Psychometric Characteristics of the Mini-Mental State Examination among Older Chinese Immigrants

Fengyan Tang[1]*, Ph.D.; Laura E. Gibbons[2], Ph.D.; Beth E. Snitz[3], Ph.D.; XinQi Dong[4], MD

[1]School of Social Work, University of Pittsburgh, Pittsburgh, PA, USA; [2]General Internal Medicine, University of Washington, Seattle, WA, USA; [3]School of Medicine, University of Pittsburgh, PA, USA; [4]Institute for Health, Health Care Policy and Aging Research, Rutgers, The State University of New Jersey, USA

ABSTRACT

Backgrounds: The recent growth of older adult immigrants and the increased burden of Alzheimer's disease and related dementias (ADRD) call for validation of the assessment tools and timely detection of cognitive impairment. In particular, older Chinese Americans are among the fastest growing populations in the U.S., who are at high risk for delayed diagnosis and deficient management of ADRD due to limited knowledge of dementias and barriers to healthcare access. The value of the MMSE as a screening tool in detecting ADRD is well documented in western countries. Yet, little is known about the psychometric characteristics of a Chinese version of the MMSE (CMMSE) in older Chinese immigrants, whose cognitive test performances may relate to cultural and linguistic differences. **Purpose:** The study aims to assess factor structure and psychometric properties of CMMSE among older Chinese immigrants in the United States. **Methods:** Participants were recruited from the greater Chicago area in the Pop-ulation-based Study of Chinese Elderly (PINE). The sample included 3,126 respondents aged 60 or over, who selected Chinese dialects in completing face-to-face interviews. A bifactor analysis was conducted to test unidimensionality of the CMMSE. Rasch modeling was used to assess scale properties. Age, sex, and education were tested for dif-ferential item functioning (DIF). **Results:** A bifactor analysis identified a unidimensional general factor with 30 items, reflecting a single un-derlying construct of cognitive capacity. We also found five concurrent subdomains with 19 items. The general factor explained 71 percent of the variance in the CMMSE. The Rasch model provided evidence of construct validity with acceptable item fit statistics. The DIF analysis showed that item biases were generally negligible. **Conclusion:** The CMMSE appeared to be a valid, reliable screening instrument for cog-nitive impairment in the US Chinese older adults. Findings provided support for the use of the CMMSE in detecting cognitive impairment among older Chinese immigrants. Although the five-factor solution has been replicated, item compositions of subdomains remain controver-sial. Future efforts are needed to improve neuropsychological assess-ment accuracy by developing a culturally and linguistically appropriate assessment tool in minority older populations.

INTRODUCTION

The Mini-Mental State Exam (MMSE)[1] is one of the most widely used screening instruments for detecting cognitive impairment and quantifying its severity. [2,3] The MMSE assesses five cognitive domains: orientation to time and place, registration, attention, recall, and language. [1] The scale is available in many languages and has been validated in various study populations. [2] Numerous investigations have been conducted to test its factor structure and predictive value. [4,5] The predictive validity and reliability are found to be generally acceptable using classical test theory (CTT) approaches. [5] However, its psychometric properties, particularly its internal structure remains controversial across countries and cultural groups. [6,7]

Previous studies using factor analysis procedures have documented two-, three-, or five-factor structure for the MMSE in various study samples. [6,8,9,10] Overall, cognitive abilities or subdomains of orientation, attention, and recall have been commonly found with various item combinations, but registration and language have not been consistently reproduced. [9] Different factor structures are partly attributable to cultural difference and language of test administration. [11] Ethnic origin and education also affect performance on the cognitive tests and contribute to overestimation of cognitive impairment among racial/ethnic minorities, the less educated, or non-English speakers. [12]

The recent growth of older adult immigrants and the increased burden of Alzheimer's disease and related dementias (ADRD) call for validation of the assessment tools and timely detection of cognitive impairment in the older population. In particular, Chinese Americans are among the fastest growing populations in the U.S., and 14 percent of them are aged 65 or older. [13] They are at high risk for delayed diagnosis and deficient management of ADRD due to limited knowledge of dementias and barriers to timely diagnosis, treatment, and healthcare access. [14]

The value of the MMSE as a screening tool in detecting ADRD is well documented in western countries. [3] Yet, little is known about the psychometric characteristics of a Chinese version of the MMSE (CMMSE) in older Chinese immigrants, whose cognitive test performances may relate to cultural and linguistic differences in addition to the well-known age and educational effects. It is noted that cognitive assessment of minority older adults present unique challenges due to the varying extents of exposure to the dominant culture and proficiency in the language utilized in the test administration. [15] Therefore, it is especially important to evaluate the psychometric properties of the translated MMSE and analyze item performance in a particular older immigrant population.

