
Old-old (80 and above)	88.9	6.85

a six-factor solution with a moderately high 61.57% cumulative percentage explained (Table A2). In summary, both the six-factor and seven-factor models were two more likely exploratory factor models, suggested by the majority of the fit indicators.

3.3. CFA

Based on the theoretical cognitive function classification of MMSE together with the results of EFA, the base reference factor model was set to the seven-factor model while the six-factor model was specified as the supporting model. The main difference between the six-factor and seven-factor was mainly due to the further breakdown of the six MMSE items into two constructs of language and instruction. This led to stating five CFA models with the seven-factor oblique model

as the hypothesized model and the rest as competing CFA models. The first proposed competing CFA was a one-factor model. This model indicated all the 23 MMSE items were formed into a single MMSE construct. The six-factor oblique CFA was the next proposed competing CFA model. While the fit of the seven-factor was better than the six-factor solution, the orthogonal seven-factor was included to examine whether an unrelated of the seven-factor constructs of MMSE was better than the seven-factor oblique model. While the one-factor CFA model was that all the 23 items were grouped under one overall MMSE construct, a second-order seven-factor CFA also indicated an overall MMSE construct however loaded under the 7 MSSE constructs instead of the 23 MMSE items.

Table 2 shows the results of the proposed seven-factor oblique CFA model and the four competing CFA models. The fit indices CFI and TLI of the hypothesized seven-factor oblique model (1.0; 1.0) were higher than the four competing models. Similarly, the RMSEA and SRMR of the hypothesized model also indicated a better fit than the four competing models with the lowest values (0.009; 0.033). These results indicated the proposed seven-factor oblique model fitted better than the four competing models. Table 3 displays the latent correlations of the seven factors together with their reliability indicator, Zumbo's alpha. Since all the MMSE items were binary coded, the ordinal Zumbo's alpha was reported (Zumbo, Gadermann, and Zeisser, 2007). The factor correlations for all the seven domains of MMSE were all positive. These positive coefficients indicated an older adult that possessed a high MMSE construct in one cognitive function domain tended to also possess high in another domain. For instance, the short recall was moderately correlated with language with a positive correlation coefficient of 0.64, indicating that an older adult that has a high recall was also high in language capability. These results supported the proposed theoretical CFA that these seven MMSE domains were distinct yet associated. The Zumbo's alpha coefficients for all the six MMSE constructs were all high in reliability with values all higher than 0.89.

Figure 1 shows the graphical representation of the seven-factor oblique CFA model. These seven factors were orientation, short recall, delayed recall, calculation, language, comprehend instruction, and visuospatial. The standardized factor loading coefficients were indicated on top of the arrow that ran from the construct to the MMSE items. All the factor loadings were high in value except for visuospatial. These high loadings indicated the high association of the MMSE items to the respective latent factor MMSE construct. The error residuals were printed after the items, on top of the arrow that ran from error terms (E1 to E23) to the MMSE items. These error residuals were low in value also indicating that the MMSE items were low in measurement errors when they were loaded into the appropriate MMSE construct. The double arrows that ran between the seven constructs indicated it represents an oblique CFA model.

Table 2. Fit indices of hypothesized cfa and competing models.

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
Hypothesized seven-factor oblique	341*	209	1.000	1.000	0.009	0.034
Competing seven-factor second order	605*	223	1.000	0.999	0.014	0.044
Competing six-factor oblique	609*	215	1.000	0.999	0.015	0.050
Competing one-factor	9196*	230	0.976	0.973	0.069	0.123
Competing seven-factor orthogonal	77470*	230	0.789	0.768	0.201	0.467

χ^2, Chi-square statistics; df, degrees of freedom; CFI, comparative fit index, TLI, Tucker-Lewis index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square. *$P < 0.01$.

Table 3. Factor correlations and Zumbo's alpha for the seven domains of MMSE.

MMSE construct	1.	2.	3.	4.	5.	6.	7.	Zumbo's alpha
1. Orientation								0.93
2. Short recall	0.53							0.96
3. Delay recall	0.48	0.57						0.94
4. Calculation	0.57	0.58	0.50					0.98
5. Language	0.56	0.64	0.54	0.63				0.91
6. Instruction	0.32	0.40	0.34	0.40	0.51			0.89
7. Visuospatial	0.38	0.37	0.35	0.50	0.39	0.29		-

