ORIGINAL ARTICLE

Measurement equivalence of the Kessler 6 Psychological Distress Scale for Chinese and Korean immigrants: Comparison between younger and older adults

En-Jung Shon [ORCID]

Department of Family Science and Social Work, College of Education, Health and Society, Miami University in Ohio, Oxford, Ohio

Correspondence
En-Jung Shon, Department of Family Science and Social Work, College of Education, Health and Society, Miami University in Ohio, 101E McGuffey Hall, Oxford, OH.
Email: shone@miamioh.edu

Abstract

Objectives: The Kessler 6 (K6) Psychological Distress Scale is a well-known instrument to screen for psychological distress of general populations. It is critical to perform the equivalence test of the K6 for Asian immigrant subgroups.

Methods: The 2012 California Health Interview Survey data were used (N = 1,210; Chinese = 640, Koreans = 570). Among 1,210, 734 were younger (18–64 years) and 476 were older (65+) adults. It was examined whether parameters in the measurement model is equivalent across the two groups, using multiple-group analysis. The equivalence tests for Chinese and Koreans were separately performed based on different age groups (younger [18–64] vs. older [65+]).

Results: The younger group had good model fit (X^2 = 41.27 [df = 16, p = .001], X^2/df = 2.58, Comparative Fit Index [CFI] = 0.99, Goodness of Fit Index [GFI] = 0.98, root mean square error or approximation [RMSEA] = 0.05, standardized root mean residual [SRMR] = 0.03), and the older group also showed good model fit (X^2 = 41.70 [df = 16, p < .001], X^2/df = 2.61, CFI = 0.98, GFI = 0.97, RMSEA = 0.06, SRMR = 0.04). The model for older group indicated measurement noninvariance between Chinese and Korean immigrants (ΔX^2 = 17.86, Δdf = 5, p = .003, CFI = 0.972, ΔCFI = 0.009). The items "hopeless," "restless," and "depress," were significantly nonequivalent between the two groups.

Conclusions: Clinicians/researchers should be aware of the potential risk for misclassification when they screen psychological distress of Chinese or Korean older immigrants. Professionals should pay attention to cross-cultural comparability when interpreting results from the K6.

KEYWORDS
Asian immigrant subgroups, K6, measurement equivalence, psychological distress, younger and older Asian immigrants

1 | INTRODUCTION

1.1 | Psychometric properties of Kessler 6

Asian immigrants are more exposed to risks of psychological distress compared to nonimmigrant populations because resettling in an alien society could cause stressful experiences in the processes of gaining employment, maintaining family cohesion, and recreating a social network (Drapeau, Marchand, & Beaulieu-Prévost, 2012). In order to address psychological distress issues of Asian immigrants, accurate screening using representative scales should be preceded.

The Kessler 6 (K6) Psychological Distress Scale (Kessler et al., 2002) is a well-known instrument that was developed to screen general populations for psychological distress through national surveys, such as the National Health Interview Survey and the National Household Survey on Drug Abuse (Kim, DeCoster, Bryant, & Ford, 2015). According to Kessler et al. (2002), the K6 offers an "alternative to lengthy diagnostic interviews for measuring severity of symptoms and overall levels of distress." Prior studies have demonstrated that the K6 is strongly related to other gold-standard measures, such as the Diagnostic and Statistical Manual of Mental Disorders (DSM), that evaluate diverse psychological distresses (e.g., mood and anxiety disorders) (Furukawa, Kessler, Slade, & Andrews, 2003; Kessler et al., 2010; Slade, Grove, & Burgess, 2011; Sunderland, Hobbs, Anderson, & Andrews, 2012). As the population in the United States has become culturally diverse (e.g., the growth in immigrant and refugee resettlement; Stolk, Kaplan, & Szwarc, 2014), examining usability of the K6 for culturally diverse groups is very important. The term "culturally diverse" could be used as a collective term for ethnic, racial, migrant, and linguistic groups (The United Nations Educational, Scientific and Cultural Organization, 2009). The K6 scale was initially developed in English; later, researchers translated the K6 to evaluate the psychological distress of diverse target populations, including Arabic, Chinese, Italian, Japanese, and Spanish populations. However, many scholars have raised concerns about the cross-cultural comparability of the K6 (Andersen et al., 2011; Kim, Bryant, & Parmelee, 2012).

Although several studies (Andersen et al., 2011; Kim, 2010; Kim et al., 2015; Mitchell & Beals, 2011) have evaluated the performance of the K6 in screening for psychological distress in diverse racial or ethnic groups, few studies included disaggregated Asian subgroups as target populations; rather, many prior studies treated Asian subgroups as just one "Asian" group. Besides, the findings from the prior studies including aggregated Asian groups have been inconclusive. Kim (2010) and Mitchell and Beals (2011) performed validity evaluations for culturally diverse populations and reported that the K6 was an appropriate screening measure for capturing a level of psychological distress in certain racial and ethnic groups (Asian or American Indian; Kim, 2010; Mitchell & Beals, 2011). In contrast, Andersen et al.'s (2011) psychometric property examination (i.e., measurement equivalence test) concluded that the K6 was less equivalent in screening for the psychological distress of diverse racial and ethnic groups (White, African American, Asian, and Indian). Kim et al., (2015) found evidence for the measurement nonequivalence of the K6 across diverse racial and

ethnic groups (18 years and older, White, African American, Hispanic/Latino, Asian, American Indian): measurements of the functioning of items asking about "nervous," "hopeless," "restless," and "depressed" feelings varied across the different racial and ethnic groups.

Regarding the K6's measurement nonequivalence, Crockett, Randall, Shen, Russell, and Driscoll (2005) argued that the functioning of psychological distress measures could be influenced by cultural predictors (country of origin background, community circumstances, or family environment), which vary by racial and ethnic subgroup (Crockett et al., 2005). With that measurement nonequivalence of the K6, the US Department of Health and Human Services (2001) additionally proposed that the functioning of the psychological distress measures could be sensitive to different age ranges. When examining psychological distress among different age groups, some prior studies have indicated that aging is associated with decreases in the prevalence of mood disorders based on evaluation using the DSM, and this aspect has been replicated from K6 evaluations (Kessler et al., 2010; Trollor, Anderson, Sachdev, Brodaty, & Andrews, 2007). Kim and Choi (2010) also found that the unadjusted 12-month prevalence rate of psychological distress such as depression among older (60 years or older) Asian Americans was significantly lower than for younger people. However, other study has suggested different finding: Kim et al. (2012) compared Asian Americans' psychological distress across different age ranges using the K6 scale, and the findings showed that adults 75 or older indicated significantly more distress than adults aged 65–74.