The Current Study

The purpose of the current study was to assess factor structure and psychometric properties of the CMMSE in a sample of older Chinese immigrants. Specifically, the first objective of the study is to test factor structure by fitting a bifactor model in the CMMSE. A bifactor structural model specifies that the covariance among a set of item responses is accounted for by a single general factor that reflects the common variance among all scale items and by specific factors that reflect additional variance among cluster of items. [16] Bifactor modeling has been found as an effective approach to modeling unidimensionality that arises in a broad construct, at the same time allowing specific subdomains of item contents. [16] That is, bifactor models account for residual correlations when a general factor contains a couple of related subdomains, and explore the extent to which items reflect both a general factor and subdomains as well, [17,18] thus providing superior parameter recovery in comparison with unidimensional models to handle dependencies of local items. [19] Results from the bifactor models will indicate whether the data support the scale as a unidimensional construct or the use of a multidimensional structure and scale. [16]

Next, to the extent that a unidimensional scale with related subdomains is supported in the CMMSE data, our second objective is to test item difficulty and construct validity using the item response theory (IRT) model. In recent years, an increasing number of studies have applied IRT models to conduct psychometrical evaluations and assess dimensional constructs. IRT is believed to be more efficient and accurate than CTT in a wide variety of substantive contexts, providing considerably more psychometric information at the item response level. [18] In the case of the MMSE, IRT models provide more accurate estimations of cognitive impairment by including the level of item difficulty in the design and scaling of the measurement. [20,21] The Rasch or one-parameter model has been applied in the MMSE, assessing and establishing a hierarchy of item difficulty and the extent to which respondents are separated by cognitive ability level. [7]

The third objective is to detect possible differential item functioning (DIF) for bias of items across different age, sex, and education groups. DIF is another function provided by IRT, which helps determine the presence of bias in assessment of cognitive abilities across various subgroups when the subgroups have the same scores on the latent trait. [12] For example, when a group of men and women have the same sum score on the MMSE, they should have the same probability of endorsing any specific test item; otherwise, DIF occurs when the parameters of a given item differ across groups, indicative of item bias and lack of validity. [7] Previous studies have documented DIF in the MMSE attributable to educational attainment, sex, age, or language of administration. [7,10,22,23] Overall, IRT approaches are appropriate for identifying and eliminating language- and culture-related biases at both item and test levels, constructing tests with equivalent reliability at all ability levels. [12]

METHODS

Chinese Version of the MMSE (CMMSE)

The 30-item MMSE was translated from English to Chinese, then translated back to English and compared with the original version by a bilingual research team. [24] Items with discrepancies were further modified. The test was administrated in English, Mandarin, or Cantonese, according to the participant's request. [24] A previous study showed that a Cantonese version with some item adaptations has been validated in a sample of older adults in Hong Kong, and the scale demonstrated good reliability (Cronbach's alpha = 0.86; test-retest reliability = 0.78; inter-rater reliability = 0.99), and discrimination validity against clinical diagnoses (canonical correlation = 0.94). [25]

Participants

Participants were recruited from the greater Chicago area in the Population-based Study of Chinese Elderly (PINE), conducted between 2011and 2013. [26] Guided by a community-based participatory research approach, the PINE is a population-based epidemiological study of the health and well-being among adults aged 60 and over, who self-identified as Chinese in the U.S. (N=3,157). [26] The study sample included 3,126 (99%) respondents who selected Chinese dialects in completing face-to-face interviews, and those selecting English in interviews were excluded ($n = 31$). The average age of the sample respondents was 72.8 years ($SD = 8.3$, range: $59 - 105$), and about 39% were 75 years or older. The average years in school was 8.7 ($SD = 5.1$, range: $0 - 26$), and 58% were female.