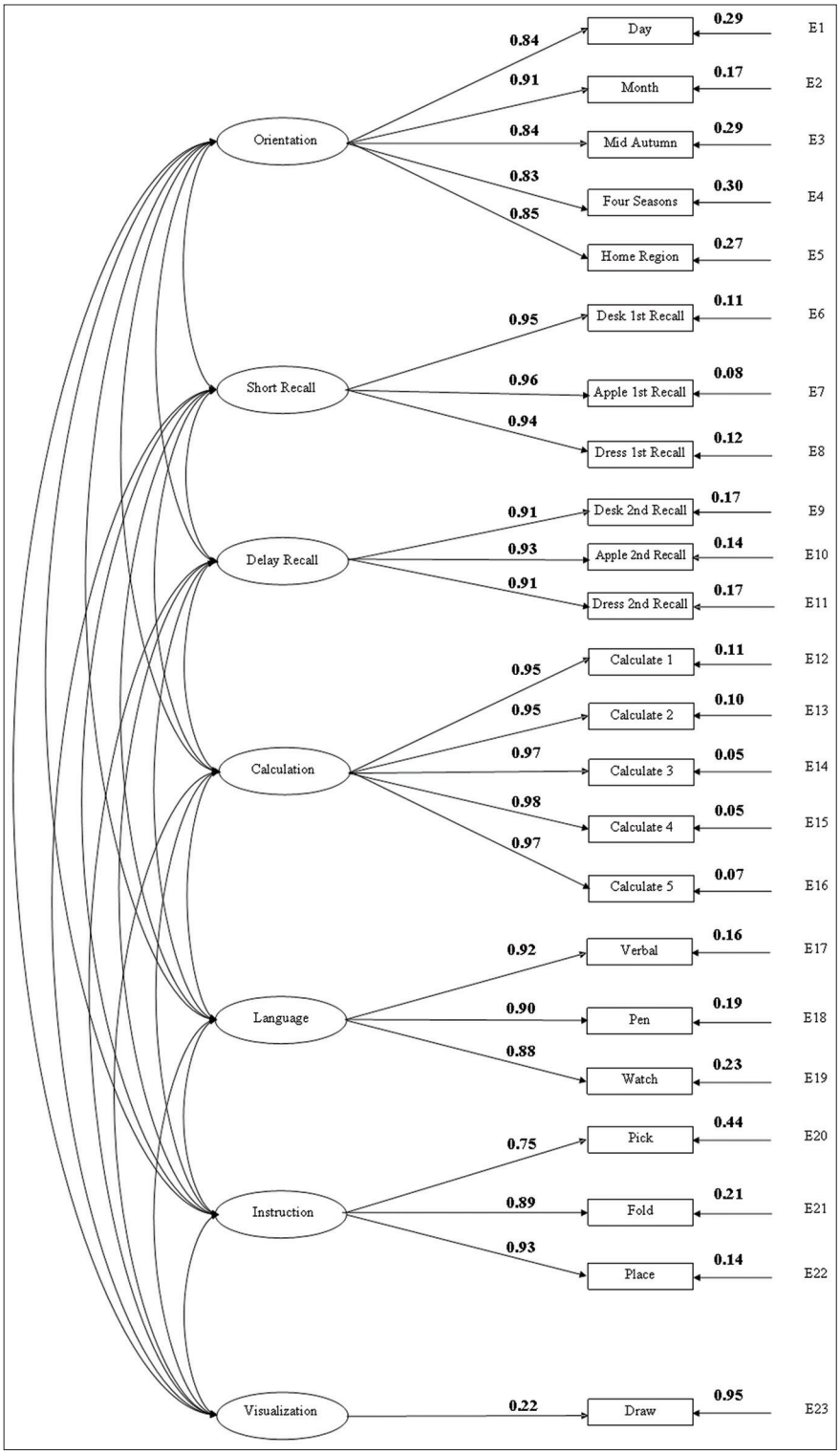

Figure 1. Mini-Mental State Examination seven-factor oblique confirmatory factor model.

3.4. Factorial Invariance

Table 4 summarizes the results of factorial invariance of the two age groups. The CFI for all the four invariance conditions was all at a high value of 0.999. Similar results were also found for the TLI. Both RMSEA and SRMR

Table 4. Factorial invariance – young-old and old-old.

Factorial invariance	CFI	TLI	RMSEA	SRMR
Configural	0.999	0.999	0.009	0.057
Metric	0.999	0.998	0.015	0.066
Scalar	0.999	0.999	0.010	0.060
Strict	0.999	0.999	0.012	0.059

CFI, comparative fit index; TLI, Tucker-Lewis index; RMSEA, root mean square error of approximation; SRMR, standardized root mean squared residual.

were low at a satisfactory level. These results indicated that the level of invariance for the MMSE was extremely good.

