



Published in final edited form as:

Asian Am J Psychol. 2012 December ; 3(4): 230–253. doi:10.1037/a0025628.

Culturally Sensitive Depression Assessment for Chinese American Immigrants: Development of a Comprehensive Measure and a Screening Scale Using an Item Response Approach

Rose Wong,

School of Social Welfare, University of California, Berkeley

Rufina Wu,

School of Social Welfare, University of California, Berkeley

Carmen Guo,

School of Social Welfare, University of California, Berkeley

Julia K. Lam, and

Department of Molecular and Cell Biology, University of California, Berkeley

Lonnie R. Snowden

School of Public Health, University of California, Berkeley

Abstract

The present mixed methods study developed a comprehensive measure and a screening scale of depression for Chinese American immigrants by combining an emic approach with item response analysis. Clinical participants were immigrants diagnosed by licensed clinicians who worked in the community. Qualitative interviews with clinicians and clinical participants (N = 63) supported the definition of the construct of depression—which guided scale development—and a 47-item pilot scale. Clinical and community participants (N = 227) completed the pilot scale and measures of neurasthenia and acculturative stress, and the Patient Health Questionnaire Depression Module (PHQ-9). A Rasch Partial Credit Model of 42-items—representing psychological, somatic and interpersonal domains of distress—best fit the data. Twenty-three items overlapped with the DSM-IV symptoms of major depression. Twenty-seven items were biased by acculturation-related variables. Nine items appropriate for self-report screening in primary care and community organizations were chosen to form a brief scale. Both measures showed strong reliability and concurrent and convergent validity. The 9-item scale had better content validity than the PHQ-9. Implications regarding the impact of culture for assessment are highlighted.

Keywords

depression; assessment; acculturation; immigrants; Chinese American

Correspondence concerning this article should be addressed to Rose Wong, Department of Psychiatry, University of California, San Francisco, 401 Parnassus Avenue, Box 0894-OVS, San Francisco, CA 94143-0894. rosewong@berkeley.edu. Rose Wong is now at the Department of Psychiatry, University of California, San Francisco.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/aap

Chinese Americans underutilize mental health services more than any other ethnic group in the U.S. (Abe-Kim et al., 2007). Immigrant Chinese Americans especially have little knowledge of mental illness, delay treatment until their condition is very severe, and seek help in primary care and the community rather than consulting mental health specialists (Hong, Lee, & Lorenzo, 1995; Kung, 2004). With regard to depression, a problem of poor recognition by health care providers (Chung et al., 2003) and low detection with translated self-report scales was observed (Huang, Chung, Kroenke, Delucchi, & Spitzer, 2006; Yeung et al., 2008). The current study combined an emic approach with item response analysis to develop a comprehensive measure of the construct of depression for research and education—the 42-item Chinese American Depression Scale (CADS-42)—and a 9-item scale (CADS-9) for screening in primary care and community organizations.

Assessing Depression in Chinese Americans via Standard Measures

The authors of several widely used depression assessment instruments reported that their scales had adequate psychometric properties when assessing Chinese Americans. These include the PHQ-9, Beck Depression Inventory (BDI) and Center for Epidemiological Studies Depression Scale (CES-D).

Patient Health Questionnaire Depression Module

The PHQ-9 is a DSM-IV criterion-based instrument for screening in primary care (Kroenke, Spitzer, & Williams, 2001). The PHQ-9 was evaluated in two studies of Chinese American immigrants (T. M. Chen, Huang, Chang, & Chung, 2006; Huang et al., 2006; Yeung et al., 2008). The first reported strong reliability ($\alpha = .91$) and validity—including sensitivity of .81 and specificity of .98 (Yeung et al., 2008). The performance estimates, however, were likely biased due to exclusion of participants who obtained low scores (< 15); these participants were not administered the semi-structured clinical interview. Correcting for this bias by extrapolating the rate of depression detected with the clinical interview to those who were excluded yields a revised sensitivity of .39.

The second study found satisfactory reliability ($\alpha = .79$), but identified cultural differences between Chinese American immigrants and other U.S. ethnic groups (Huang et al., 2006). First, three items—sleep, appetite and psychomotor disturbances—showed differences in item functioning, suggesting that the symptoms had a different meaning or were poorly understood. Second, the PHQ-9 detected a suspiciously low rate of depression for Chinese, especially men.