Statistical Analysis

To test factor structure and unidimensionality, two bifactor models were estimated, comparing the originally designed structure with five subdomains [1] and a modified four-subdomain solution with 29 items. [5] The bifactor model fits the data when the test items are loaded on two separate factors: the general factor and one specific subdomain. Comparison of model fit was based on Chi-square, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), Akaike's Information Criterion (AIC), and Bayesian information criterion (BIC). Adequate

fit is indicated by the CFI and TLI values greater than 0.95, [27] and the RMSEA less than 0.06. [28] Smaller AIC and BIC values indicate better model fit.

We further estimated essential unidimensionality using the explained common variance (ECV), that is, the percentage of explained variance by the general factor out of the total variance explained by both the general factor and the specific factors. [16] For this purpose, we also applied the index coefficient omega hierarchical (ωH), an appropriate model-based reliability index when item response data are consistent with a bifactor structure. [16] It estimates the amount of variance that can be attributed to a single general factor. Higher ECV and ωH values indicate a stronger general factor relative to the subdomain factors and thus, a more confidence in applying a unidimensional IRT model. [29]

For the IRT analysis, we used the one-parameter Rasch model to assess item difficulty and construct validity. The Rasch model predicts the probability of a correct response to an item given the individual ability and the item difficulty. The assumptions of Rasch models include unidimensionality (i.e., the items are explained by a single underlying construct) and local independence (i.e., shared item variance is accounted for by the unidimensional construct. [30] We assessed the fit of each item using infit and outfit values, which indicate how accurately data fit the model, or how well item difficulty and individual ability contribute to the latent trait of the test. [31,32] According to Linacre (2002), [29] infit and outfit values between 0.5 and 1.5 were deemed acceptable, values of between 1.5 and 2.0 were deemed moderate, and values less than 0.5 or greater than 2.0 were considered unacceptable.

Lastly, we conducted DIF analysis to assess the measurement equivalence and examine whether the probability of passing an item is different for various subgroups, when the subgroups have the same or matched scores on the latent trait. The logistic regression model was carried out for an overall test of DIF effects between the comparison groups of age (older than 75 versus 75 or younger), sex (male versus female), and education (less than nine years versus nine years or above). The logistic regression model is powerful and flexible in detecting items with DIF. [33] The DIF estimates are classified according to effect size as "A" (negligible effect), "B" (moderate effect), and "C" (large effect"). [32]

RESULTS

Following the theoretical framework and empirical evidence, we estimated and compared two bifactor structural models: five subdomains with 30 items, [1] and four subdomains (without registration) with 29 items (without Item 11-repeat: apple). [5] Although the CFI, TLI, and RMSEA were comparable between two models, the 5-subdomain model (χ^2 = 2345.56, df = 376, CFI = 0.90, TLI = 0.89, RMSEA = 0.04, AIC = 12391.82, BIC = 12925.21) was considered better fit given the smaller AIC and BIC values as compared with the 4-subdomain model (χ^2 = 2213.18, df = 350, CFI = 0.90, TLI = 0.89, RMSEA = 0.04, AIC = 17865.44, BIC = 18369.20). When further examining the parameter estimators, we found that the subdomain factor structures did not hold, particularly for the Orientation to Time and Language, as items 1-5, 22-24, 26-28 had negative factor loadings, indicating the model had some difficulty converging. We removed these items from the subdomains but not from the general factor, which would have little impact on the remaining measurement model, since these items contributed very little variance to the model. The final five-subdomain solution fit the data with 30 items in the general factor and19 items in the subdomains (see Figure 1). The robust maximum likelihood estimation was then used and model fit indices were improved (χ^2 = 736.03, df = 386, CFI = 0.99, TLI = 0.99, RMSEA = 0.02).

Table 1 presents that the majority of items had sizable slopes that defined their relationships with the general factor, with standardized

Table 1. Bifactor Analysis and a Rasch Model of the CMMSE.

