2018

Measurement Equivalence of English versus Native Language versions of the Kessler 6 (K6) Scale: An Examination in Three Asian American Groups

Yuri Jang

Daniel A. Powers
University of South Florida, dpowers@usf.edu

Hyunwoo Yoon

Min-Kyoung Rhee

Nan Sook Park
University of South Florida, nanpark@usf.edu

See next page for additional authors

HHS Public Access

Author manuscript

Asian Am J Psychol. Author manuscript; available in PMC 2019 September 01.

Published in final edited form as:

Asian Am J Psychol. 2018 September ; 9(3): 211–216. doi:10.1037/aap0000110.

Measurement Equivalence of English Versus Native Language Versions of the Kessler 6 (K6) Scale: An Examination in Three Asian American Groups

Yuri Jang, PhD[1], **Daniel A. Powers, PhD**[2], **Hyunwoo Yoon, PhD**[3], **Min-Kyoung Rhee, PhD**[1], **Nan Sook Park, PhD**[4], and **David A. Chiriboga, PhD**[5]

[1]Steve Hicks School of Social Work, Population Research Center, The University of Texas at Austin

[2]Department of Sociology, Population Research Center, The University of Texas at Austin

[3]School of Social Work, Texas State University

[4]School of Social Work, University of South Florida

[5]Department of Child and Family Studies, University of South Florida

Abstract

The use of languages other than English in population-based surveys is necessitated by the linguistic diversities in the United States. However, inclusion of multiple languages in survey data collection raises concerns about whether an instrument administered in different languages functions equivalently across groups. Using the Kessler Psychological Distress Scale 6 (K6), the present study examined differential item functioning (DIF) between surveys conducted either in English or the native language of the groups of Chinese Americans (n = 622), Korean Americans (n = 471), and Vietnamese Americans (n = 513). DIF analyses using a series of multiple indicator multiple cause (MIMIC) models showed that there were substantial differences between English and non-English versions in the endorsement of the K6 items, with patterns that differed by ethnicity. The K_4 (depressed) showed DIF in all three groups: non-English survey users consistently showed a higher degree of endorsement compared to their English using counterparts. It is speculated that its translated expression in Asian languages may carry less associations with illness/disorder than the English word, thereby making it easy to endorse among Asian language survey users. Findings suggest a lack of measurement equivalence between the K6 administered in English and Asian languages and call for caution in cross-linguistic contexts.

Keywords

Differential item functioning; Kessler 6; survey language; Asian Americans

Corresponding Author Yuri Jang, Ph.D., Steve Hicks School of Social Work, The University of Texas at Austin, 1925 San Jacinto Blvd., D 3500, Austin, TX 78712, Telephone number: 512-471-1702, Fax number: 512-471-9600.

Disclosure: The authors have reported no conflicts of interest.

Introduction

Individuals with language barriers are often excluded from population-based studies. With at least 350 languages being spoken in the United States, language serves as a major component of diversity (U.S. Census Bureau, 2015). It is striking that 47 million Americans do not speak English as their primary language, and 25.2 million Americans report that they speak English less than 'very well' (Pandya, McHugh, & Batalova, 2011). These facts suggest that there may be an under-representation of language minorities in published research. Under-representation may be particularly salient for Asian Americans, among whom linguistic isolation is common (Islam, Khan, Kwon, Jang, Ro, & Trinh-Shevrin, 2010; Jang, Park, Chiriboga, & Kim, 2017; Pew Research Center, 2013).

In order to capture the linguistic diversity of the population, some population-based studies have conducted surveys/interviews in languages other than English. For example, in the 2015 California Health Interview Survey (CHIS), 8.4% of adult participants were interviewed in languages other than English: Spanish, Chinese [Mandarin and Cantonese dialects], Vietnamese, Korean, and Tagalog (UCLA Center for Health Policy Research, 2016). The effort enabled individuals not fluent in English to participate in the study but raised concerns about measurement equivalence: whether an instrument administered in different languages addresses the same constructs and functions equivalently across the groups (Cole, Kawachi, Maller, & Berkman, 2000; Gallo, Anthony, & Muthén, 1994; Kim, Chiriboga, & Jang, 2009).