Research conducted in Asia found similar indications of poor validity. Studies in Taiwan (Liu et al., 2011), Hong Kong (Lai, Tang, Lee, Yip, & Chung, 2010) and China (S. Chen et al., 2009) reported sensitivity from .85 to .92. These indicators, however, were also likely biased due to the exclusion of a part of the sample. The studies also found lower optimal cutoff scores for the PHQ-9 (S. Chen et al., 2009; Lai et al., 2010) and PHQ-2 (Liu et al., 2011), including a cutoff of only four points for men on the PHQ-9 (Lai?). These lower cutoff scores imply that Chinese endorsed fewer items overall or interpreted items unlike other respondents.

Beck Depression Inventory

The BDI is a 21-item measure of the severity of depression (Beck, 1961). The Chinese version of the BDI (CBDI) was determined to be impractical for self-administration due to its length and complexity (Yeung, Howarth, et al., 2002). The instrument's sensitivity of .79, and its other reported favorable psychometric properties, were also likely biased upward due to the exclusion of many subjects who obtained low scores (< 16); they were not given the

semi-structured clinical interview and not included in validity tests. Extrapolating the rate of depression detected with the clinical interview yielded a sensitivity of .62.

A sensitivity of .78 was also reported for a subset of the sample based on an empirically determined cutoff score (≥ 13 points) (Yeung, Neault, et al., 2002). Again, this reported sensitivity was likely biased upward. Cases of depression detected by the clinical interview but “missed” by the CBDI, in addition, were clustered in a very mild range. These findings suggest a manifestation of depression in the mild range that is not captured by the instrument.

Studies conducted in Asia also showed the CBDI’s poor content validity. A study in China found seven items that correlated poorly with total scores (Y. Zheng, Wei, Goa, Zhang, & Wong, 1988). A study in Hong Kong detected five items that did not function equally well in the CBDI vs. the BDI when administered to bilingual college students (D. W. Chan, 1991). Factor analyses in China, furthermore, found un-interpretable factors, whereas the Hong Kong study—of college students with exposure to Western culture—extracted factors similar to those found among European Americans.

Center for Epidemiologic Studies Depression Scale

The 20-item CES-D measures depressive symptoms and is designed for use in the general population (Radloff, 1977). Although good reliability among Chinese American immigrants—.77 (Y.-W. Ying, 1988) and .92 (Rankin, Galbraith, & Johnson, 1993)—was reported, poor validity was observed. First, factor analyses suggested low construct validity (Gupta & Yick, 2001; Kuo, 1984; Y.-W. Ying, 1988). Depressed affect and somatic items, and depressed affect and interpersonal items, loaded together to form mixed, un-interpretable factors. Lower validity was also associated with lower acculturation; factors derived for bilingual college students were similar to those found for U.S. samples (Y. Ying, Lee, Tsai, Yeh, & Huang, 2000). A second indication of low validity is that positive affect and positive self-concept items had little cultural relevance (Cheung & Bagley, 1998; Kuo, 1984; Y.-W. Ying, 1989). Research in Asia also found evidence for mixed factors rather than clearly differentiated factors and for invalid items (Boey, 1999; Cheung & Bagley, 1998; Yen, Robins, & Lin, 2000).

Cultural Expressions of Depression and Scales Adapted to Accommodate Cultural Differences

In China, the majority of key terms drawn from U.S. depression assessment scales were found not to have fully-equivalent Chinese terms (Y. Zheng, Xu, & Shen, 1986). Study participants, as they interpreted key symptoms, demonstrated a cultural style of expression. They emphasized behavior- or mind-focused descriptions of functioning, philosophical and descriptive statements, and somatic and interpersonal expressions: “Uncomfortable inside heart” expressed “depressed”, and “intolerable inside heart” expressed “agitated”. Interpersonal expressions—referring to being judged, social comparisons and cutoff relationships—expressed psychological terms: “Feel less capable than others” expressed “failure”, and “no one understands or cares about me” and “friendless” expressed “hopelessness”. Key terms such as “weight loss” and “suicidal interest”, also, largely lacked equivalence.