Item #/ Factor	Item	Bifactor Loadings		Rasch Model Statistics		
		General factor	Subdomain factors	Difficulty β	Outfit MSQ	Infit MSQ
Orientation						
6	State	0.56	0.43	0.92	0.73	0.89
7	County	0.35	0.37	3.41	0.99	1.00
8	City	0.39	0.11	-1.07	0.92	0.97
9	Room	0.29	0.08	-1.49	1.05	1.1
10	Address	0.44	0.32	1.00	1.08	1.06
Registration						
11	Repeat: Apple	0.31	0.65	-3.12	**0.16**	0.80
12	Repeat: Newspaper	0.40	0.43	-1.62	0.64	0.90
13	Repeat: Train	0.46	0.20	-0.94	0.78	0.82
Attention-calculation						
14	100-7	0.52	0.13	-1.03	0.51	0.81
15	93-7	0.45	0.46	1.62	0.94	0.96
16	86-7	0.47	0.54	1.48	0.86	0.82
17	79-7	0.50	0.58	1.32	0.76	0.87
18	72-7	0.48	0.55	1.72	0.83	0.89
Recall						
19	Recall: Apple	0.43	0.11	0.37	0.96	1.04
20	Recall: Newspaper	0.40	0.22	1.83	1.09	1.09
21	Recall: Train	0.33	0.20	2.51	1.24	1.19
Language						
25	Follow direction	0.63	0.42	0.01	**0.48**	0.80
29	Write a sentence	0.60	0.53	0.73	0.75	0.85
30	Drawing copy	0.51	0.17	0.87	0.94	0.98
Not in subdomains						
1	Year	0.62	NA	-0.34	0.57	0.79
2	Season	0.44	NA	-0.58	0.98	0.99
3	Date	0.40	NA	1.01	1.24	1.15
4	Day of week	0.47	NA	-0.91	0.79	0.91
5	Month	0.52	NA	-1.08	0.56	0.80
22	Recognition: wristwatch	0.29	NA	-2.70	0.66	0.91
23	Recognition: pencil	0.19	NA	-3.46	0.70	0.94
24	Immediate repetition "NO IFS, ANDS, OR BUTS"	0.44	NA	-0.16	0.99	1.04
26	Paper in right hand	0.24	NA	-0.03	**1.64**	1.29
27	Fold paper in half	0.45	NA	-0.09	0.86	1.02
28	Place paper in lap	0.44	NA	-0.19	0.95	1.01