Differential item functioning (DIF) may occur when respondents systematically differ in their endorsement of the same item despite being similar with respect to the ability or attribute that item is intended to measure (Dorans & Holland, 1993; Zumbo, 1999). A sizable body of literature has addressed response biases associated with race/ethnicity for specific instruments (e.g., Iwata & Buka, 2002; Jang, Kwag, & Chiriboga, 2010; Kim et al., 2009; Teresi, Ramirez, Lai, & Silver, 2008); however, only a few studies focused on survey language as a source of DIF. For example, Hahn and colleagues (2014) confirmed the absence of DIF by survey language in their development of linguistically equivalent English and Spanish measures of social health. On the other hand, studies focusing on affective and cognitive measures report the presence of DIF between English and Spanish versions of an instrument (e.g., Azocar, Arean, Miranda, & Muñoz, 2001; Kim, DeCoster, Bryant, & Ford, 2016; Jones, 2006), elucidating potential reporting bias that may stem from linguistic differences in nuances and connotations.

Using the Kessler Psychological Distress Scale 6 (K6) as a target instrument, the present study examined survey language-associated measurement equivalence in Asian Americans. The K6 was developed as a screening tool for non-specific mental distress and serious mental illness (SMI) (Kessler et al., 2002, 2003). The scale measures the frequency of experiencing six different manifestations of psychological distress over the past 30 days: (1) nervous, (2) hopeless, (3) restless, (4) depressed, (5) worthless, and (6) everything was an effort. Due to its brevity, ease of administration, and ability to detect the possibility of diagnosable cases of SMI, the K6 has been widely used in national and international population-based studies (Kessler et al., 2003, 2010; Stolk, Kaplan, & Szwarc, 2014). It has

been translated into many Asian languages, and its psychometric properties have been validated in various samples of Asians and Asian Americans (e.g., Kang et al., 2015; Lee et al., 2012; Min & Lee, 2015).

One of the population-based studies to use the K6 has been the California Health Interview Survey (CHIS). The CHIS has employed the K6 since 2005 and contributed to a wealth of research on the K6 in diverse racial/ethnic groups, including Asian Americans (e.g., Nguyen & Goel, 2015; Sorkin, Nguyen, & Ngo-Metzger, 2011). However, these studies combined data from participants who were interviewed in English and in their native language, thus leaving open the possibility that the interview language could cause measurement bias. Substantiating this possibility, a recent study found that the measurement structure of the K6 in the CHIS participants who were interviewed in English was different from those interviewed in Spanish (Kim et al., 2016). Analyses showed that the structure of the non-Hispanic White group was different from that of Hispanics who were interviewed in either Spanish or English, but there was no difference between the two language versions for Hispanics. With regard to Asian Americans, no meaningful analyses were performed due to the disproportionally small numbers of the participants interviewed in Asian languages. It is notable that only 1.3% of the participants in the 2015 CHIS were interviewed in an Asian language (UCLA Center for Health Policy Research, 2016), further evidencing the persistent under-representation of non-English speaking Asian Americans in population-based studies.

Using an Asian American sample collected via methodological strategies designed to capture cultural and linguistic diversities, the goal of the present study was to assess the measurement equivalence of the K6 based on the selected survey language (English or one's native language). Based on previous studies on language-based equivalence in affective and cognitive measures (e.g., Azocar et al., 2001; Kim et al., 2016; Jones, 2006), we hypothesized that DIF would exist between English and Asian language versions of the K6.

Method

Sample

Data were drawn from the 2015 Asian American Quality of Life (AAQoL) survey. As part of a city-funded initiative, the target population was self-identified Asian Americans aged 18 and older living in Austin, Texas. In order to reach the broadest possible audience, the survey was conducted with culturally and linguistically sensitive approaches that included: (1) providing both English and Asian language versions of the survey questionnaire, (2) using research personnel (e.g., recruiters and survey assistants) who shared the languages and cultures of the target populations, and (3) building a strong partnership between the research team and key individuals and organizations within ethnic communities. More information on our culturally and linguistically sensitive recruitment strategy is available elsewhere (City of Austin, 2017).