1.2 | Purpose of this study

Although some studies have conducted measurement equivalence tests across different racial or ethnic groups using the K6, few studies have focused on Asian immigrant subgroups (Andersen et al., 2011; Banh et al., 2012; Kim et al., 2015). Notably, few studies have examined whether the K6 is measurement invariant across Asian immigrant subgroups (Stolk et al., 2014). In the absence of clear information about the measurement equivalence of the K6 across Asian immigrant subgroups, the current study examined whether parameters in the measurement model (K6) are equivalent across the two Asian immigrant groups (Chinese and Korean) using structural equation modeling (SEM; Kim, 2010; Kim et al., 2012).

In addition, few studies were sensitive to age effects when performing measurement equivalence tests of the K6. So, the equivalence tests of the K6 targeting two different Asian immigrant groups were separately performed for two different age groups (young–middle-aged [18–64 years] vs. older adults [65 and older]).

2 | METHODS

2.1 | Measures

The K6 scale consists of six items to assess psychological distress (Kessler et al., 2002). Respondents were asked to respond to the

following question: "During the past 30 days, approximately how often did you feel nervous (Item 1), hopeless (Item 2), restless or fidgety (Item 3), so depressed that nothing could cheer you up (Item 4), that everything was an effort (Item 5), and worthless (Item 6)?" (Kessler et al., 2002). The Cronbach's alpha score for young–middle-aged groups was .85, and the score for older adults was .88.

2.2 | Target sample

The top five countries of birth with over 1 million immigrants living in the United States are China, India, Philippines, Vietnam, and Korea (Gryn & Gambino, 2012). Of the top five Asian immigrant groups in the United States, Chinese and Korean immigrants are coherently considered as East Asian populations. Given that these two groups form the largest portion of immigrants in the United States and have regional similarity (i.e., country of origin), this study included Chinese and Korean immigrants. The current study tried to reduce ethnic heterogeneity in order to conserve statistical power, yet still include groups where measurement noninvariance could be anticipated.

The 2011–2012 California Health Interview Survey (CHIS) was used to obtain an adequate sample size for the target groups (Chinese and Korean immigrants) and relevant response rates for the K6 measure. The CHIS is a population-based random-digit-dialing survey of California residents age 18 or older. The CHIS is a cross-sectional (not panel) design representative of the California population, including diverse racial and ethnic groups (CHIS, 2014). Specific inclusion criteria were as follows: (a) adults ages 18 years and older; (b) foreign-born Chinese or Korean immigrants who responded to the question on ethnicity by indicating only one of the two ethnic groups (initial question: "Would you describe yourself as Native Hawaiian, other Pacific Islander, American Indian, Alaska Native, Asian, African American, or White?" Follow-up question: "You responded with Asian; what specific ethnic group are you, such as Chinese, Filipino, or Vietnamese?"). CHIS 2012 data included 1,210 Chinese and Korean immigrants. Of those, 640 were foreign-born Chinese and 570 were foreign-born Koreans. Out of the total of 1,210, 734 were young–middle-aged (18–64 years) and 476 were older adults (65 and older).

2.3 | Sample description

Table 1 shows sample description. (a) The demographic information for *young–middle-aged adults (less than 65 years old)* is as follows:

TABLE 1 Demographic information of target groups

| | Young adults (18–64 years) | | | | | | Older adults (65 years +) | | | | | |
| | Chinese | | Korean | | | | Chinese | | Korean | | | |
	n	%	*n*	%	X^2	*p* Value	*n*	%	*n*	%	X^2	*p* Value
Male	207	44.5	94	34.9	6.45	0.013	67	38.3	103	34.2	0.80	0.374
Female	258	55.5	175	65.1			108	61.7	198	65.8		
Married	311	66.9	192	71.4	1.60	0.217	97	55.4	146	48.5	2.12	0.155
Unmarried	154	33.1	77	28.6			78	44.6	155	51.5		
Employed	323	69.5	165	61.3	5.05	0.028	13	7.4	16	5.3	0.86	0.427
Unemployed	142	30.5	104	38.7			162	92.6	285	94.7		
Uninsured	74	15.9	99	36.8	42.41	0.000	3	1.7	6	2	2.35	0.308
Public insurance	60	12.9	20	7.4			167	95.4	292	97		
Private	331	71.2	150	55.8			5	2.9	3	1		
0–99% FPL	73	15.7	38	14.1	0.94	0.816	58	33.1	159	52.8	25.00	0.000
100–199% FPL	77	16.6	51	19			49	28	83	27.6		
200–299% FPL	61	13.1	33	12.3			28	16	22	7.3		
300% + FPL	254	54.6	147	54.6			40	22.9	37	12.3		
Good at English	211	57.8	89	38	22.30	<.001	31	20.1	49	16.7	0.83	0.366
Not good at English	154	42.2	145	62			123	79.9	245	83.3		
Born in the United States	129	27.7	46	17.1	10.63	0.001	18	10.3	5	1.7	17.9	0.000
Foreign born	336	72.3	223	82.9			157	89.7	296	98.3		
<15 years	71	21.1	40	17.9	0.86	0.387	13	8.3	9	3	6.10	0.020
≥15 years+	265	78.9	183	82.1			144	91.7	287	97		
	M	SD	M	SD	t	*p* Value	M	SD	M	SD	t	*p* Value
Education (1 = no to 9 = PhD+)	5.27	2.46	5.66	2.08	−2.26	0.024	4.34	2.77	4.04	2.49	1.18	.238

Abbreviations: FPL, federal poverty level; X^2, chi square.

men and women were relatively evenly distributed for Chinese participants, but there was a higher percentage of female in Korean participants. There were more participants who were married, employed, and insured on private insurance for both the Chinese and Korean groups. More than half of Chinese and Korean participants lived at 300% or higher of the federal poverty level (FPL). On average, both Chinese and Korean participants had associate of arts or associate of science levels of education (1 = no formal education~9 = PhD or equivalent). Regarding acculturation status, more Chinese and Korean participants had lived in the United States for 15 years or longer (compared to less than 15 years) and were foreign born. While more Chinese participants perceived their English proficiency as good, more Korean participants perceived their English proficiency as poor. (b) The demographic characteristics of older *adults* (*65 years or older*) are as follows: there were more female and unemployed participants in both ethnic groups. Regarding marital status, there was a relatively even distribution of married and unmarried individuals for both Chinese and Korean participants. A majority of Chinese and Korean participants were covered by public insurance programs. While half of the Chinese participants were under 0–199% of the FPL, almost 80% of Korean participants were under 0–199% of FPL. On average, both ethnic groups had some college-level education (1 = no formal education~9 = PhD or equivalent). In terms of acculturation status, almost 80% of both ethnic groups perceived their English proficiency as poor, and almost 90% of them were foreign born and had lived for 15 years or longer in the United States.