4. Conclusion and Discussion

This study explored and validated the factor structure of MMSE through EFA, CFA, and the factorial invariance test. The validation results indicated a seven-factor oblique CFA best fitted the MMSE inventory, specifying seven cognitive functions inherently within it: Orientation, short recall, delayed recall, calculation, language, comprehend instruction, and visuospatial. These results were similar to MMSE literature on their cognitive functions but differed in the number of dimensions. The factorial invariance confirmed the hypothesized CFA was at a high invariance level showing almost no measurement differences between the young-old and old-old. The reliability results also indicated that these 23 MMSE indicators that formed the seven factors were of high reliability.

One main finding of the present study is the statistical results of the CFA indicated the alignment between the 23 MMSE items with the theoretical expected seven cognitive functions. For instance, the three items of registration of desk, apple, and dress were fittingly grouped under the short recall cognitive function, and the five subtraction calculation items properly formed the calculation cognitive function. This finding did not appear in the MMSE literature that generally reported a low dimension (e.g., Fillenbaum, Heyman, Wilkinson, *et al*., 1987). One plausible reason is that the MMSE literature was not predominantly to establish the dimensions of the MMSE inventory with the main purpose being to relate the MMSE items to the respective cognitive functions they belonged to (e.g., Fillenbaum, Heyman, Wilkinson, *et al*., 1987; Tinklenberg, Brooks, Tanke, *et al*., 1990) but reluctantly used the EFA to confirm a low dimension to form a summated MMSE score. Another probable reason is that more updated appropriate EFA procedures to determine the number of dimensions used in the present paper were not adopted even for the more recent papers (e.g., Baek, Kim, Park, *et al*., 2016).

The present paper also revealed the characteristics of the MMSE inventory possess a duality factorial structure that could be viewed as a seven dimension of cognitive functions, and also as a general higher-order all-inclusive cognitive function construct within which the seven dimensions of cognitive functions were grouped under it. This duality property was reflected in the results of CFA. The seven-factor CFA model turned out as the best model in which the MMSE items were grouped under their respective cognitive functions according to the theoretical grouping showed not only the alignment with theory expectation but also indicating the first property of this inventory that these seven cognitive functions were associated but were separate constructs. The second best fit CFA, the second-order CFA with a slightly lower fit, indicated the possibility to view this inventory as a second-order cognitive function construct. The moderate to high factor correlations of the seven constructs within the seven oblique CFA further showed evidence that these seven constructs were positively associated but were different in their cognitive functionality. Similarly, the low fit of the seven-factor orthogonal CFA also indicated the unlikeliness that the seven MMSE constructs were unrelated. The practical implication of this duality property is that MMSE can be viewed as an overall indicator or as separate seven distinct but related cognitive functions.

Another crucial conclusion and inference from the CFA result is that it indicates the routine way of generating an MMSE score, whether it is viewed as separate seven constructs or an overall higher-order construct, the summated score to generate an overall MMSE score or subscale MMSE scores by summing the items is not an appropriate procedure. The earlier studies on the validation of MMSE often use an all-inclusive MMSE summated score to represent an index in measuring the level of cognitive function by summing the MMSE items according to the number of items correctly answered (e.g., Park, Kwon, Jung, *et al*., 2012 used domain MMSE summated scores). The limitations of the summated score were well noted in the measurement literature. The main limitation is that summated score does not take care of measurement errors. When the summated score is further subdivided into a few categorical levels using a cutoff

arbitrary decision to distinguish the risk level of cognitive impairment, additional measurement errors are introduced and converting a continuous score to a categorical variable produces a loss of information. However, this procedure was commonly found in studies that used MMSE (Deb and Braganza, 1999; Osterweil, Mulford, Syndulko, *et al*., 1994; Yi and Vaupeo, 2002). More seriously, the deriving of an optimal cutoff score (e.g., Wong and Fong, 2009) has nothing to do with the validity of the MMSE inventory and its dimensionality, and the relationship to the cognitive functions but a vaguely derived score is generated that contains measurement errors.

The results of the factorial invariance indicated the MMSE inventory possessed a high degree of invariance between the two age groups which were seldom found in the factorial invariance literature. The configural invariance indicated that both the young-old and old-old have the same factorial structure. This gave the conclusion that the structural form between MMSE items and their cognitive functions is the same for the two age groups. The metric invariance further provided evidence that the factor loadings were also invariant across the two age groups. That is the weights that indicated the association between the MMSE items and the cognitive function construct were also maintained, reflected by the equality of their respective factor loadings. Scale invariance further qualified the latent MMSE construct was also with the same degree of measurement errors across the two age groups. The last invariance, the strong factorial invariance, is a prerequisite to testing for the equality of latent means. In the presence of this invariance, the comparison of latent means becomes unambiguous (Cheung and Rensvold, 2002), indicating that the MMSE could be used with high confidence for latent model analysis, where systematic group differences in means matrices are due to group differences in common factor score distributions (Yoon and Millsap, 2007). In summary, the results of the factorial invariance assured the use of the MMSE as a seven latent dimensions construct and suggested that moving away from the commonly used summated score approach which contained measurement errors to latent modeling is a better direction for further analysis using latent models.