The psychometric properties of measures were improved by removing invalid items and adding culture-specific ones. The newly introduced content included symptoms of somatic and interpersonal distress (T.-A. Cheng & Williams, 1986; Lin, 1989) as well as psycholinguistically equivalent expressions (B. Chan, Parker, Tully, & Eisenbruch, 2007; Y. Zheng & Lin, 1991). The lone culturally adapted measure tested in the U.S., the 48-item

Chinese Depression Inventory (Y. Zheng & Lin, 1991), performed similarly to the CBDI, but both scales were unsuited for self-report due to patients' unwillingness to complete them (Yeung, Neault, et al., 2002). The authors noted, furthermore, that research assistants may have introduced bias by assisting respondents to understand the items.

Phenomenological Research for Describing Depression in Chinese Americans

Immigrants' explanatory models of illness—intertwining psychological, somatic and interpersonal causes and symptoms—point to a construct with multiple domains (Yeung, Chang, Gresham, Nierenberg, & Fava, 2004; Y.-W. Ying, 1990). Female community-based participants gave explanations that contained interpersonal triggers and consequences, whether they were more psychological or more somatic (Y.-W. Ying, 1990). Primary care patients, similarly, explained depression in terms of physical complaints, but attributed its cause to stress and psychological problems related to interpersonal difficulties (Yeung et al., 2004). Patients with lower acculturation, furthermore, presented stronger levels of somatic symptoms and did not acknowledge depressed mood spontaneously.

Research on somatization clarified the role of somatic complaints in a mild stage of depression and the symbolic function of somatic expressions. Chinese patients' somatic emphasis and psychological de-emphasis are associated with milder depression (H. Chen, Guarnaccia, & Chung, 2003) and lower acculturation and education (W. Mak & Zane, 2004; Parker, Chan, Tully, & Eisenbruch, 2005). Research on the metaphorical qualities of Chinese language supported this interpretation, showing that commonly used expressions with body words communicate psychological and interpersonal difficulties (Tung, 1994).

Epidemiological research also pointed to an influence of acculturation on symptom expression and the relevance of neurasthenic, somatic and anxiety-based expressions. The national survey of Asian Americans found complex patterns of risk based on immigration-related factors and gender that suggested differences in the experience or report of depression. A large-scale community survey of Chinese American immigrants found that depression overlapped significantly (23.1%) with neurasthenia (Y.-P. Zheng et al., 1997)—a DSM-IV culture-bound syndrome considered a predecessor of depression in Chinese societies (Parker et al., 2005). Immigrants with depression, in addition, scored very high on somatization and as high on measures of anxiety, interpersonal sensitivity, obsessive-compulsiveness and paranoid ideation as immigrants with anxiety disorders.

Self-construal research highlighted the centrality of socially based distress. Asian Americans, especially immigrants, were shown to have high levels of interdependent self-construal and low levels of independent self-construal, which changed slowly with acculturation (Kwan, Bond, & Singelis, 1997; Singelis, Bond, Sharkey, & Lai, 1999). Interdependent self-construal was associated with a cognitive focus on social relations (Diener & Diener, 1995) Kwan et al., 1999), social anxiety and distress (Okazaki, 1997), with less importance given to individually oriented self-esteem concepts (Singelis et al., 1999).

Independent self-construal was recognized as a premise in the CES-D and BDI, based on the negative relationship of independent self-construal and depression scores (Norasakkunkit & Kalick, 2002; Okazaki, 2000). Interdependent self-construal, also, was indirectly related to depression via a social personality variable and anxiety. Specifically, it predisposed a person for sociotropy—a trait of high relational dependence, need to please and social threat—which was related to depression via heightened anxiety (W. W. S. Mak, Law, & Teng, 2011).

Purpose of the Present Study

Consistent with the view that mental illness manifests itself uniquely in each culture (Marsella, Kinzie, & Gordon, 1973), the current study sought to gain knowledge of the cultural manifestation of depression in Chinese American immigrants. Ascertaining culture-based symptoms and differences among immigrants also was a precursor to evaluating the goodness-of-fit and efficacy of Western measures and adapting assessment for cultural sensitivity. The comprehensive measure would support research on depression and its correlates from an emic perspective. It would also serve for developing short, unbiased scales and for educating clinicians and the community. The screening scale, as a first 'indigenous' measure, would be a tool in interventions that promote linkage to care. With these considerations, the study aimed to quantify the impact of culture on assessment of less acculturated Chinese in the U.S.