2.4 | Sociodemographic differences between young–middle-aged and older groups

Considering the fact that the K6 measure could be sensitive to different age groups, the current study evaluated possible differences on sociodemographic characteristics between younger (18–64 years) and older (65 years+) groups. Except residential location (region), all remaining sociodemographic characteristics were significantly different between younger and older groups: sex (male or female; X^2 = 3.40, p = .037), education (t = 8.59, p < .001; range = 1 [no formal education]–9 [PhD+]; younger M = 5.41, older M = 4.15), marital (unmarried or married; X^2 = 37.31, p < .001), employment (unemployed or employed; X^2 = 430.36, p < .001), FPL (X^2 = 222.48, p < .001), insurance status (uninsured, public insurance, or private insurance; X^2 = 855.69, p < .001), region (nonurban or urban; X^2 = 1.62, p = .205), perceived English proficiency (poor or good; X^2 = 115.12, p < .001), years lived in the United States (less than 15 years or 15 years or longer; X^2 = 49.32, p < .001). As most of the sociodemographic characteristics were significantly different between younger and older groups, it would be critical to separate/divide the two groups to perform further statistical analyses (i.e., multiple-group analyses). Analyses details are presenting in following sections.

2.5 | Preliminary data evaluation

Univariate frequencies, descriptive statistics, and histograms as well as bivariate scatterplots were examined for outliers, adequate variability, skewness (<2), and kurtosis (<21) (Curran, West, & Finch, 1996): there were no outlier and no missing values for the K6 items, and there was no severe violation in normality (the highest skewness score was 3.33 on the "feeling worthless" item among Chinese), following Curran et al.'s (1996) guidelines. Mardia's normalized estimate of multivariate kurtosis was used to evaluate normality status (Byrne, 2016): in a multivariate normality test with the configural model, kurtosis was 80.65 (critical ratio [CR] = 104.20) in Chinese immigrants and 52.80 (CR = 64.32) in Korean immigrants. These values indicated no problems in the normality of the K6 measure.

Assumptions and conditions for performing SEM were evaluated (Lomax & Hahs-Vaughn, 2013). Specifically, noncollinearity (Pearson's r < .80 or variance inflation factor [VIF] <10) and nonproblematic multivariate outliers (cases with nonextreme discrepancies across the squared Mahalanobis distance scores; Kline, 2016) were expected: bivariate correlation matrices, including sizable r scores (i.e., all r > .30), did not show multicollinearity in Chinese and Korean immigrants. The Mahalanobis distance showed that both groups included one or two potential outliers. However, the sensitivity analysis did not show a significant difference (i.e., sample with the potential outliers vs. sample without the potential outliers), so this study kept the two cases including potential outliers. All VIF values in ordinary linear square (OLS) regression models (e.g., six different OLS models) were under 2.5. Scatterplots between the standardized residuals and scores on the six items showed that residuals fell randomly within a band of ±2 *SDs*. Therefore, the assumption of independence of residuals was satisfied (Lomax & Hahs-Vaughn, 2013).

2.6 | Configural model and covariance-based multiple-group invariance test

2.6.1 | Model specification and identification

A single-factor structure was selected for further analyses in the current study: all loadings of error were fixed as 1, and one-factor loading (i.e., everything is an effort) was fixed as 1. In accordance with prior studies and empirical suggestions (Arnaud et al., 2010; Brooks, Beard, & Steel, 2006), the current study gave a pair of correlations on the error variances of "nervous" and "restless." For identification, the single-factor rule (i.e., at least three indicators) and the T rule, "if n is the number of observed variables in the model, the number of observations equals n (n + 1) 2 when means are not analyzed," were applied (i.e., known information > unknown information indicates overidentification, known information = unknown information indicates just identification, and known information < unknown information indicates underidentification; Kline, 2016). The single-factor rule for model specification was satisfied since K6 includes more than three items. Based on the T rule for identification, the chosen model was overidentified (known = 21 > unknown = 13).

2.6.2 | Configural model and fully constrained model

For the multiple-group invariance test, the configural model (no equality constraints imposed) was initially evaluated. Global model fit, including X^2 with p value, Comparative Fit Index (CFI), Goodness of Fit Index (GFI), root mean square error or approximation (RMSEA), and standardized root mean residual (SRMR; Kline, 2016), and the parameter estimations (magnitudes of standardized factor loading scores with p value, error variance, and squared multiple correlations of the measurements) were also evaluated. The configural model, targeting all age groups, showed that overall model fit was good, and magnitude of the factor loadings was high (X^2 = 71.53 [df = 16, p < .001], X^2/df = 4.45, CFI = 0.98, GFI = 0.98, RMSEA = 0.05 [90% CI = 0.04–0.07], and SRMR = 0.04). The model that was constrained with all factor loadings equal showed that ΔX^2 was not significantly different across the two groups (Chinese vs. Korean; p = .53). This result indicates measurement invariance status for whole model including both younger and older adults (Kline, 2016).

However the separate models (younger [young–middle aged] vs. older groups) showed different aspects: the younger group (18–64 years) had good model fit (X^2 = 41.27 [df = 16, p = .001], X^2/df = 2.58, CFI = 0.99, GFI = 0.98, RMSEA = 0.05 [90% CI = 0.03–0.06], and SRMR = 0.03), and the older age group (65 and older) also showed good model fit (X^2 = 41.70 [df = 16, p < .001], X^2/df = 2.61, CFI = 0.98, GFI = 0.97, RMSEA = 0.06 [90% CI = 0.04–0.08], and SRMR = 0.04). When constrained for all factor loadings, *only the model for older age group* indicated measurement noninvariance status between Chinese and Korean immigrants (ΔX^2 = 17.86, Δdf = 5, p = .003, CFI = .972, ΔCFI = .009). Given the findings of noninvariance on the fully constrained model for older adults, additional testing for the invariance of each factor loading proceeded separately. This procedure was focused on the extent to which items were similar or different across the two groups (Byrne, 2016).

2.7 | IRB approval

The Miami University (Oxford, OH) IRB reviewed this study and determined that this study is regulated human subjects research; however, the description meets the criteria of at least one exempt category included in 45 CFR 46 and associated guidance (the project reference number: 03220e).