The 10-page questionnaire for the AAQoL was originally developed in English and then translated into the languages of major Asian groups (e.g., Chinese, Vietnamese, Korean, Hindi, Gujarati, and Tagalog). In the case of Chinese, both traditional and simplified versions were prepared. The initial translations were conducted by 8 professional translators

and graduate-level bilingual researchers. For each language, the translated version was reviewed for accuracy by two or more bilingual volunteers. Upon refinement of the questionnaire, each language version was pilot tested with 3–5 community members who were representatives of the target group and spoke the target language.

The surveys were completed using a paper and pencil questionnaire in the participants' preferred languages. Recognizing that Asian Americans are often difficult to locate using standard recruitment strategies and that reliance on a single source can increase the chances for bias, multiple potential survey sites were contacted. In addition, the project was publicized through media and ethnic community sources, and referrals for individuals, groups, and organizations were actively sought. A total of 76 survey sessions took place at various locations and events across the City of Austin (e.g., churches, temples, grocery stores, small group meetings, and cultural events) from August to December, 2015. While the surveys were self-administered, bilingual research assistants at each survey site provided survey assistance. It took about 20 minutes to complete the 10-page questionnaire, and respondents were each paid US $10 for their participation. The AAQoL project was approved by the Institutional Review Board.

A total of 2,614 individuals participated in the AAQoL survey, about half of whom used non-English versions of the questionnaire. The present analyses focused on Chinese (n = 622), Korean (n = 471), and Vietnamese (n = 513) participants whose distribution of English survey users and native language users allowed meaningful comparisons.

Measures

The Kessler Psychological Distress Scale 6 (K6).

The K6 measures the frequency of experiencing 6 different symptoms of psychological distress over the past 30 days: (1) nervous, (2) hopeless, (3) restless, (4) depressed, (5) worthless, and (6) everything was an effort. Each item is rated on the 5-point scale ranging from 0 (none of the time) to 4 (all of the time). Responses were summed to create a composite score, ranging from 0 to 24. A score of 6 or greater is indicative of mental distress, and 13 or greater suggests SMI (Kessler et al., 2003). The K6 has been translated into Chinese, Korean, and Vietnamese, and its psychometric properties have been validated (e.g., Kang et al., 2015; Lee et al., 2012; Min & Lee, 2015). The internal consistency of the 6-item scale was high in the present samples (αs = .85 for Chinese surveyed in English, .87 for Chinese surveyed in Chinese, .87 for Koreans surveyed in English, .89 for Koreans surveyed in Korean, .91 for Vietnamese surveyed in English, and .87 for Vietnamese surveyed in Vietnamese).

Survey language.—English was coded as 0, and non-English (Chinese, Korean or Vietnamese) was coded as 1.

Covariates.—Covariates considered in the present analysis included age group (0 = 18–39, 1 = 40201359, 2 = 60 and older), sex (0 = male, 1 = female), marital status (0 = married, 1 = not married), education (0 = ≥ high school graduation, 1 = < high school graduation), nativity (0 = U.S.-born, 1 = foreign-born), and duration of residence in the U.S. (0 = ≥ 10

years, 1 = < 10 years). Dichotomy of the duration of residence was based on the immigration literature suggesting the tenth year as a marker of adaptation (Alegria et al., 2004; Beiser & Edwards, 1994).

Analytic Strategy

DIF analyses examine the extent to which group membership itself affects the probability of endorsing particular items on a scale. DIF approaches assess the probability that an item response for one group will differ from that of another group when a common variable is held constant. In the present study, Multiple Indicator Multiple Cause (MIMIC) models were used to explore DIF between survey language groups. MIMIC models can examine the direct effects of group membership on individual item responses with simultaneous factor analysis and regression of a latent trait on group differences while controlling for covariates (Muthén & Muthén, 2009). Because of their ability to control for the level of a latent trait, MIMIC models have been increasingly used as a method of detecting DIF (e.g., Gallo et al., 1994; Jang et al., 2010).

In the present analyses, a series of MIMIC models compared the direct effect of survey language (English vs. Non-English) on the endorsement pattern of the individual items of the K6 in each ethnic group. We estimated the direct effect as a contrast (or difference) in the level of endorsement of each of the K6 items between English and non-English survey users in each ethnic group, while accounting for the effects of other covariates.