For further analyses and follow-up after validation, the recommendation is to set CFA as the base to establish the measurement component of MMSE. When the structural component is considered, a latent approach is recommended to use the various latent models for further analyses. For instance, using the structural equation model to relate MMSE items and the cognitive function constructs to a group of covariates or examining the effect of cognitive function constructs on a medical condition. The latent modeling followed after CFA has several advantages. First, the number of MMSE items to include in an MMSE inventory becomes a researcher's choice that the option to choose a subset of MMSE items that are more related to the study does not restrict to a complete list of MMSE items. Second, the creation of the various cognitive functions allows for addressing the effects separately from the overall cognitive function or an individual cognitive subscale. For instance, examining the effect of the overall cognitive function on a medical condition and concurrently exploring the effect of the calculation subscale on the same medical condition. Third, removing the measurement errors is automatically inbuilt into the latent model.

In summary, this paper recommends validation of MMSE using EFA, CFA, and factorial invariance test and showed the results of the validation that MMSE is more appropriate than the MMSE literature that only concentrated on EFA. While this systematic approach was commonly used and already established in the measurement, psychology, and education literature, it is recommended for future use of MMSE. These procedures avoid all the limitations discussed in the paper. It takes into account measurement errors, relates the MMSE items more appropriately to the theoretical cognitive function setting, creates competing CFA models that are appropriately set up before testing, and tests for invariances.

Funding

This study was not supported by any grant.

Conflicts of Interest

No conflicts of interest were reported by all others.

Authors' Contributions

TKT conducted the analysis and drafted the introduction, methods, results, and discussions of the manuscript. QF reviewed, amended, and provided recommendations on the organization of the manuscript.

Ethical Approval

The human data used in our study are a publicly available survey dataset that can be downloaded from the webpage: https://cpha.duke.edu/research/chinese-longitudinal-healthy-longevity-survey-clhls.

Availability of Supporting Data

The CLHLS dataset is in open access on the webpage: https://cpha.duke.edu/research/chinese-longitudinal-healthy-longevity-survey-clhls.

References

Arevalo-Rodriguez I, Smailagic N, Roqué-Figuls M, *et al*. (2021). Mini-mental state examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, 7(7):CD010783. https://doi.org/10.1002/14651858.CD010783.pub3

Baek MJ, Kim K, Park YH, *et al*. (2016). The validity and reliability of the MMSE-2 for detecting mild cognitive impairment and Alzheimer's disease in a Korean population. *PLoS One*, 11(9):e0163792. https://doi.org/10.1371/journal.pone.0163792

Bartlett MS. (1951). The effect of standardization on a Chi-Square approximation in factor analysis. *Biometrika*, 38:337-344. https://doi.org/10.1093/biomet/38.3-4.337

Bentler PM. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107:238-246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler PM. (1995). *EQS 5* [*Computer Program*]. Encino, CA: Multivariate Software Inc.

Bollen K and Lennox R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110:305-314. https://doi.org/10.1037/0033-2909.110.2.305

Bollen KA. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Brown TA. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.

Browne MW and Cudeck R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21:230-258. https://doi.org/10.1177/0049124192021002005

Burke SL, Grudzien A, Burgess A, *et al*. (2021). The utility of cognitive screeners in the detection of dementia spectrum disorders in Spanish-speaking populations. *Journal of Geriatric Psychiatry and Neurology*, 34(2):102-118. https://doi.org/10.1177/0891988720915513

Cattell RB. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1:245-276. https://doi.org/10.1207/s15327906mbr0102_10

Cheung GW and Rensvold RB. (2002). Evaluating goodness of fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2):233-255. https://doi.org/10.1207/s15328007sem0902_5

CLHLS. (2020). *Chinese Longitudinal Healthy Longevity Survey, Data Collection and Quality Control Procedures*. Available from: https://sites.duke.edu/centerforaging/programs/chinese-longitudinal-healthy-longevity-survey-clhls/design-method-study-population/design-method-data-collection-quality-control-procedures. [Last accessed on 2022 Jan 26]. https://doi.org/10.1007/978-1-4020-6752-5_2

Comtois D. (2021). *summarytools: Tools to Quickly and Neatly Summarize Data. R package version 1.0.0*. Available from: https://CRAN.R-project.org/package=summarytools.[Last accessed on 2022 Jan 26].