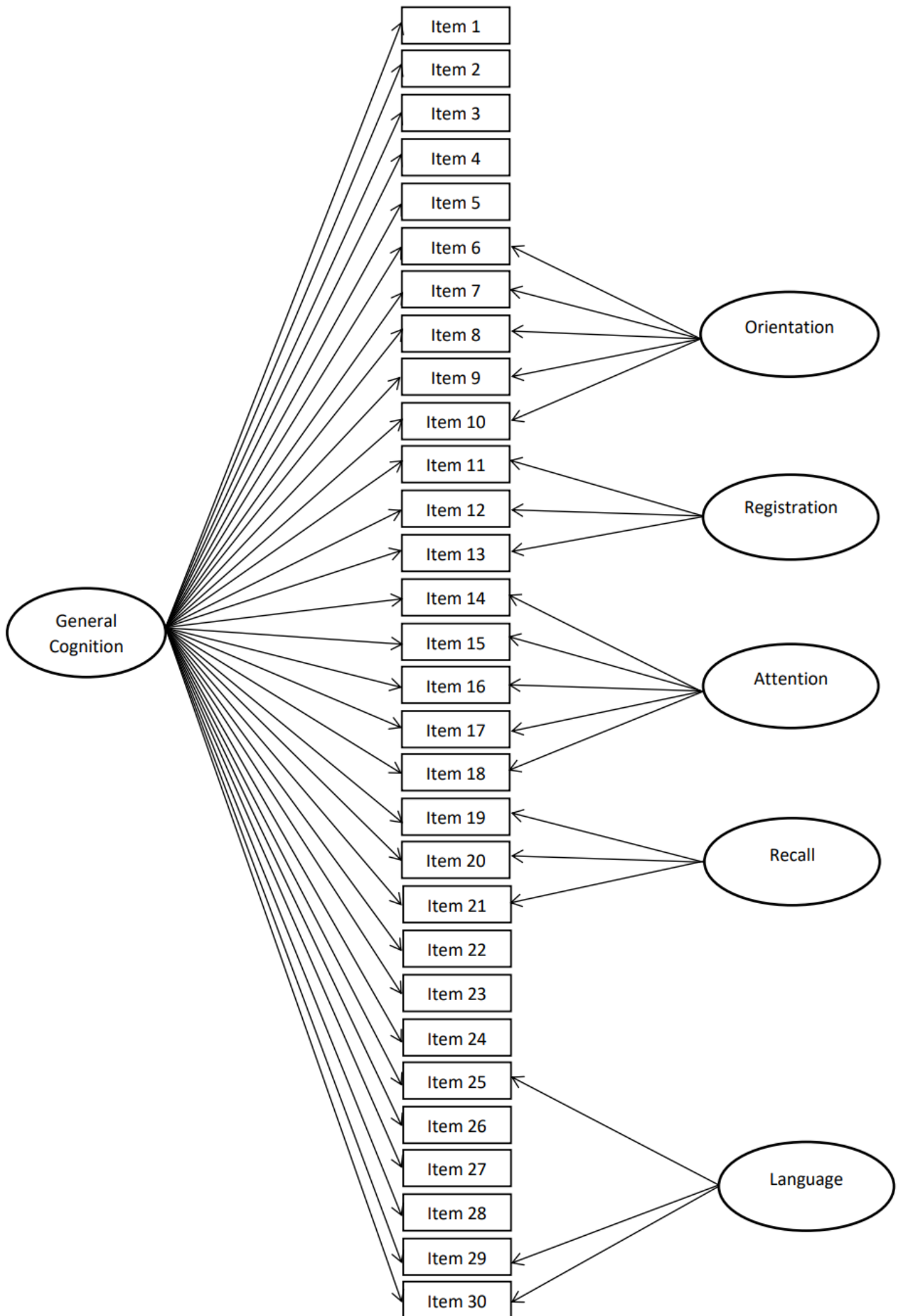Note: Bold statistic indicated outfit values were not within acceptable range.

Figure 1. A Bifactor Model of the CMMSE

coefficients ranging from 0.19 to 0.62. Only Item 23 had a factor loading below 0.20. The factor loadings for five subdomains varied from 0.08 to 0.65. It seemed that three items (19 – 21) did not load well on the Recall subdomain due to the low factor loadings (0.11 – 0.22). Some items (e.g., Items 14, 19, 30) had higher loadings on the general factor than on the subdomain, suggesting they might share more in common with the general factor than with the specific subdomain. Overall, the general factor was strong in comparison with the specific subdomain factors (ECV = 0.69). The general factor of cognitive ability was reliable (ω estimate = 0.91), indicating 91 percent of the variance in the observed scores was due to all sources of the common variance. The general factor alone explained about 71 percent of the variance (ωH = 0.71).

Table 1 also presents the summary of the Rasch model results. Item 7 ("what county are we in?") was the most difficult, followed by Item 21 (recall: train) and Item 20 (recall: newspaper). By contrast, Item 23 (recognition: pencil) was the easiest, followed by Item 11 (repeat: apple), and Item 22 (recognition: wristwatch). Regarding item fit, Item 11 (immediate recall: apple), Item 25 (follow direction: read words on card), and Item 26 (follow direction: paper in right hand) showed indication of bad fit due to the outfit statistics (either <0.5 or >1.5), indicating that they were less productive for the measurement construction. But infit indices for all items were acceptable.

The DIF analysis showed that 14 items exhibited DIF by age groups, 16 items with DIF by sex, and 23 items with DIF by education groups. However, the effect sizes of the results were negligible ("A") for all items, except that Item 23 had a large effect size ("C") by sex groups. It seemed that the overall DIF or item biases are negligible in the CMMSE.