3 | RESULTS

3.1 | Configural model and measurement invariance test for older adults

Although several prior studies found noninvariance status on measurement of K6 across different racial or ethnic groups, the current study findings indicated that the K6 measure is effective when targeting young–middle-aged Chinese and Korean immigrants. In other words, given items in the K6 measure are effective to evaluate the psychological distress of young–middle-aged Chinese and Korean immigrants. However, the model including only older immigrants showed different aspects from the model including only younger immigrants. Table 2 shows "Goodness of fit statistics for test of multiple-group invariance targeting older adults." Specifically, the second model (measurement comparison Model B in Table 2), which constrained "nervous" factor loading as equal, reported that ΔX^2 was not significantly different across the two groups. The "nervous" item was invariant across the two groups. The third model (Model C in Table 2),

TABLE 2 Goodness-of-fit statistics for test of multiple-group invariance (older adults)

	X^2	df	ΔX^2	Δdf	p Value	CFI	ΔCFI
1. Configural model (no equality constraints imposed)	41.70	16			<.001	.981	
2. Measurement model							
A. All factor loadings constrained equal			19.54	6	.003	0.981	0.010
B. Factor loading for only the item of nervous constrained equal	44.89	18	3.19	2	.203 (NS)	0.98	0.001
C. Factor loadings for nervous and *hopeless* constrained equal	**57.01**	**19**	**15.31**	**3**	**.002**	**0.972**	**0.009**
D. Factor loadings for nervous and *restless* constrained equal	**53.3**	**19**	**11.60**	**3**	**.009**	**0.975**	**0.006**
E. Factor loadings for nervous and *depress* constrained equal	**50.60**	**19**	**8.92**	**3**	**.030**	**0.977**	**0.004**
F. Factor loadings for nervous and worthless constrained equal	45.70	19	3.99	3	.262 (NS)	0.981	0.000

Abbreviations: CFI, Comparative Fit Index (CFI), ΔCFI, changes in Comparative Fit Index; Δdf, changes in degree of freedom; NS, nonsignificant; X^2, chi square; ΔX^2, changes in chi square.

Note: Models A–F indicate measurement comparisons with the configural model. Everything is an efforts was a fixed item. Significant p value (based on .05 level) is indicated in boldface.

which constrained both "nervous" and "hopeless" factor loadings as equal, found that ΔX^2 was significantly different across the two groups; these results indicate that the "hopeless" item could be problematic. The fourth model (comparison Model D in Table 2) constrained "nervous" and "restless" factor loadings as equal and showed significant difference in ΔX^2; these items were noninvariant across the two groups. The item "restless" could thus be problematic. The fifth model (comparison Model E in Table 2) constrained factor loadings of "nervous" and "depressed" as equal and found that ΔX^2 was significantly different, which means the "depressed" item could be problematic. Finally, the sixth model (comparison Model F in Table 2) constraining "nervous" and "worthless" factor loadings as equal did not show significantly different ΔX^2. In conclusion, the three items of "hopeless," "restless," and "depressed" were nonequivalent across the two groups (i.e., Chinese vs. Korean immigrants 65 years of age and older). Given factor loadings on Models C–E, the magnitude of factor loadings for the three items was different between older Chinese and Korean immigrants. Overall, the "hopeless" item was more reflected in the K6 of older Chinese adults, the "restless" item was more reflected in the K6 of older Korean adults, and the "depressed" item was far more reflected in the K6 of older Korean adults (the factor loadings were consistently 0.80 or higher on older Korean adults in Models C–E). Appendices (Figures A1–A4) present detail information of factor loadings of each item.

4 | DISCUSSION

This study addressed whether the K6 is measurement-equivalent for ethnicity by comparing two different ethnic groups (Chinese immigrants and Korean immigrants). Notably, only *older* Chinese and Korean immigrants showed measurement nonequivalent status on the K6. This result seems to be evidence for configural and measurement variance across older Asian immigrant subgroups in the United States. when the K6 is modeled to have a single-factor structure. In particular, the "hopeless," "restless," and "depressed" items varied between older Chinese and older Korean immigrants. Given findings indicated the fact that older adults could have more complicated experiences or more accumulated stress factors than younger adults.

A prior study (Jang et al., 2018) found that each Asian subgroup (18 years and older Chinese, Korean, and Vietnamese) showed different endorsement of K6 items based on language format (English format vs. non-English format): "Non-English survey users consistently showed a higher degree of endorsement compared with their English survey user counterparts" (Jang et al., 2018, p. 211); this finding highlighted the importance of understanding differences in cross-linguistic contexts. In the current study, 53% of young–middle-aged Chinese and Korean immigrants reported that they are good at speaking English (i.e., can do so well or very well), while only 17.9% of older adults perceived that they are good at speaking English. Although the CHIS allowed respondents to select Asian language options (e.g., Mandarin and Cantonese dialects, Korean, Vietnamese, and Tagalog; CHIS, 2014), there are still concerns about whether the questions administered in different

languages would address the same meanings equivalently for the two (Chinese and Korean) older groups (Cole, Kawachi, Maller, & Berkman, 2000). The two groups (older Chinese and Korean immigrants) could have different ways of thinking and delivering or articulating their emotional condition (hopeless, restless, and depressed). Moreover, the interpretation of given words and sentences could be influenced by different nuances and connotations stemming from linguistic differences (Kim et al., 2012; Kim et al., 2015).

Regarding the meaning of the depression item, there could be ethnocultural variations of depressive disorders or symptomatology in the measurement of depressive experiences (Kalibatseva & Leong, 2011). Scholars have argued that the existing measures for evaluating depressive symptoms have limited cultural validity, and this limitation reduces their clinical utility with foreign-born immigrants. The symptomatology of depression described by the K6 might not be culturally sensitive to depressive experience (i.e., might be endorsed differently) in different older immigrant subgroups (Kessler & Bromet, 2013; Marsella, Sartorius, & Jablensky, 1985). Notably, the current study found that "depressed" item was highly reflected in the K6 of older Korean immigrants as compared to their counterparts (older Chinese immigrants). This difference could be related to older Chinese adults' tendency to be more likely to present their psychological problems as physical complaints (Mak & Zane, 2004). In general, "somatization is defined as the presentation of one or more medically unexplained physical symptoms without reference to their possible psychodynamic origins" (Wu & Kelley, 2007, p. 14). Scholars (Parker, Cheah, & Roy, 2001; Parker, Gladstone, & Chee, 2001; Suen & Tusaie, 2004) have indicated that older Chinese adults prefer to express their emotional feelings somatically. Under Chinese culture, obedience, deference, self-control, and conformity are encouraged, while boldness, personal desire, and expressing emotions are considered as negative (Suen & Tusaie, 2004). Thus, older Chinese adults might be more likely to suppress their expression of depressive symptoms and, instead, use more somatic symptoms to express depressive symptoms.