Davey RJ and Jamieson S. (2004). The validity of using the mini-mental state examination in NICE dementia guidelines. *Journal of Neurology Neurosurgery and Psychiatry*, 75:343-344.

Deb S and Braganza J. (1999). Comparison of rating scales for the diagnosis of dementia in adults with Down's syndrome. *Journal of Intellectual Disability Research*, 43(5):400-407. https://doi.org/10.1046/j.1365-2788.1999.043005400.x

Dinno A. (2009). Exploring the sensitivity of Horn's parallel analysis to the distributional form of simulated data. *Multivariate Behavioral Research*, 44(3):362-388. https://doi.org/10.1080/00273170902938969

Fillenbaum GG, Heyman A, Wilkinson WE, *et al*. (1987). Comparison of two screening tests in Alzheimer's disease. The correlation and reliability of the mini-mental state examination and the modified blessed test. *Archives of Neurology*, 44:924-927. https://doi.org/10.1001/archneur.1987.00520210026014

Finch WH. (2020). Using fit statistic differences to determine the optimal number of factors to retain in an exploratirt factor analysis. *Educational and Psychological Measurement*, 80(2):217-241. https://doi.org/10.1177/0013164419865769

Folstein MF, Folstein SE, and McHugh PR. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12:189-198. https://doi.org/10.1016/0022-3956(75)90026-6

Fox, J. (2019). *polycor: Polychoric and Polyserial Correlations. R package version 0.7-10*. Available from: https://CRAN.R-project.org/package=polycor. [Last accessed on 2022 Jan 26].

Garrodp LE, Abad FJ, and Ponsoda V. (2013). A new look at Horn's parallel analysis with ordinal variables. *American Psychological Association*, 18(4):454-474. https://doi.org/10.1037/a0030005

Glorfeld LW. (1995). An Improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55(3):377-393. https://doi.org/10.1177/0013164495055003002

Gu D. (2005). *Assessment of Data Quality of the Chinese Longitudinal Healthy Longevity Survey in 2002*. The CLHLS Technical Report No.2005-1.

Gu D, Feng Q, Chen H, et al. (2021). Chinese longitudinal healthy longevity survey. In: D Gu and ME Dupre (eds.). Encyclopedia of Gerontology and Population Aging. Switzerland AG, Cham: Springer Nature. p.957-970.

Hayton JC, Allen DG, and Scarpello V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2):191-205. https://doi.org/10.1177/1094428104263675

Horn JL and McArdle JJ. (1992). A practical and theoretical guide to MI in aging research. *Experimental Aging Research*, 18:117-144. https://doi.org/10.1080/03610739208253916

Horn JL. (1965). A rationale and a test for the number of factors in factor analysis. *Psychometrika*, 30:179-185. https://doi.org/10.1007/bf02289447

Hsieh SW, Wu DW, Wang CW, *et al*. (2021). Poor cognitive function is associated with obstructive lung diseases in Taiwanese adults. *International Journal of Environmental Research and* Public Health, 18:2344. https://doi.org/10.3390/ijerph18052344

Hu L and Bentler PM. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6:1-55. https://doi.org/10.1080/10705519909540118

Jackson DJ, Gillaspy JA, and Purc-Steogebsib R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1):6-23. https://doi.org/10.1037/a0014694

Jöreskog KG. (1993). Testing structural equation models. In: KA Bollen and JS Long (eds.). *Testing Structural Equation Models*. Newbury Park, CA: Sage, p.294-316.

Kaiser HF and Rice J. (1974). Little jiffy, mark Iv. *Educational and Psychological Measurement*, 34(1):111-117. https://doi.org/10.1177/001316447403400115

Kaiser HF. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20:114-151. http://dx.doi.org/10.1177/001316446002000116

Kaiser HF. (1970). A second generation little jiffy. *Psychometrika*, 35(4):401-415. https://doi.org/10.1007/BF02291817

Lane FC, Anderson B, Ponce HF, *et al*. (2012). Factorial invariance of LibQUAL as a measure of library service quality over time. *Library and Information Science Research*, 34:22-30. https://doi.org/10.1016/j.lisr.2011.07.004

Langa KM, Llewellyn DJ, Lang IA, *et al*. (2009). Cognitive health among older adults in the United States and in England. *BMC Geriatrics*, 9:23. https://doi.org10.1186/1471-2318-9-23

Lezak MD, Howieson DB, and Loring DW. (2004). *Neuropsychological Assessment*. 4th ed. New York: Oxford University Press.