DISCUSSION

The 30-item CMMSE appears to be a valid, reliable instrument for screening cognitive impairment in the US Chinese older adults. As suggested, a bifactor model was estimated to handle dependencies of scale items and provide accurate model estimates. [5] Our study documented that the bifactor model fit the CMMSE data well, identifying a general factor that reflects a single underlying construct of broad cognitive capacity, as found in previous research. [2,5] Therefore, we can conclude that the CMMSE is essentially unidimensional, but with five specific subdomains as well. The findings of a strong general factor and concurrent subdomains may explain the mixed findings regarding factor structure in previous factor analysis studies. [6,8,9,10]

Although the five-factor solution has been replicated in this study, item compositions of five subdomains still remain controversial. Five items of orientation to time were dropped from the subdomain, yet they still contributed to the general factor. We had conducted additional analysis to test whether a six-factor solution fit the data, with the added orientation-to-time subdomain, but findings did not support the structure. Thus, we retained the original five-factor structure, dropping the Orientation-to-time items as well as six items from the Language subdomain. Findings indicated that the Language subdomain is less cohesive or less cross-cultural replicable. For example, the phase "No ifs, ands, or buts" is not easily translated, and its articulation complexity can be lost in the translation into other languages. [3] Lack of consistency in item compositions warrants further research into distinct components and cross-cultural equivalence of the MMSE. [9,10]

The Rasch model provided further evidence of construct validity, and it appeared to fit well for the majority of the items, except a few items with unacceptable outfit statistics. Yet, it should be noted that outfit statistics are susceptible to the influence of outliers and that for this reason, infit statistics are generally more emphasized over the outfit statistics. [32] Our analysis showed that all infit values were within the recommended acceptable range, suggesting that the CMMSE

fit well with the Rasch model and that the items indeed assess the latent trait of cognitive status. [7]

The IRT analysis also established a hierarchy of item difficulty, with Item 7 (naming the residence county) being the most difficult. This is probably because of the issues related to language translation and transfer across cultures that can have a major impact on the psychometric properties of a test. [34] Compared with the state, city, and street address, a county name may be less mentioned or familiar to older adult immigrants. The item bias occurred as it could be more difficult for a non-English speaker to memorize a foreign county name which does not make sense to them. The language issues may exist in other items of orientation to place, making it distinct from orientation to time. The use of ten orientation items may be excessive and invalid. [35] In addition, delayed recall and attention items were found to be more difficult than other subdomain items in this study sample of older Chinese immigrants, which is consistent with previous studies in other populations. [7,35] The easiest items included recognition of objects and immediate repeating words, also corroborating previous findings. [7,35] Although we found the presence of DIF in some items that may be attributable to age, sex, and education, the effect sizes were generally negligible, suggesting that item bias does not appear to be a major source of observed differences in cognitive status by these groups. [22] Therefore, the CMMSE can be considered a valid assessment tool.

Some limitations of this study should be noted. First, the study sample was restricted to the greater Chicago area, thus limiting the external validity of the research. Second, immigration experience and cultural adaptation process may affect cognitive test performance, for example, some orientation and language items could be challenges for recent immigrants due to cultural, idiomatic, and geographic differences. In addition, the CMMSE was administrated with different Chinese dialects, including Mandarin,

Cantonese, and Taishanese, which may have impact on test performance across language groups. The MMSE was always translated with linguistic and cultural adaptations when it was administrated in a foreign country to assure cultural competence with the instrument. [3,36] Although it may not be the case for immigrants to English-speaking host countries, a direct translation, as used in the present study, may make performance on the MMSE less meaningful for those with little or no education. [3,36]

In summary, this study is the first comprehensive psychometric investigation of the CMMSE in a representative sample of older Chinese immigrants. Findings bearing on the factor structure and item characteristics have both clinical and research implications since the study adds to the body of the literature on the validity and reliability of the scale. We found support for a bifactor model that indicated a generally unidimensional cognitive ability and supported the use of a sum or a latent score across all items. We also found five concurrent subdomains, as theorized by Folstein et al. [1] Identification of the underlying cognitive subdomains that reflect the cognitive process may help clinicians to detect different stages of cognitive decline and subtypes of cognitive impairment. [35] The IRT analysis identified few statistically problematic items, and item bias is generally ignorable across age, sex, and education groups. The order of item difficulty is mostly comparable to the findings in other populations. Future efforts are needed to improve neuropsychological assessment accuracy for identifying cognitive decline and dementias through the use of a culturally and linguistically appropriate assessment tool in diverse older populations.