A qualitative study, which focused on symptom manifestation among older Korean immigrants in the United States. (Lee-Tauler et al., 2016), found that expression of depressive symptoms included all different domains (i.e., complex domains) of individuals. Specifically, their depression reflected the combination of social discrimination, social isolation, and suicide in extreme circumstances. These immigrants also attributed depression to not achieving social and material success in America and strained relationships with their children. These interpretations seem to suggest more existential and complicated emotional reactions, rather than somatic reactions.

For the remaining two items, hopelessness and restlessness, more future research should be performed to discover reasons for the measurement invariance status between older Chinese and Korean immigrants. Possibly, different experiences from immigration or acculturation could directly or indirectly influence the measurement nonequivalence of the two items. For example, the two older immigrant groups (Chinese and Korean) could face different aspects of cultural conflicts or prejudice. Additionally, arriving in the United States at different time points could impact the measurement nonequivalence of

the two items. A prior study argued for the possibilities of longitudinal nonequivalence with the K6 measure (Drapeau et al., 2010). Depending on target groups' own experiences of different timing in immigrating to the United States, the meaning of "hopeless" and "restless" could be differently interpreted by older Chinese and Korean immigrants.

As the current study discussed at the beginning of the discussion section, the English proficiency effect should be more thoroughly considered. Concerning the potential effects of English proficiency among older adults, the current study performed additional data analyses. The K6 measurement invariance test (Chinese older adults vs. Korean older adults) was separately performed for the group having poor English proficiency and the other group having good English proficiency. Older groups having good English proficiency ($n = 80$) showed measurement equivalent status across the Chinese and Korean groups ($\Delta df = 6$, $\Delta X^2 = 5.02$, $p = .54$). However, older groups with poor English proficiency ($n = 368$) presented measurement variance problems ($\Delta df = 6$, $\Delta X^2 = 19.32$, $p = .004$). When the factor loadings were constrained manually, hopeless and restless were still showed nonequivalent status across Chinese and Korean older adults (comparison p values on nervous = .42, hopeless = .002, restless = .012, depressed = .07, and worthless = .26). Even though the sample size for the two groups (depending on English proficiency) did not show adequate distribution status (fluent older group $n = 80$ vs. nonfluent older group $n = 368$), this additional analysis would be beneficial in double-checking the potential effects of English proficiency on K6 measurement nonequivalence. As this additional analysis could just discuss the potential effects of English proficiency on K6 measurement nonequivalent status, it would be great if future research could address the limitations of the current study (e.g., inadequate distribution on sample size between the English fluent group and the nonfluent group).

In addition, the current study needs to think about whether the scoring strategy of K6 could be a potential matter of the nonequivalent status of several items or not. Yet, the K6 measure does not have an optimal cut point score of each item (i.e., Items 1–6); instead, there is an optimal cut point for the total score on K6. If future research could find same findings (targeting more broad range of Asian subgroups: not only for California residences but also for other State residences) with the current study findings, then scholars and professionals could try to suggest or modify the current scoring strategy to enhance cultural sensitivity/competency. However, more future studies should be performed; and professionals need to careful to drop off specific items among six items reflecting K6 structure.

Optimal cut point on the K6 is 6–18 versus 19+ (possible range 0–24; Kessler et al., 2002). The observed total score range from given metrics (of the current study) was 0–24. Among older Chinese and older Korean groups, only four individuals reported higher levels of the total score based on the optimal cut point (i.e., two individuals scored 20, one individual scored 21, and one individual scored 24). The current study performed the same analysis excluding these four individuals: analysis outcomes were exactly same with the things that the current study found originally, regardless of K6 total scores of individuals. When the factor loadings were constrained manually,

hopeless, restless, and depressed were still showed nonequivalent status across Chinese and Korean older adults (comparison of p values on nervous = .22, hopeless = .001, restless = .02, depressed = .04, and worthless = .28). This additional analysis would be beneficial to check whether individuals approaching the threshold scores on K6 significantly influence measurement invariance status or not.

5 | LIMITATIONS AND SUGGESTIONS ON FUTURE RESEARCH

The current study has several limitations. The generalizability of this study could be limited by the use of CHIS data collected only from California. Thus, future investigations should examine the equivalence of the K6 across diverse target groups using nationally representative data. Furthermore, other important factors such as gender, socioeconomic status, and acculturation that can possibly contribute to observed differences were not controlled. In particular, acculturation may be directly or indirectly related to the portion of one's lifetime spent in the United States as well as to language proficiency, and these factors might affect differential responses to items to measure psychological distress.

Measurement nonequivalence in the K6 between older Chinese and Korean immigrants may imply true population-based differences in addition to methodological artifacts. Given the findings from the current study, we may raise concern about the use of the K6 as a screening assessment for older Asian immigrant subgroups: as the K6 has been frequently used for diverse populations, researchers should pay attention to cross-cultural comparability when interpreting results from the K6, especially when the K6 measure is utilized to provide prevalence of psychological distress among racially, ethnically, or linguistically diverse older populations. Furthermore, clinicians or researchers need to be aware of the potential risk for misclassification when they try to screen for psychological distress in diverse older Asian subgroups. But again, dropping off specific items or modifying/adjusting total-score calculation strategies should be carefully performed. The decision-making toward ethnic relevant and culturally sensitive approaches should be based on replicated and valid empirical findings from future research.

REFERENCES
Andersen, L. S., Grimsrud, A., Myer, L., Williams, D. R., Stein, D. J., & Seedat, S. (2011). The psychometric properties of the K10 and K6 scales in screening for mood and anxiety disorders in the South African Stress and Health study. *International Journal of Methods in Psychiatric Research*, *20*(4), 215–223. https://doi.org/10.1002/mpr.351

Arnaud, B., Malet, L., Teissedre, F., Izaute, M., Moustafa, F., Geneste, J., ... Brousse, G. (2010). Validity study of Kessler's psychological distress scales conducted among patients admitted to French emergency department for alcohol consumption-related disorders. *Alcoholism: Clinical and Experimental Research*, 34(7), 1235–1245. https://doi.org/10.1111/j.1530-0277.2010.01201.x

Banh, M. K., Crane, P. K., Rhew, I., Gudmundsen, G., Stoep, A. V., Lyon, A., & McCauley, E. (2012). Measurement equivalence across racial/ethnic groups of the mood and feelings questionnaire for childhood depression. *Journal of Abnormal Child Psychology*, 40(3), 353–367. https://doi.org/10.1007/s10802-011-9569-4

Brooks, R. T., Beard, J., & Steel, Z. (2006). Factor structure and interpretation of the K10. *Psychological Assessment*, 18(1), 62–70. https://doi.org/10.1037/1040-3590.18.1.62

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Taylor & Francis.