Liau AK, Chow D, Tan TK, *et al*. (2010). Development and validation of the Personal strengths inventory using exploratory and confirmatory factor analyses. *Journal of Psychoeducational Assessment*, 29:14-26. https://doi.org/10.1177/0734282910365648

Lim M, Li Z, Xie H, *et al*. (2019). An Asian study on clinical and psychological factors associated with personal recovery in people with psychosis. *BMC Psychiatry*, 19:256. https://doi.org/10.1186/s12888-019-2238-9

Meredith W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58:525-543. https://doi.org/10.1007/bf02294825

Mitchell AJ. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, 43(4):411-431. https://doi.org/10.1016/j.jpsychires.2008.04.014

Mullen M. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 3:573-596. https://doi.org/10.1057/palgrave.jibs.8490187

Osterweil D, Mulford P, Syndulko K, *et al*. (1994). Cognitive function in old and very old residents of a residential facility: Relationship to age, education, and dementia. *Journal of the American Geriatric Society*, 42(7):766-773. https://doi.org/10.1111/j.1532-5415.1994.tb06539.x

Park MH, Kwon DY, Jung JM, *et al*. (2012). Mini-mental status examination as predictors of mortality in the elderly. *Acta Psychiatrica Scandinavica*, 127:298-304. https://doi.org/10.1111/j.1600-0447.2012.01918.x

Pett MA, Lackey NR, and Sullivan JJ. (2003). *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781412984898

Pomplun M and Omar MH. (2001). The factorial invariance of a test of reading comprehension across groups of limited English proficient students. *Applied Measurement in Education*, 14:261-283. https://doi.org/10.1207/s15324818ame1403_4

Putnick DL and Bornstein MH. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Development Review*, 41:71-90. https://doi.org/10.1016/j.dr.2016.06.004

Reilly S, Challis D, Burns A, *et al*. (2004). The use of assessment scales in old age psychiatry services in England and Northern Ireland. *Aging and Mental Health*, 8(3):249-55. https://doi.org/10.1080/13607860410001669787

Revelle W and Rocklin T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, 14:403-414. https://doi.org/10.1207/s15327906mbr1404_2

Revelle W. (2021). *psych: Procedures for Personality and Psychological Research*. Evanston, Illinois, USA: Northwestern University. Available from: https://CRAN.R-project.org/package=psych Version = 2.1.9. [Last accessed on 2022 Jan 26].

Revuelta J, Franco-Martinez A, and Ximénez C. (2021). Nominal factor analysis of situational judgment tests: Evaluation of latent dimensionality and factorial invariance. *Educational and Psychological Measurement*, 81(6):1054-1088. https://doi.org/10.1177/0013164421994321

Rosseel Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1-36. https://doi.org/10.18637/jss.v048.i02

Schürer S, van Ophuysen, S, and Behrmann, L. (2021). A new instrument for assessing cohesion in primary and grammar school classes factorial structure and measurement invariance of the GruKo4. *Journal of Psychoeducational Assessment*, 39(3):271-285. https://doi.org/10.1177/0734282920966825

Schwartz G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461-464.

Sclove LS. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52:333-343. https://doi.org/10.1007/bf02294360

Shigemori K, Ohgi S, Okuyama E, *et al*. (2010). The factorial structure of the mini-mental state examination (MMSE) in Japanese dementia patients. *BMC Geriatrics*, 10:36. https://doi.org/10.1186/1471-2318-10-36

Shulman KI, Herrmann N, Brodaty H, *et al*. (2006). IPA survey of brief cognitive screening instruments. *International Psychogeriatrics*, 18:281-294. https://doi.org/10.1017/s1041610205002693

Tinklenberg J, Brooks JO, Tanke ED, *et al*. (1990). Factor analysis and preliminary validation of the mini-mental state examination from a longitudinal perspective. *International Psychogeriatrics*, 2(2):123-134. https://doi.org/10.1017/s1041610290000382

Tucker LR and Lewis C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1):1-10. https://doi.org/10.1007/BF02291170

van der Meulen M, Dirven L, Bakunina K, *et al*. (2021). MMSE is an independent prognostic factor for survival in primary central nervous system lymphoma. *Journal of Neurooncology*, 152:357-362. https://doi.org/10.1007/s11060-021-03708-8

Vandenberg RJ and Lance CE. (2000). A review and synthesis of the MI literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3:4-69. https://doi.org/10.1177/109442810031002

Velicer WF and Jackson DN. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25:1-28. https://doi.org/10.1207/s15327906mbr2501_1

Velicer WF. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41:321-327. https://doi.org/10.1007/bf02293557

Wong SS and Fong NK. (2009). Reliability and validity of the telephone version of the cantonese mini-mental state examination

(T-SMMSE) when used with elderly patients with and without dementia in Hong Kong. *International Psychogeriatrics*, 1(2):345-353. https://doi.org/10.1017/s1041610209008588