Acknowledgement

REFERENCE

[1]. Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method

for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

[2]. Schultz-Larsen, K., Kreiner, S., & Lomholt, R. K. (2007). Mini-Mental Status Examination: Mixed Rasch model item analysis derived two different cognitive dimensions of the MMSE. *Journal of Clinical Epidemiology*, *60*, 268–279. http://dx.doi.org/10.1016/j.jclinepi.2006.06.007

[3]. Shim, Y.S., Yang, D.W., Kim, H-J, Park, Y.H., and Kim, S. (2017). Characteristic differences in the mini-mental state examination used in Asian countries. BMC Neurology, 17:141, 1-9. DOI 10.1186/s12883-017-0925-z

[4]. Teresi, J.A. (2007). Mini-Mental State Examination (MMSE): Scaling the MMSE using item response theory (IRT). *Journal of Clinical Epidemiology, 60*, 256-259. https://doi:10.1016/j.jclinepi.2006.06.009

[5]. Rubright, J.D., Nandakumar, R., & Karlawish, J. (2016). Identifying an appropriate measurement modeling approach for the Mini-Mental State Examination. *Psychological Assessment, 28*, 125-133. http://dx.doi.org/10.1037/pas0000146

[6]. Ideno, Y., Takayama, M., Hayashi, K., Takagi, H., and Sugai,Y. (2012). Evaluation of a Japanese version of the mini-mental state examination in elderly persons. Geriatrics and Gerontology International, 12, 310-316. https://doi:10.1111/j.1447-0594.2011.00772.x

[7]. Melo, D. M., Barbosa, A.J.G., Castro, N.R. et al. (2020). Mini-mental state examination in Brazil: An item response theory analysis. *Psychological Evaluation, 30*, e3014, 1-10. http://dx.doi.org/10.1590/1982-4327e3014

[8]. Brugnolo, A., Nobili, F., Barbieri, M. P., et al. (2009). The factorial structure of the Mini Mental State Examination (MMSE) in Alzheimer's disease. *Archives of Gerontology and Geriatrics, 49*, 180–185. http://dx.doi.org/10.1016/j.archger.2008.07.005

[9]. Castro Costa, E., Dewey, M. E., Uchôa, E., Firmo, J. O., Lima-Costa, M. F., & Stewart, R. (2014). Construct validity of the mini mental state examination across time in a sample with low-education levels: 10-year followup of the Bambui Cohort Study of Ageing. *International Journal of Geriatric Psychiatry, 29*, 1294-1303. https://doi:10.1002/gps.4113

[10]. Jones, R. N., & Gallo, J. J. (2000). Dimensions of the Mini-Mental State Examination among community dwelling older adults. *Psychological Medicine, 30*, 605–618. http://dx.doi.org/10.1017/S0033291799001853

[11]. Shyu, Y.L. & Yip, P. (2001). Factor structure and explanatory variables of the Min-Mental State Examination (MMSE) for elderly persons in Taiwan. *J Formos Med Assoc*, 100, 676-683. PMID: 11760373

[12]. Mungas, D., Reed, B. R., Marshall, S. C., & González, H. M. (2000). Development of psychometrically matched English and Spanish language neuropsychological tests for older persons. *Neuropsychology, 14,* 209–223. http://dx.doi.org/10.1037/0894-4105.14.2.209

[13]. U.S. Census Bureau (2016). Asian American Data Links. Retried January 11, 2018 from www.census.gov/about/partners/cic/resources/data-links/asian.html

[14]. Woo, B.K. (2017). Dementia health promotion for Chinese Americans. Cureus, 9, e1411. https://doi:10.7759/cureus.1411

[15]. Wrobel, N.H., and Farrag, M.F. (2007). Preliminary validation of an Arabic version of the MMSE in the elderly. Clinical Gerontologist, 75-93. https://doi.org/10.1080/07317110802072223

[16]. Reise, S. P. (2012). The rediscovery of bifactor measurement models. Multivariate Behavioral Research, 47, 667–696.