General Method

Overview

The research is described in three parts. Study 1 describes the development of a construct map and pilot items using qualitative methods. Study 2 describes the development of CADS-42 using item response analysis. Study 3 describes the development and initial validation of CADS-9, constructed of items from CADS-42. Study 1 employed a qualitative sample ($N = 60$) of clinical participants and clinicians. Studies 2 and 3 employed an empirical sample ($N = 227$) of clinical and community participants, non-overlapping with the qualitative sample.

The research was conducted in the region of San Francisco. Participants had oral fluency in Chinese, study materials were in Chinese and English, and procedures were conducted in Chinese unless otherwise noted. Research assistants were university students with oral and written fluency in Chinese and English. The institutional review board of the University of California, Berkeley approved all study procedures.

Emic Approach: Insuring Cultural Awareness

The study's emic approach considered the manifestation of depression in natural settings in three ways. First, the samples were diverse and the criterion was diagnosis of major depression (MD) or dysthymia by clinicians experienced with treating Chinese American immigrants in the community. Second, expert-based qualitative inquiry was the basis of defining the theoretical construct and constructing a large pool of pilot items in Study 1. A group of eight experts, clinicians who participated in Study 1, provided qualitative analyses and consultation for all studies. Third, concurrent validity was assessed by relating the scales to the criterion—clinical diagnosis by culturally aware clinicians.

Emic Approach: Item Response Evaluation

The study applied Wilson's (2005) item response approach, a bottom-up method of scale development based on the use of item response analysis. Following its four logical steps ensured the development of measures from an emic perspective. These first three steps concern Study 1. The first step was to define the construct based on the population's conceptualization of illness. The definition, presented visually in a construct map, consisted of a description of the persons who experience depression and the content that represented their experience along a continuum of severity. The definition also considered the scale's purpose—to detect clinically relevant risk in the community.

The second step was to design items according to the description of content. The third step was to define the ‘outcome space’, a set of responses to items that was meaningful to respondents and provided adequate information to separate respondents by their levels of depression severity. The fourth step was to use Rasch analysis to fit a model to the data that represented the defined construct (Study 2), and then fit another model to the data that met the purpose of screening (Study 3).

Rasch (1960) analysis was suited to the study’s objectives because of its utility for understanding the role of items in levels of severity, detecting differences between socio-demographically and culturally-defined groups, and in developing unbiased and efficient scales. Rasch analysis estimates the locations of items on a continuum of severity and it detects, with the analysis of differential item functioning (DIF), differences in the endorsability of items between subgroups after controlling for their levels of depression. DIF captures differences in experience, willingness to report, interpretation, or other factors concerning the relevance of the symptom to each socio-demographic or cultural subgroup. With this information, it is possible to determine whether, for example, an item assessing suicidal ideation represents the same level of depression severity for immigrants with better and poorer English proficiency. It also becomes possible to choose unbiased items—selecting the fewest conceptually non-repetitive items that cover the clinically relevant range—to form an efficient assessment instrument. The choice of items can also target a specific purpose—such as to screen for somatic presentations of depression.

For other reasons too, Rasch analysis was suited to studying the population of concern—a minority group that is difficult to reach due to the stigma of mental illness. With an assumption of the local independence of responses to items, a reliable model with objective measures—which are not dependent on the instrument or the participants of the study—can be obtained with a small, unrepresentative sample (Bond & Fox, 2001). Rasch modeling, in addition, assumes unidimensionality for the practical purpose of creating a useful measurement model (Smith, 1996). All items that form the instrument support the measurement of the latent trait under study.

Study 1 – Construct Map and Pilot Items

Method

Sample—The pilot scale was developed with 34 clinical participants and 29 clinicians. The clinical participants (18 women, 16 men, age range: 25–59) had diagnoses of depression (26 MD, 8 dysthymia) given by licensed health and mental health providers in the community. The median of years in the U.S. was 15; the median of years of education was nine. Six participants reported co-morbid diagnoses (3 anxiety disorder, 2 schizophrenia, 1 bipolar disorder).