Figure 1 illustrates the logic of DIF analysis. A measurement model relates the K6 items to a continuous latent variable (η_0) representing a latent trait of mental distress. A regression model relates survey language and covariates ($X_1 \ldots\ldots X_p$) to the latent trait of mental distress (η_0). The dashed line represents a direct effect that captures residual variations in item responses associated with a non-English survey use. DIF (item bias) is present if respondents surveyed in different languages but at the same level of a trait do not have the same level of endorsing a particular item.

For each of the K6 items, 6 sub-models were tested producing estimates $\gamma_{1,1} \ldots\ldots \gamma_{1,6}$ for each ethnic group. These parameters represent differences in the degree of endorsement of each of the K6 items between English and non-English survey users in each ethnic group and provide estimates of DIF through these direct effects. Positive estimates suggest a higher endorsement in non-English survey users compared to English survey users. Analyses were conducted using Stata 14 (StataCorp, 2015).

Results

Sample Characteristics

Sample characteristics are summarized in Table 1. The sample includes 622 Chinese, 471 Koreans, and 513 Vietnamese. Over 65% of the Chinese sample (n = 424), 79% of the Korean sample (n = 371) and 71% of the Vietnamese sample (n = 366) used non-English surveys. Across all ethnic groups, the rate of non-English survey use was consistently higher in older adults (aged over 60), those who were married, less educated, and foreign-born, and

those who had a shorter stay in the U.S. (< 10 years). Table 1 also presents descriptive characteristics of both individual and total scores of the K6.

MIMIC Analyses

Depicted as $\gamma_{0,1}\ldots\ldots\gamma_{0,p}$ in Figure 1, the DIF analysis carried out in each ethnic group yielded 144 estimates (6 × 3 × 8) of the effects of covariates ($X_1\ldots\ldots X_p$) on the latent trait (not presented in tabular format). The main findings of DIF are summarized in Table 2. The overall findings suggested that there were substantial differences in item endorsements of the individual items of the K6 between English and non-English survey users with patterns that differed by ethnicity.

Vietnamese were the only group presenting DIF in the K_1 (nervous); those who used the Vietnamese version showed a higher degree of endorsement than those who used the English version. Both Chinese and Vietnamese non-English survey respondents showed about equal levels of low endorsement of the K_2 (hopeless) compared to their English user counterparts. A lower endorsement in the K_3 (restless) among non-English survey users was observed in Koreans and Vietnamese. DIF of the K_4 (depressed) was observed in all three ethnic groups; non-English survey users of all three groups showed a higher degree of endorsement compared to their English user counterparts. Chinese non-English survey users showed a lower degree of endorsement of the K_5(everything was an effort) and higher degree of endorsement of the K_6(worthless), compared to their English survey user counterparts.

Discussion

Responding to the linguistic diversities in the United States and the need to include language minorities in population-based surveys (Islam et al., 2010; Jang et al., 2017; U.S. Census Bureau, 2015), there has been an increasing effort to use languages other than English in survey/interview administration (UCLA Center for Health Policy Research, 2016). The most common language being offered to non-English speaking individuals is Spanish; however, many studies are expanding their language options to include various Asian languages. While such effort enables many individuals with language barriers to be included, of concern is whether there is a lack of measurement equivalence between surveys using English and native language.

Although a few studies examined measurement equivalence of an instrument between English-using and Spanish-using Hispanics (e.g., Azocar et al., 2001; Hahn et al., 2014; Kim et al., 2016; Jones, 2006), Asian Americans have received little attention mostly due to the lack of data that include sufficient number of participants interviewed/surveyed in Asian languages. The Asian American Quality of Life (AAQoL) survey, which employed multiple strategies to capture cultural and linguistic diversities of Asian Americans (e.g., use of Asian language versions of the questionnaire, bilingual and bicultural research personnel, and community partnerships), included many non-English speaking participants, offering an optimal opportunity to explore measurement equivalence between English and non-English (Chinese, Korean, and Vietnamese) versions of the same instrument.

It is notable that well over half of Chinese American, Korean American, and Vietnamese American participants in the AAQoL survey responded in the native language version of the questionnaire. At a descriptive level, differences in demographic characteristics between English and non-English survey users were found in each ethnic group. Individuals who used their native language version of the survey questionnaire were more likely to be older, married, less educated, foreign-born, and with a shorter length of stay in the U.S. across all groups. These compositional differences are accounted for by including the potential confounders of non-English survey language as control variables in our models that assess DIF.