California Health Interview Survey. (2014). *CHIS 2011–2012 methodology series: Report 2—Data collection methods*. Los Angeles, CA: UCLA Center for Health Policy Research.

Cole, S. R., Kawachi, I., Maller, S. J., & Berkman, L. F. (2000). Test of item-response bias in the CES-D scale. Experience from the New Haven EPESE study. *Journal of Clinical Epidemiology*, 53(3), 285–289.

Crockett, L. J., Randall, B. A., Shen, Y. L., Russell, S. T., & Driscoll, A. K. (2005). Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: A national study. *Journal of Consulting and Clinical Psychology*, 73(1), 47–58. https://doi.org/10.1037/0022-006x.73.1.47

Curran, J. P., West, G. S., & Finch, F. J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 14–29. https://doi.org/10.1037//1082-989X.1.1.16

Drapeau, A., Beaulieu-Prevost, D., Marchand, A., Boyer, R., Preville, M., & Kairouz, S. (2010). A life-course and time perspective on the construct validity of psychological distress in women and men. Measurement invariance of the K6 across gender. *BMC Medical Research Methodology*, 10, 68. https://doi.org/10.1186/1471-2288-10-68

Drapeau, A., Marchand, A., & Beaulieu-Prévost, D. (2012). Epidemiology of psychological distress. In L. Labate (Ed.). Mental illnesses: Understanding, prediction and control: InTech.

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of mental health and well-being. *Psychological Medicine*, 33(2), 357–362.

Gryn, T., & Gambino, C. (2012). *The foreign born from Asia: 2011*. Washington, DC: U.S. Department of Commerce: Economics and Statistics Administration. Retrieved from https://www.census.gov/library/publications/2012/acs/acsbr11-06.html

Jang, Y., Powers, D. A., Yoon, H., Rhee, M.-K., Park, N. S., & Chiriboga, D. A. (2018). Measurement equivalence of English versus native language versions of the Kessler 6 (K6) scale: An examination in three Asian American groups. *Asian American Journal of Psychology*, 9(3), 211–216. https://doi.org/10.1037/aap0000110

Kalibatseva, Z., & Leong, F. T. L. (2011). Depression among Asian Americans: Review and recommendations. *Depression Research and Treatment*, 2011, 9. https://doi.org/10.1155/2011/320902

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., ... Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959–976.

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, 34, 119–138. https://doi.org/10.1146/annurev-publhealth-031912-114409

Kessler, R. C., Green, J. G., Gruber, M. J., Sampson, N. A., Bromet, E., Cuitan, M., ... Zaslavsky, A. M. (2010). Screening for serious mental illness in the general population with the K6 screening scale: Results from the WHO World Mental Health (WMH) survey initiative. *International Journal of Methods in Psychiatric Research*, 19(1), 4–22. https://doi.org/10.1002/mpr.310

Kim, G. (2010). Measuring depression in a multicultural society: Conceptual issues and research recommendations. *Hallym International Journal of Aging*, 12, 27–46.

Kim, G., Bryant, A. N., & Parmelee, P. (2012). Racial/ethnic differences in serious psychological distress among older adults in California. *International Journal of Geriatric Psychiatry*, 27(10), 1070–1077. https://doi.org/10.1002/gps.2825

Kim, G., DeCoster, J., Bryant, A. N., & Ford, K. L. (2015). Measurement equivalence of the K6 scale: The effects of race/ethnicity and language. *Assessment*, 23(6), 758–768. https://doi.org/10.1177/1073191115599639

Kim, J., & Choi, N. G. (2010). Twelve-month prevalence of DSM-IV mental disorders among older Asian Americans: Comparison with younger groups. *Aging & Mental Health*, 14(1), 90–99. https://doi.org/10.1080/13607860903046461

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Lee-Tauler, S. Y., Lee-Kwan, S. H., Han, H., Lee, H. B., Gallo, J. J., & Joo, J. H. (2016). What does depression mean for Korean American elderly? A qualitative follow-up study. *Psychiatry Investigation*, 13(5), 558–565. https://doi.org/10.4306/pi.2016.13.5.558

Lomax, R. G., & Hahs-Vaughn, D. L. (2013). *An introduction to statistical concept* (3rd ed.). New York, NY: Taylor & Francis.

Mak, W. W. S., & Zane, N. W. S. (2004). The phenomenon of somatization among community Chinese Americans. *Social Psychiatry and Psychiatric Epidemiology*, 39(12), 967–974. https://doi.org/10.1007/S00127-004-0827-4

Marsella, A. J., Sartorius, N., & Jablensky, A. (1985). Cross-cultural studies of depressive disorders: An overview. In *Culture and depression studies in the anthropology and cross-cultural psychiatry of affect and disorder*. Berkeley and Los Angeles, California: University of California Press.

Mitchell, C. M., & Beals, J. (2011). The utility of the Kessler screening scale for psychological distress (K6) in two American Indian communities. *Psychological Assessment*, 23(3), 752–761. https://doi.org/10.1037/a0023288

Parker, G., Cheah, Y. C., & Roy, K. (2001). Do the Chinese somatize depression? A cross-cultural study. *Social Psychiatry and Psychiatric Epidemiology*, 36(6), 287–293.

Parker, G., Gladstone, G., & Chee, K. T. (2001). Depression in the planet's largest ethnic group: The Chinese. *The American Journal of Psychiatry*, 158(6), 857–864. https://doi.org/10.1176/appi.ajp.158.6.857

Slade, T., Grove, R., & Burgess, P. (2011). Kessler psychological distress scale: Normative data from the 2007 Australian National Survey of mental health and wellbeing. *Australian and New Zealand Journal of Psychiatry*, 45(4), 308–316. https://doi.org/10.3109/00048674.2010.543653

Stolk, Y., Kaplan, I., & Szwarc, J. (2014). Clinical use of the Kessler psychological distress scales with culturally diverse groups. *International Journal of Methods in Psychiatric Research*, 23(2), 161–183. https://doi.org/10.1002/mpr.1426

Suen, L.-J. W., & Tusaie, K. (2004). Is somatization a significant depressive symptom in older. *Geriatric Nursing*, 25(3), 157–163. https://doi.org/10.1016/J.GERINURSE.2004.04.005

Sunderland, M., Hobbs, M. J., Anderson, T. M., & Andrews, G. (2012). Psychological distress across the lifespan: Examining age-related item bias in the Kessler 6 Psychological Distress Scale. *International Psychogeriatrics*, 24(2), 231–242. https://doi.org/10.1017/S1041610211001852

The United Nations Educational, Scientific and Cultural Organization. (2009). *Investing in cultural diversity and intercultural dialogue: UNESCO world report*. Paris: UNESCO.