The clinicians (16 women, 13 men) were 9 clinical social workers, 8 psychiatrists, 5 primary care and internal medicine physicians, 3 TCM physicians, 2 family therapists, and 2 psychologists. They had 5–27 years of post-licensure clinical experience with Chinese American immigrants. All provided treatment in Chinese except for two who used translators. All, also, were immigrants except for three who were U.S.-born children of immigrants.

Interviews—Clinical participants and clinicians gave an explanatory model of illness (20 clinical participants, 9 clinicians) or an item generation (14 outpatients, 20 clinicians) interview. This study used Kleinman’s (1986) nine-question semi-structured explanatory model interview, which elicits the name of illness, cause, impact, chief problem, severity, most feared consequence, course, appropriate treatment and desired outcome. Two questions

eliciting interpersonal distress (“How do you get along with your family members now?” and “How do you get along with other people in your life?”), added by Ying (1990), were also used. The nine clinicians who gave explanatory model interviews were interviewed about the explanatory models of 14 clinical participants who were their patients.

The item generation interview entailed evaluating a list of 60 symptom concepts in English drawn from the CES-D (Y.-W. Ying, 1988); BDI (Yeung, Neault, et al., 2002), PHQ-9 (Huang et al., 2006), and four culturally adapted scales (T. A. Cheng, Wu, Chong, & Williams, 1990; Lin, 1989; North East Medical Services, 1999; Y. Zheng & Lin, 1991). Clinical participants identified their current symptoms, symptoms at the onset of their first depressive episode, and symptoms from the list that they had not named spontaneously. The interviewer noted the symptom expressions they used. Clinicians first rated each concept as *uncommon*, *common* or *very common* for their clients and provided typical Chinese language expressions, then identified relevant concepts not listed and concepts with differential usage by gender.

A master-level research assistant conducted interviews with clinical participants in Cantonese or Mandarin. She took notes during interviews and wrote transcripts of explanatory models after interviews. Clinical participants received US\$25 for 30–45 minute one-on-one interviews. The PI conducted one-on-one interviews with clinicians, who were not remunerated, in English and Cantonese. Data was collected April of 2007 to January of 2008.

Analysis of data—First, to define the construct map, the PI and the master-level research assistant formulated descriptions of persons and items based on the explanatory model interviews. The description of items focused on the nature of the psychological, somatic and interpersonal (or social) domains of distress. The description of items also included experts’ assignment of pilot items into domains and levels of severity, which would represent their theoretical understanding of the role of items.

Second, to generate pilot items, the PI and five undergraduate research assistants developed a system for coding symptom concepts in explanatory model interviews, coded independently, then met to resolve differences by consensus. Concepts common to at least three participants were retained. For item generation interviews, symptom concepts that at least one-half of clinicians endorsed as *uncommon* or *very common* or at least two clinical participants cited were retained. A group of three experts, with information from interviews provided by the research assistant and PI, chose Chinese expressions for 47 retained concepts to comprise a pilot scale, CADS-47. They were instructed to choose commonly used expressions—understandable to speakers of Cantonese and Mandarin—that could be self-administered by a person with some middle school education. The PI and research assistant translated items to English. CADS-47 was pre-tested with eight community members and five social workers.

With the pilot items, five experts completed the theoretical description of items by independently classifying them into domains and levels of severity. The PI then communicated with each expert to share discrepant classifications and request reclassifications to achieve consensus. The PI also independently identified items that overlapped with the DSM-IV definition of MD. Experts confirmed the face validity of the final construct map.

Third, to define the response set, the focus group adapted the Likert scale of the PHQ-9. The PHQ-9 was chosen to maintain consistency and the possibility of comparison with a measure that operationalizes the DSM-IV construct of depression.