Yi Z and Vaupel JW. (2002). Functional capacity and self-evaluation of health and life of oldest old in China. *Journal of Social Issue*, 58:733-748. https://doi.org/10.1111/1540-4560.00287

Yoon M and Millsap RE. (2007). Detecting violations of factorial invariance using data-based specification searches: A monte carlo study. *Structural Equation Modeling*, 14(3):435-463. https://doi.org/10.1080/10705510701301677

Zumbo BD, Gadermann AM, and Zeisser C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1):21-29. https://doi.org/10.22237/jmasm/1177992180

Publisher's note

Appendix

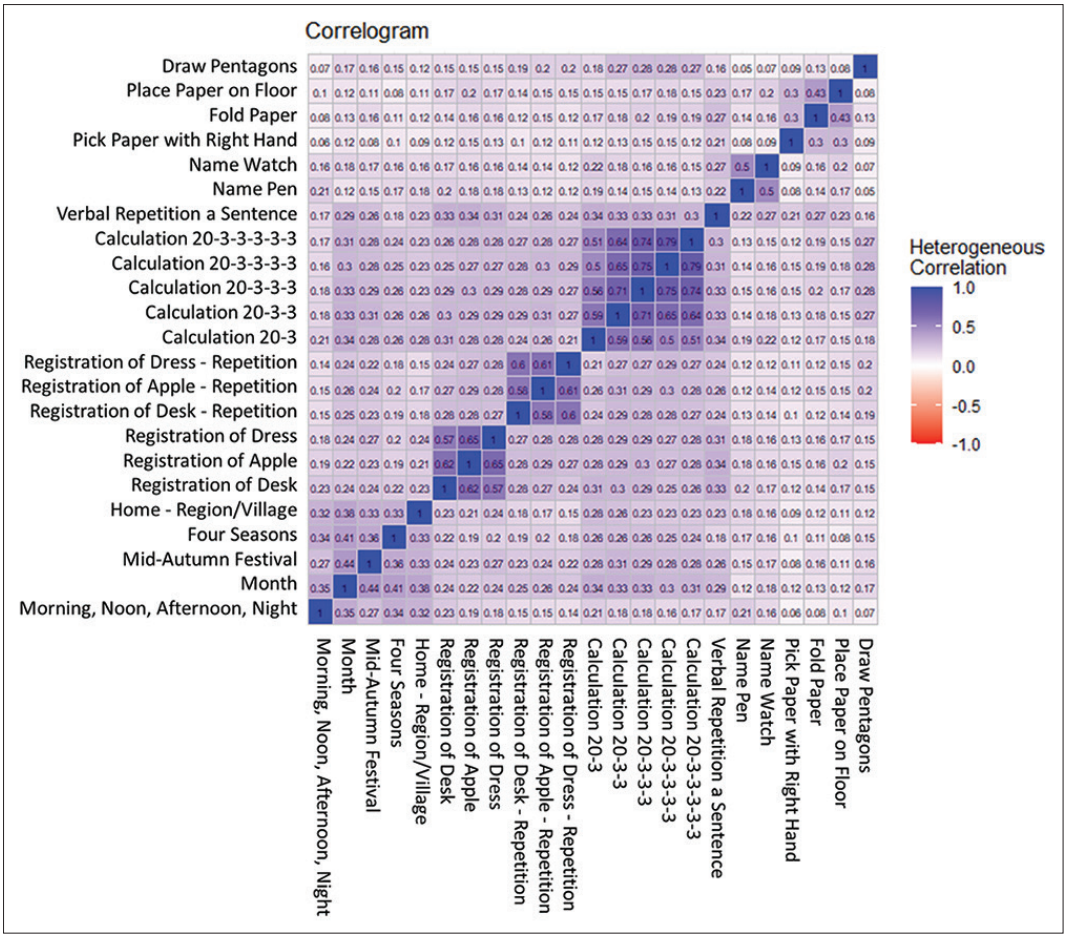

Figure A1. Correlogram with correlation coefficient printed.