Moving beyond the descriptive level of differences, the intent of the present study was to explore differential item functioning (DIF) through use of multiple indicator multiple cause (MIMIC) models. These models allowed examination of response bias associated with survey language while controlling for the underlying effect of the latent trait and covariates (Gallo et al., 1994; Jang et al., 2010). Results demonstrated substantial differences in the endorsement of the individual items of the K6 between English and non-English survey users. Importantly, DIF was observed in all six items in at least one of the ethnic groups. Higher endorsement among non-English survey users was found in the K_1 (nervous) in Vietnamese and the K_6 (worthless) in Chinese. The opposite pattern of a higher endorsement among English survey users was observed in the K_2 (hopeless) in Chinese and Vietnamese, K_3 (restless) in Korean and Vietnamese, and K_5 (everything was an effort) in Chinese. The K_4 (depressed) was the only DIF item found across all three groups: those who responded in their native language were more likely to endorse to this item compared to those who responded in English.

The different patterns of endorsement between English survey users and native language survey users may result from the fundamental differences in understanding, processing, and expressing the symptoms of mental distress in different language groups. The most evidence for the hypothesis can be found in that fact that all three ethnic groups manifested language-based DIF in the K_4 (depressed). The reason for this systematic difference is unclear. It is speculated that its translated expression in Asian languages may carry less associations with illness/disorder than the English word, thereby making it easy to endorse among Asian language survey users. It may also be relevant that the word is passive in English but active in the three Asian language translations. Overall, the differences in meanings, connotations, and perceived intensity in all six items intended to measure mental distress require further exploration from the linguistic and cultural perspectives.

The generalizability of these findings is limited by at least two facts. First, due to the difficulty of implementing a random probability sampling strategy with a population that is generally difficult to identify, our recruitment strategy focused on identifying multiple sites and sources for soliciting a volunteer sample. Second, since the targeted populations were geographically restricted to Central Texas, regional variations could exist. Given the non-representative and regionally-defined nature of the sample, caution should be exercised in generalizing the findings to the larger population of Asian Americans. Future studies should not only include more representative samples but also further explore linguistic and cultural explanations for the response bias identified.

Despite these concerns, the results imply that researchers dealing with Asian populations to consider the potential for response bias of the survey instruments administered in different languages and call for caution in the cross-linguistic contexts. For those in practice, it is of course equally important to provide a language option when conducting clinical assessments since lack of familiarity with the English version may affect results. However, both researchers and practitioners should be aware that test scores may be differentially affected by the language version of the assessing tool. More studies are needed, not only to replicate the issue of scale equivalence in different languages within and across different ethnic groups, but also to enhance our understanding of how culture and language shape an individual's understanding and expression of mental health. Developing practice guidelines to promote a better understanding of instrument equivalence across languages would be helpful for practitioners working with diverse Asian populations.

Acknowledgments

Funding Sources: This work was supported in part by a grant from the National Institute on Aging (R01AG047106, PI: Yuri Jang, Ph.D.). The support for data collection was provided by the City of Austin's Asian American Quality of Life initiative (Contract No. 26–8275-39, PI: Yuri Jang, Ph.D.).

References

Alegria M, Takeuchi D, Canino G, Duan N, Shrout P, Meng XL, … Woo M (2004). Considering context, place and culture: The National Latino and Asian American Study. International Journal of Methods in Psychiatric Research, 13(4), 208–220. [PubMed: 15719529]

Azocar F, Arean P, Miranda J, & Muñoz RF (2001). Differential item functioning in a Spanish translation of the Beck Depression Inventory. Journal of Clinical Psychology, 57(3), 355–365. [PubMed: 11241365]

Beiser M & Edwards RG (1994). Mental health of immigrants and refugees. New Directions for Mental Health Services, 61, 73–86.

City of Austin. (2017). Asian American Quality of Life Retrieved from http://austintexas.gov/department/documents-3

Cole S, Kawachi I, Maller SJ, & Berkman L (2000). Test of item-response bias in the CES-D scale: Experience from the New Haven EPESE study. Journal of Clinical Epidemiology, 53, 285–289. [PubMed: 10760639]

Dorans NJ , & Holland PW (1993). DIF detection and description: Mantel-Haenszel and standardization. In Holland PW & Wainer H (Eds.), Differential item functioning (pp. 35–66). Hillsdale, NJ: Erlbaum.