Results

Construct map—Figure 1 shows the construct map. Four characteristics described persons with depression. “Reduction in efficiency and functional activity” referred to losses in efficiency with regard to work, study, maintaining family and social relationships, home and self-care, and daily life tasks. “Maladaptation to environment” referred to the impact of difficulties in social, economic, cultural, physical, and religious or spiritual adjustment in the U.S., especially with regard to immigration and past traumas such as the Cultural Revolution and the loss or separation of family members. “Doubt and despair concerning social role and existence” referred to painful experiences of self-criticism, hopelessness, poor social belonging, and self-destructive impulses. Finally, “likelihood of family history of depression” concerned a genetic component of illness.

Regarding the description of items, the psychological domain concerned emotional and cognitive disturbances, with strong concern over one’s loss of ability to control disturbances. Emotional states included anxiety focused on interpersonal concerns (e.g., fear of losing one’s ability to provide for family). The somatic domain covered a range of experiences regarding diminished physical, mental and psychosocial functioning. The social domain included negative cognitions and affective and behavioral experiences regarding difficulties in social relationships and fulfilling one’s social role.

Item generation—The item generation procedure produced 47 items—with 23 items overlapping with DSM-IV symptoms of MD. For some DSM-IV items, the focus group chose expressions that were common idiomatic terms (26-*useless*, 18-*unhappy* and 17-*as-good*) and incorporated examples to assist understanding (20-*interest* and 9-*memory*). For items related to social withdrawal, the group chose terms that communicated the avoidance of interpersonal distress (35-*socialize* and 28-*talk*). For worthlessness, the expressions referred to failure in one’s social role (36-*burden*, 26-*useless*, 17-*as-good*). Experts had most difficulty with finding easy-to-understand expressions for psychomotor retardation and mental impairment (19-*thinking*, 16-*started* and 15-*decisions*).

Among items that did not overlap with DSM-IV symptoms of MD, many idiomatic expressions and culturally based concepts were employed. 7-*bored* and 3-*troubled*, with poor translations in English, are common expressions of mild distress. 34-*face* and 12-*fate* are concepts central to Chinese culture. 22-*ability-fear* and 14-*lonely* concern strong interpersonally based responsibilities and needs, respectively, but their English translations may be interpreted as self-focused emotional states. 29-*afraid*, 10-*nervous* and 4-*worry*, also, concern anxiety arising directly from interpersonal problems, but the English expressions also reflect self-focused states.

Large numbers of pilot items were assigned to each domain. The experts classified the 47 items as: 16 psychological, 14 somatic, 10 social, 5 psychological-social, 1 psychological-somatic, and 1 somatic-social. With regard to severity, they easily agreed on the severe and mild items but showed some disagreement on moderate severity items. The final classification showed psychological and social items falling along the whole continuum and somatic items mostly occupying the moderate range. The experts also predicted that men more easily expressed 23-*angry*, 7-*bored* and 32-*respect* and that women more easily expressed 9-*memory*, 39-*crying*, 33-*fault* and 30-*meaning*.

Outcome space—Qualitative descriptors and numbers of days were added to the Likert scale of the PHQ-9 to promote clarity and define distinct options, as required by an item response approach. The resulting response options were: 0 (*not at all/no days/0 days*), 1 (*a little bit/several days/1–3 days*), 2 (*quite a bit/about half the days/4–10 days*) and 3 (*extremely/nearly every day/11–14 days*). The focus group strongly recommended

dichotomous response options over the Likert scale and inquiring about symptoms over a one-week rather than two-week period to promote understanding. The researchers, however, rejected dichotomous response options because they likely would not provide enough information to discriminate adequately among respondents. The researchers also opted for two weeks for the purpose of comparisons with the PHQ-9.

The pre-testing of CADS-47 led to minor modifications. Three items were revised to increase their discriminability (e.g., from “lonely” to “very lonely”).

Discussion

Findings supported the hypothesized culture-based construct of depression with multiple domains of distress. The item generation procedure produced ample items regarding each domain—and most items did not overlap with DSM-IV symptoms of MD. Idiomatic expressions, also, were most appropriate for communicating some of the DSM-IV concepts. Items that were culture-specific in the social and psychological domains, furthermore, reflected an interdependent self-construal. Overall, findings supported the need of empirical modeling to examine the quantitative impact of culture-based expressions of depression on assessment.