[17]. Chen, F.F., West, S.G., and Sousa, K.H. (2006). A Comparison of Bifactor and Second-Order Models of Quality of Life. Multivariate Behavioral Research, 41(2), 189-225. DOI: 10.1207/s15327906mbr4102_5

[18]. Reise, S. P., Morizot, J. and Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Qual Life Res, 16,19–31. DOI 10.1007/s11136-007-9183-7

[19]. Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. Journal of Educational Measurement, 47, 361–372.

[20]. Crane, P.K., Narasimhalu, K., Gibbons, L., Pedraza, O., Mehta, K.M., Tang, X. et al. (2008). Composite scores for executive function items: Demographic heterogeneity and relationships with quantitative MRI. J Int Neuropsychol Soc., 14(5), 746–759. doi:10.1017/S1355617708081162.

[21]. Meade, A.M. and Lautenschlager, G.J. (2004). Same question, different answers: CFA and two IRT approaches to measurement invariance. Symposium presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

[22]. Jones, R. N., and Gallo, J. J. (2002). Education and sex differences in the Mini-Mental State

Examination: Effects of differential item functioning. The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 57B(6), P548–P558. http://dx.doi.org/10.1093/geronb/57.6P548

[23]. Jones, R. N. (2006). Identification of Measurement Differences Between English and Spanish Language Versions of the Mini-Mental State Examination: Detecting Differential Item Functioning Using MIMIC Modeling. Medical Care, 44 (11), S124-S133.

[24]. Chang, E-S. & Dong, X. (2014). A battery of tests for assessing cognitive function in U.S. Chinese older adults – findings from the PINE study. J Gerontol A Biol Sci Med Sci., 69A (S2), S23–S30. doi:10.1093/gerona/glu195

[25]. Chiu, H.F., Lee, H.C., Chung, W.S., and Kwong, P.K. (1994). Reliability and validity of the Cantonese version of mini-mental state examination-a preliminary study. Hong Kong Journal of Psychiatry, 4 (SP2), 25-28.

[26]. Dong, X., Wong, E., & Simon, M.A. (2014). Study design and implementation of the PINE study. *Journal of Aging and Health, 26*, 1085-1099.

[27]. Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

[28]. Browne, M.W. & Cudeck R. (1993). Alternative Ways of Assessing Model Fit. In: Bollen K, Long J, editors. Testing Structural Equation Models. Sage; Newbury Park, CA: 1993. pp. 136–162.

[29]. Linacre, J. M. (2002). What do infit and outfit, mean-squared and standardized mean? Rasch Measurement. Transactions,16, 878. https://www.rasch.org/rmt/rmt162f.htm

[30]. Edelen. M.O. and Reeeve, B.B. (2007). Appling item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res, 16, 5–18. DOI 10.1007/s11136-007-9198-0

[31]. Bond, T. G. and Fox, C. M. (2007). *Applying the Rasch Model*, 2nd ed. Mahwah, NJ: Lawrence Erlbaum.

[32]. Molde, H., Hynninen, K.M., Torsheim, T. et al. (2017). A bifactor and item response analysis of the geriatric anxiety inventory. *International Psychogeriatrics, 29,* 1647-1656. https://doi.org/10.1017/S1041610217001004

[33]. Swaminathan, H. and Rogers, H.J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement, 27 (4), 361-370.

[34]. Carnero-Pardo, C. (2014). Should the Mini-Mental State Examination be retired? Neurologia, 29 (8), 473-481. https://doi:10.1016/j.nrl.2013.07.003

[35]. McGrory, S., Doherty, J. M., Austin, E. J., Starr, J. M., & Shenkin, S. D. (2014). Item response theory analysis of cognitive tests in people with dementia: A systematic review. *BMC Psychiatry, 14*, 47. https://doi:10.1186/1471-244X-14-47

[36]. Steis, M.R. & Schrauf, R. W. (2009). A review of translation and adaptations of the Mono-Mental State Examination in languages other than English and Spanish. Research in Gerontological Nursing, 2 (3), 214-224. doi: 10.3928/19404921-20090421-06.