Gallo JJ, Anthony JC, & Muthén BO (1994). Age differences in the symptoms of depression: A latent trait analysis. Journal of Gerontology: Psychological Sciences, 49, P251–P264.

Hahn EA, DeWalt DA, Bode RK, Garcia SF, DeVellis RF, Correia H, & Cella D (2014). New English and Spanish social health measures will facilitate evaluating health determinants. Health Psychology, 33(5), 490–499. [PubMed: 24447188]

Islam NS, Khan S, Kwon S, Jang D, Ro M, & Trin-Shevrin S (2010). Methodological issues in the collection, analysis, and reporting of granular data in Asian American populations: Historical challenges and potential solutions. Journal of Health Care for the Poor and Underserved, 21(4), 1354–1381. doi: 10.1353/hpu.2010.0939 [PubMed: 21099084]

Iwata N, & Buka S (2002). Race/ethnicity and depressive symptoms: A cross-cultural ethnic comparison among university students in East Asia, North and South America. Social Science and Medicine, 55, 2243–2252. [PubMed: 12409137]

Jang Y, Kwag K, & Chiriboga DA (2010). Not saying I am happy does not mean I am not: Cultural influences on responses to positive affect items in the CES-D. Journal of Gerontology: Psychological Sciences, 65B, 684–690.

Jang Y, Park NS, Chiriboga DA, & Kim MT (2017). Latent profiles of acculturation and their implications for health: A study with Asian Americans in central Texas. Asian American Journal of Psychology doi: 10.1037/aap0000080

Jones RN (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. [PubMed: 17060819]

Kang YK, Guo WJ, Xu H, Chen YH, Li XJ, Tan ZP, … & Luo Y (2015). The 6- item Kessler psychological distress scale to survey serious mental illness among Chinese undergraduates: Psychometric properties and prevalence estimate. Comprehensive Psychiatry, 63, 105–112. doi: 10.1016/j.comppsych.2015.08.011 [PubMed: 26555498]

Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, … & Zaslavsky AM (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychological Medicine, 32, 959–976.doi: 10.1017/S0033291702006074 [PubMed: 12214795]

Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, … & Zaslavsky AM (2003). Screening for serious mental illness in the general population. Archives of General Psychiatry, 60, 184–189. doi:10.1001/archpsyc.60.2.184 [PubMed: 12578436]

Kessler RC, Green JG, Gruber MJ, Sampson NA, Bromet E, Cuitan M, … & Lara C (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, 4–22. [PubMed: 20527002]

Kim G, Chiriboga DA, & Jang Y (2009). Cultural equivalence in depressive symptoms in older White, Black, and Mexican-American adults. Journal of the American Geriatrics Society, 57, 790–796. [PubMed: 19484834]

Kim G, DeCoster J, Bryant AN, & Ford KL (2016). Measurement Equivalence of the K6 Scale: The Effects of Race/Ethnicity and Language. Assessment, 23, 758–768. [PubMed: 26282779]

Lee S, Tsang A, Ng KL, Ma YL, Guo W, Mak A, & Kwok K (2012). Performance of the 6-item Kessler scale for measuring serious mental illness in Hong Kong. Comprehensive Psychiatry, 53, 584–592. [PubMed: 22104556]

Min JW & Lee SH (2015). Validation of the K6/K10 scales of psychological distress and their optimal cutoff scores for older Koreans. The International Journal of Aging and Human Development, 80, 264–282. doi: 10.1177/0091415015590316 [PubMed: 26195502]

Muthén LK, & Muthén BO (2009). Mplus us er's guide (5th ed). Los Angeles: Muthén & Muthén.

Nguyen D, & Goel M (2015). Social determinants and the psychological distress of Vietnamese immigrants. International Journal of Culture and Mental Health, 8, 22–33.

Pandya C, McHugh M, & Batalova J (2011). Limited English proficient individuals in the United States: Number, share, growth, and linguistic diversity Washington, DC: Migration Policy Institute.