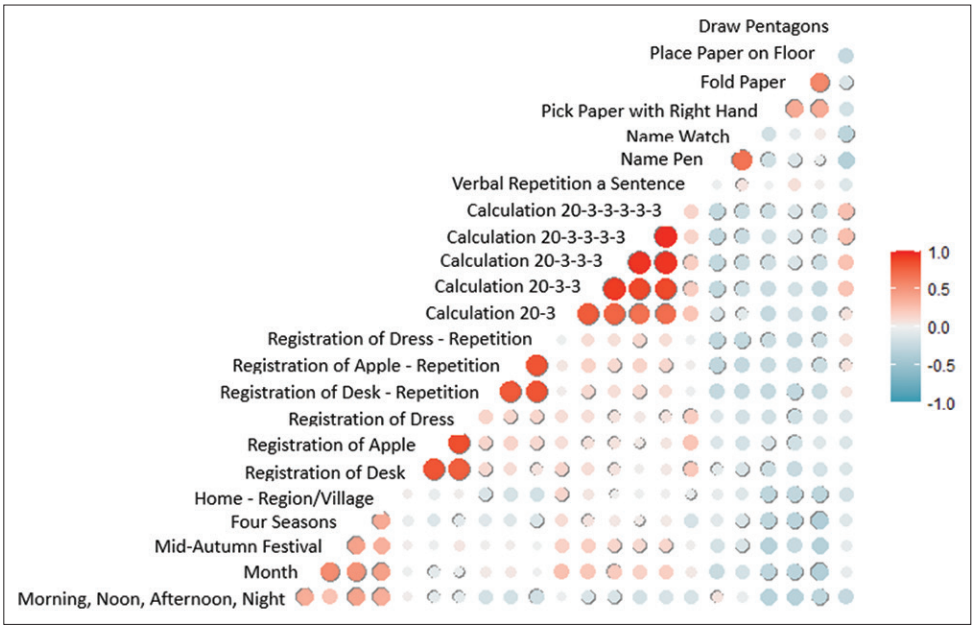

Figure A2. Correlogram without correlation coefficient.

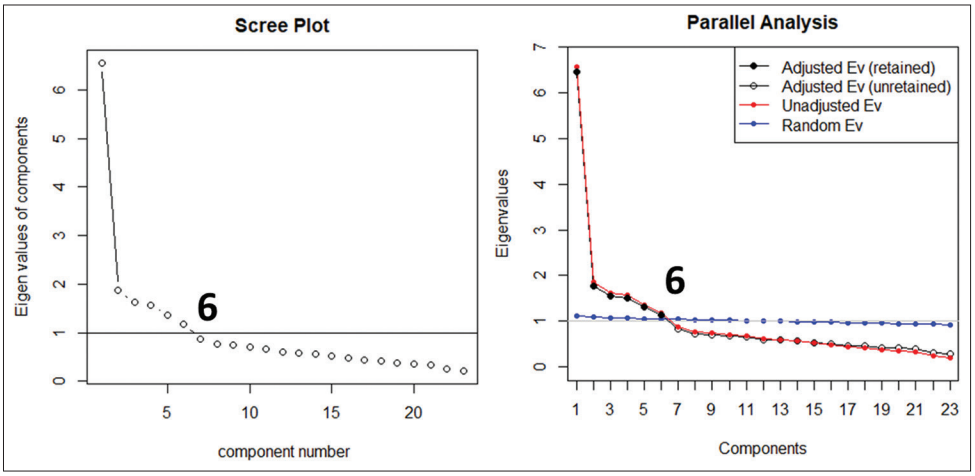

Figure A3. Scree plot and Horn's parallel analysis.

Table A1. Indicators to the determination of the number of factors.

No. of factors	RMSEA	BIC	SABIC	map	vss1	vss2
1	0.121	26100	26831	0.022	0.71	0
2	0.098	14752	15413	0.017	0.55	0.76
3	0.085	9703	10297	0.018	0.52	0.74
4	0.067	4938	5469	0.019	0.5	0.68
5	0.051	2055	2526	0.016	0.51	0.70
6	0.036	351	764	0.018	0.53	0.70
7	0.022	−444	−85	0.024	0.53	0.67

RMSEA, root mean square error of approximation; BIC, Bayesian information criterion; SABIC, sample-adjusted BIC; map, Velicer's MAP criteria; vss1, very simple structure 1; vss2, very simple structure 2.

Table A2. Number of factors, eigenvalue, % and cumulative % variance explained.

No. of factors	Eigenvalue	% of variance	Cumulative % of variance
1	6.56	28.52	28.52
2	1.87	8.12	36.63
3	1.62	7.04	43.67
4	1.57	6.82	50.49
5	1.36	5.93	56.42
6	1.19	5.15	61.57
7	0.87	3.77	65.34
8	0.76	3.30	68.64
9	0.74	3.20	71.84
10	0.70	3.03	74.87
11	0.67	2.91	77.78
12	0.61	2.63	80.42
13	0.58	2.53	82.95
14	0.56	2.45	85.40
15	0.52	2.25	87.65
16	0.48	2.07	89.72
17	0.43	1.88	91.60
18	0.42	1.80	93.40
19	0.37	1.63	95.03
20	0.36	1.55	96.58
21	0.34	1.46	98.04
22	0.24	1.05	99.10
23	0.21	0.90	100.00