Trollor, J. N., Anderson, T. M., Sachdev, P. S., Brodaty, H., & Andrews, G. (2007). Age shall not weary them: Mental health in the middle-aged and the elderly. *Australian and New Zealand Journal of Psychiatry*, 41(7), 581–589. https://doi.org/10.1080/00048670701392817

US Department of Health and Human Services (2001). Publications and reports of the surgeon general mental health: Culture, race, and ethnicity. In *A supplement to mental health, A report of the surgeon general*. Rockville, MD: Substance Abuse and Mental Health Services Administration (US).

Wu, C., & Kelley, L. S. (2007). Choosing an appropriate depression assessment tool for Chinese older adults. *Journal of Gerontological Nursing*, *33*(8), 12–22.

APPENDIX A.
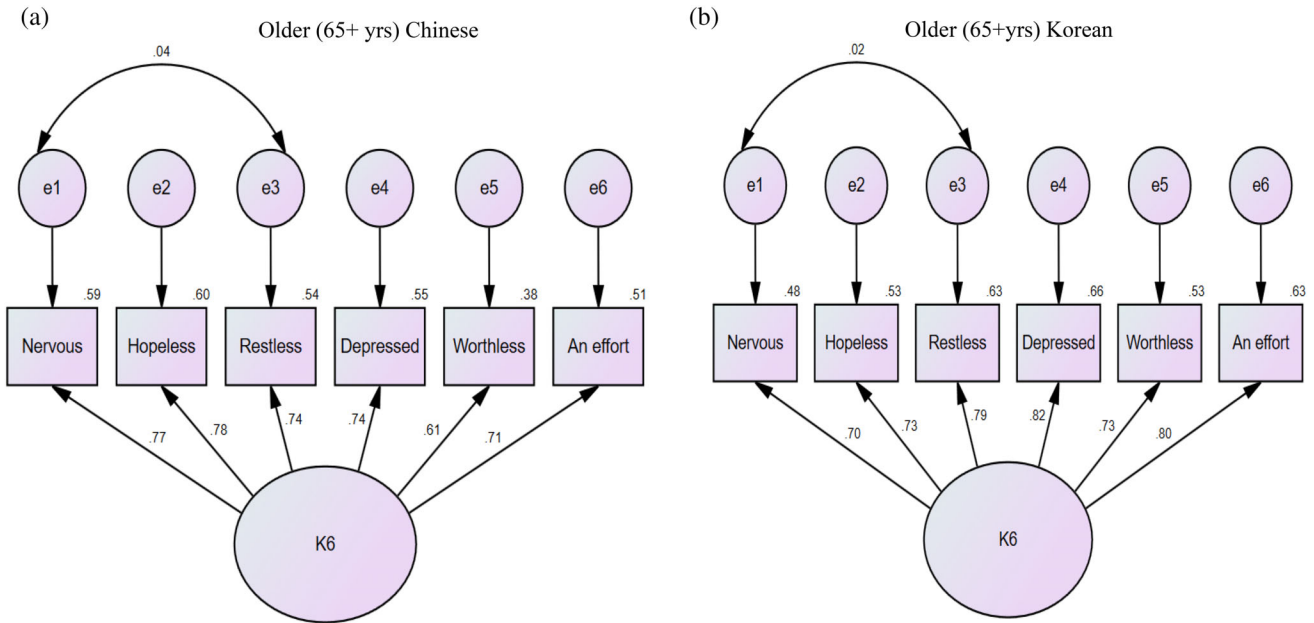
FIGURE A1 Constrained all factor loadings (older adults)
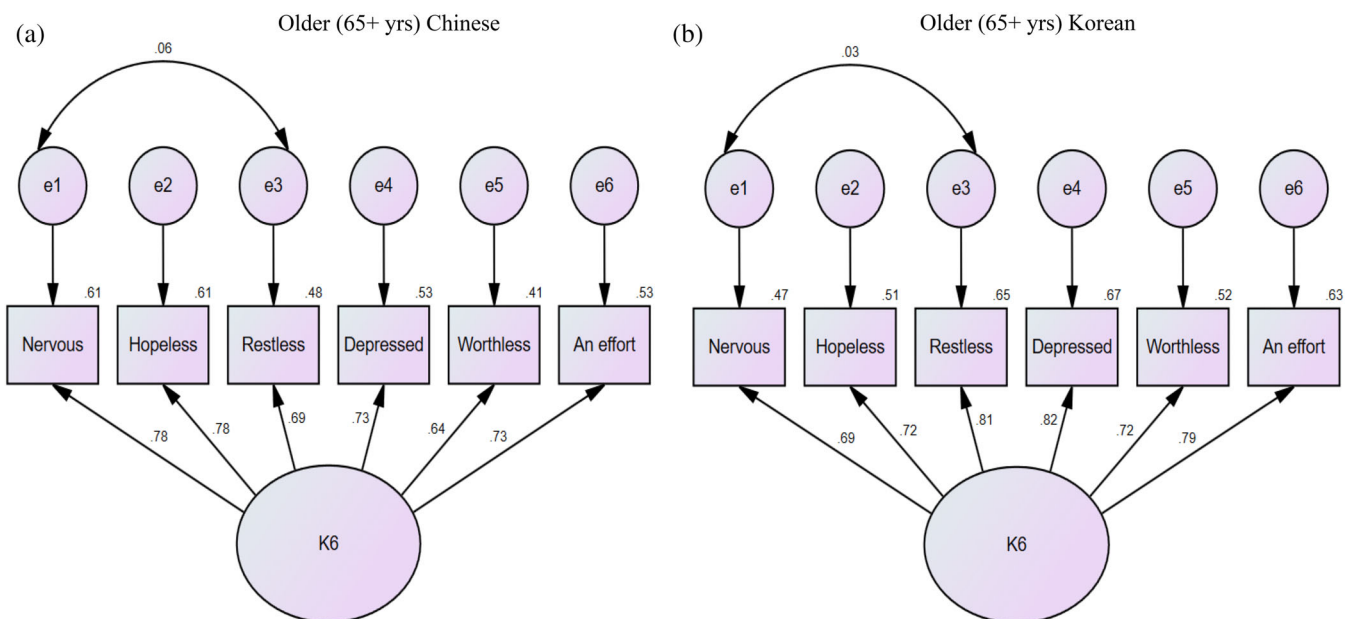
FIGURE A2 Model C (in Table 2). Note: Factor loadings for *nervous* and *hopeless* constrained equal
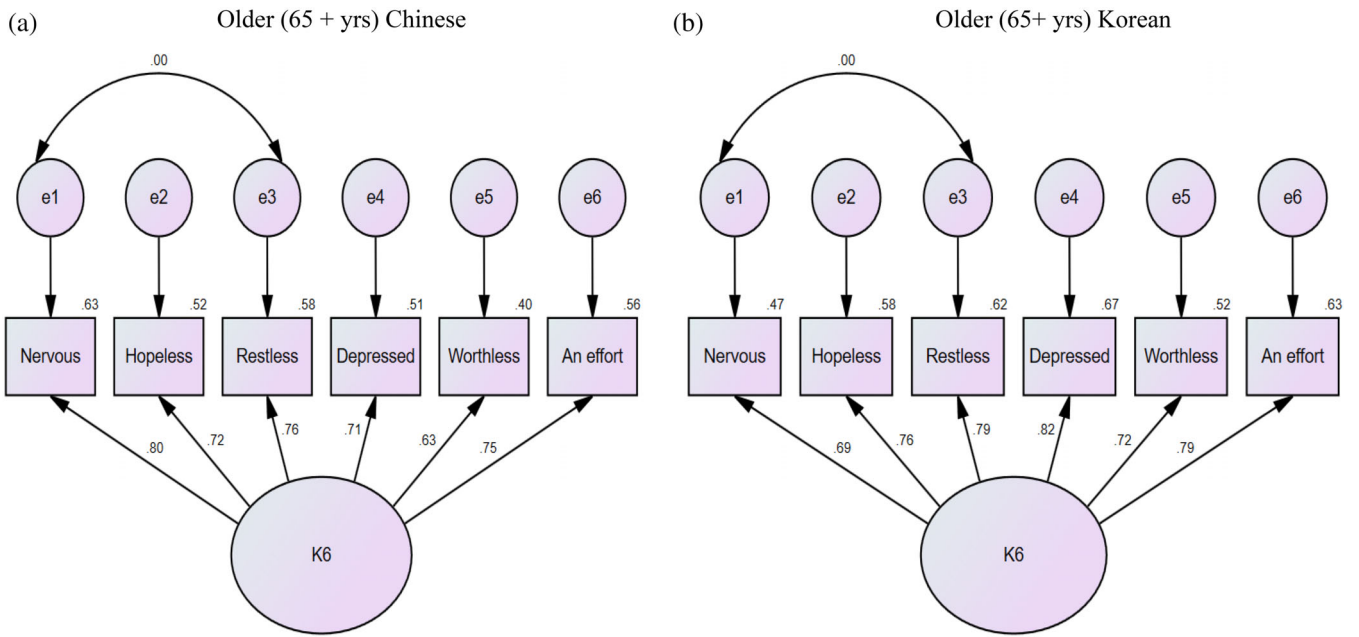
(a) Older (65 + yrs) Chinese