Pew Research Center (2013). The rise of Asian Americans Washington, DC: Pew Research Center Retrieved from http://www.pewsocialtrends.org/files/2013/04/Asian-Americans-new-full-report-04-2013.pdf

Sorkin DH, Nguyen H, & Ngo-Metzger Q (2011). Assessing the mental health needs and barriers to care among a diverse sample of Asian American older adults. Journal of General Internal Medicine, 26, 595–602. [PubMed: 21321793]

StataCorp. (2015). Stata Statistical Software: Release 14 College Station, TX: StataCorp LP.

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, 161–183. [PubMed: 24733815]

Teresi JA, Ramirez M, Lai JS, & Silver S (2008). Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. Psychology Science Quarterly, 50(4), 538. [PubMed: 20165561]

Census Bureau U.S. (2015). Census Bureau reports at least 350 languages spoken in U.S. homes Retrieved from https://www.census.gov/newsroom/press-releases/2015/cb15-185.html

ULCA Center for Health Policy Research (2016). CHIS 2015 Methodology Report 2 Retrieved from http://healthpolicy.ucla.edu/chis/design/Documents/chis-2015-short-methodology-report-2_data-collection_2016-12-13.pdf

Zumbo BD (1999). A handbook on the theory and methods of Differential Item Functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores Ottawa, Ontario, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Public Significance Statement

In recognition of the increasing use of languages other than English in population-based surveys, the present study explored measurement equivalence of the K6 by survey language in three Asian American groups. Differential item functioning (DIF) analyses showed substantial differences in the endorsement of the K6 items between English and native language versions, with patterns that differed by ethnicity. Findings suggest a lack of measurement equivalence between the K6 administered in English and Asian languages and call for caution in cross-linguistic contexts.

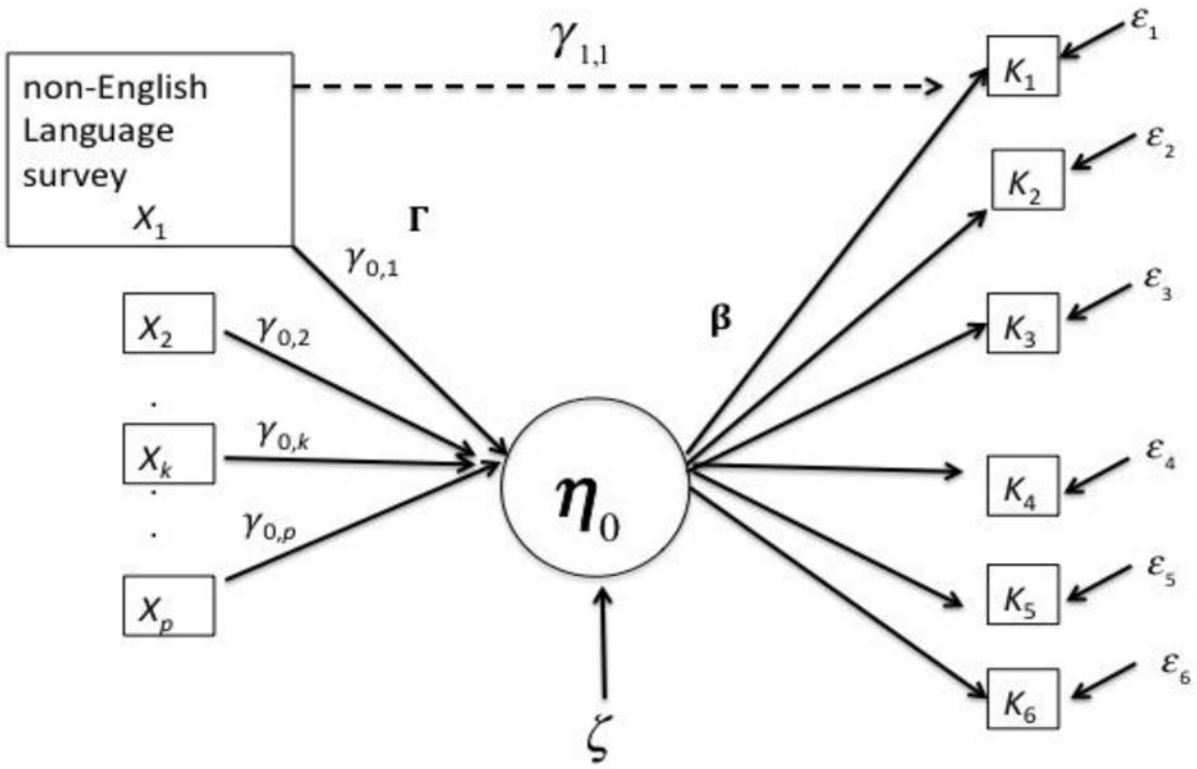

Figure 1.

Path diagram for the MIMIC model. The factor loadings of the measurement model relating the latent trait (η_0) to the K6 items ($K_1,\ldots,K_6$) are contained in β. Coefficients for the regression of survey language and covariates ($X_1,\ldots,X_p$) on latent trait are contained in Γ. The direct effects of survey language for each of the K6 items ($K_1,\ldots,K_6$) are contained in $\gamma_{1,1} \ldots \gamma_{6,1}$, with the dashed line depicting the difect effect of survey language on K_1 for purposes of illustration.

Table 1

Descriptive Characteristic of the Sample

	Chinese (n = 622)		Korean (n = 471)		Vietnamese (n = 513)	
	English Survey User (n = 198)	Chinese Survey User (n = 424)	English Survey User (n = 100)	Korean Survey User (n = 371)	English Survey User (n = 147)	Vietnamese Survey User (n = 366)
Age group						
18–39	68.7	38.4^{c}	67.0	31.4^{c}	68.0	27.3^{c}
40–59	25.8	25.5	28.0	43.5^{b}	25.2	44.1^{c}
60+	5.6	32.1^{c}	5.0	25.1^{c}	6.8	28.7^{c}
Female	52.0	58.9	64.0	59.6	56.5	58.0
Unmarried	52.3	29.5^{c}	54.0	18.1^{b}	59.2	34.6^{c}
< High school graduation	7.6	16.7^{b}	10.1	23.0^{c}	12.3	46.1^{c}
Foreign-born	71.7	98.8^{c}	73.0	98.6^{c}	66.7	98.1^{c}
Duration of residence in the U.S. (< 10 years)	23.7	49.2^{c}	19.2	42.1^{c}	13.0	31.3^{c}
K6 items						
K_1 (nervous)	1.48±0.80	1.29±0.96^{a}	1.49±0.87	1.32±0.91	1.36±1.01	1.44±0.91
K_2 (hopeless)	0.76±0.88	0.45±0.80^{c}	0.84±0.94	0.77±0.85	1.04±0.99	0.77±0.92^{b}
K_3 (restless)	1.15±0.92	0.93±0.90^{b}	1.21±0.99	0.90±0.92^{b}	1.32±1.12	0.94±0.91^{c}
K_4 (depressed)	0.61±0.80	0.66±0.80	0.70±0.88	0.82±0.88	0.82±0.92	1.09±0.96^{b}
K_5 (everything was an effort)	1.36±1.13	0.79±0.84^{c}	1.31±1.02	1.10±0.96	1.40±1.17	1.43±1.05
K_6 (worthless)	0.51±0.84	0.59±0.81	0.58±0.85	0.48±0.75	0.75±0.99	0.65±0.91
K6 total score	5.87±4.13	4.70±4.06^{b}	6.14±4.36	5.04±4.25	6.65±5.14	6.32±4.45

Note. Comparative analyses were conducted using t or χ^2 test between English survey users and non-English survey users in each ethnic group.

$^{a}p < .05$.

$^{b}p < .01$.

[c] $p < .001$.

Table 2

DIF Estimates of Survey Language from MIMIC Models Separately Estimated in Each Ethnic Group

	Chinese	Korean	Vietnamese
K_1 (nervous)	−0.007	−0.090	0.196[b]
K_2 (hopeless)	−0.184[c]	0.073	−0.234[c]
K_3 (restless)	−0.003	−0.242[b]	−0.380[c]
K_4 (depressed)	0.276[c]	0.277[c]	0.488[c]
K_5 (everything was an effort)	−0.426[c]	−0.125	0.116
K_6 (worthless)	0.275[c]	−0.023	−0.012

Note. The series of the DIF analyses yielded 144 estimates (6 × 3 × 8) of the effects of covariates ($X_1,\ldots,X_p$). Positive estimates suggest a higher endorsement in non-English survey users compared to English survey users.

[a] $p < .05$.

[b] $p < .01$.

[c] $p < .001$.