(b) Older (65+ yrs) Korean

FIGURE A3 Model D (in Table 2). Note: Factor loadings for *nervous* and *restless* constrained equal

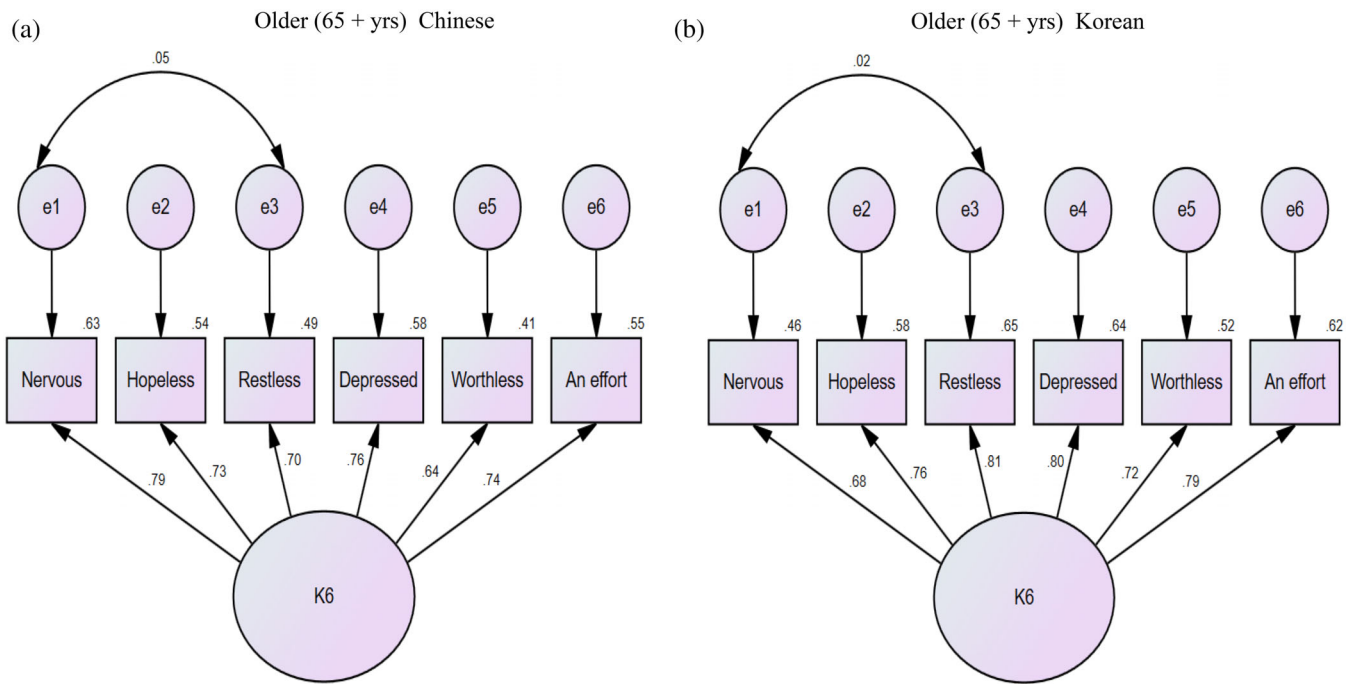(a) Older (65 + yrs) Chinese

(b) Older (65 + yrs) Korean

FIGURE A4 Model E (in Table 2). Note: Factor loadings for *nervous* and *depress* constrained equal