Study 2 – Comprehensive Measure

Method

Participants—The sample was designed to include a broad representation of the community, similar numbers of clinical and community men and women, and people with all levels of depression severity. The inclusion criteria were: immigrant, oral Chinese fluency, and 21–60 years of age. Clinical participants had a current diagnosis of MD or dysthymia given by a licensed clinician in the community. Participants with comorbid psychiatric diagnoses were not excluded. People at risk of severe distress due to participation were excluded.

Twenty-three clinicians and service providers who worked in 12 study sites (7 clinical, 5 community) were recruiters. Clinicians recruited clients to be clinical participants. Service providers recruited clients and others associated with their organizations to be community participants. In a snowballing procedure, community participants recruited their acquaintances. Recruiters did not disclose participants’ identifying or health information. Participants thus self-reported their diagnoses, which were not confirmed by recruiters or by independent assessment.

Analysis of Data—The Rasch (1960) partial credit model (PCM)—appropriate for analyzing Likert-scale response sets of psychological phenomena (Embretson & Reise, 2000)—was used to analyze the data (Masters & Wright, 1996). In Rasch analysis, the probability a person will respond in a certain way to a particular item is modeled as a logistic function of the relative distance between the item and person locations on the latent trait (Bond & Fox, 2001). Maximum likelihood estimation is used to transform ordinal raw scores into linear, interval scale measurement units called *logits* (logarithm of odds). Item difficulty and person ability are thus placed on the same logit scale.

Empirical models were derived through a process of item reduction and recalibration to improve measurement properties. For each calibration, item fit, the proper usage of response options and overall fit were evaluated (Smith, 2000). Infit and Outfit item fit statistics are based on mean squared standardized residuals between observed responses and the model probability. With a mean of 1.0, values of .66 to 1.33 were considered to fit the model (Wilson, 2005); a more lenient rule of .5 to 1.5 was applied when justified substantively.

(Linacre & Wright, 1994). Higher fit statistics indicate poorer fit and lower statistics indicate closer fit than expected. Primary importance was given to Infit—weighted estimates that give more value to on-target observations (Linacre, 2002). Outfit statistics—unweighted estimates that show the influence of off-target observations—were applied if more information to determine item exclusion was needed. Verifications of the usage of response option followed standard procedures (Masters & Wright, 1996). Improvements in the overall fit of models were indicated by reductions in total deviance and increases in reliability. Rasch reliability coefficients, conceptually analogous to Cronbach's alpha (Bond & Fox, 2001), were considered adequate from .80 to .90 and optimal if greater than .90 (Traub, 1994). Reliability coefficients based on expected a posteriori/plausible values rather than maximum likelihood estimation were reported given their higher accuracy for small sample sizes.

A substantive evaluation of the final model was conducted by comparing a mapping of the estimated distributions of items and persons, called a *Wright map*, and the construct map. Observations were made regarding differences in the locations of items, the number of items per level of severity and per domain, and the number of items that overlap with the DSM-IV definition of MD.

DIF was analyzed by gender, age, education, English and years in the U.S. The latter three variables, as proxies for acculturation, were interpreted as a unit. For English proficiency, an average score of self-report questions on reading, writing and speaking was calculated. Subgroups for each variable were determined by substantive rationale or empirical analysis of cutoff values that maximized the detection of items with statistically significant DIF. In DIF analysis, separate models are estimated for each subgroup to observe the size of differences between item estimates. Differences in item difficulty were tested for significance at $\alpha = 0.05$ by using the joint standard errors to calculate a *t*-statistic. Differences of item estimates, called *effect sizes*, were classified as large (>.638), intermediate (.426–.638), or small (<.426) (Longford, Holland, & Thayer, 1993). Intermediate and large effect sizes were considered an important source of bias; small effect sizes were considered a minimal source (Paek, 2002).

Item response analyses were conducted with ACER ConQuest Version 2.0 (Wu, Adams, Wilson, & Haldane, 2007). All other analyses were conducted with PASW Statistics 18 (SPSS Inc., 2009).

Instruments—Three instruments were used to assess convergent validity: Neurasthenia Questionnaire (NTQ), Acculturative Stress Questionnaire (ASQ-14) and PHQ-9.

Acculturative stress is considered a proximal risk factor of depression

NTQ is a diagnostic measure of current neurasthenia based on the criteria of the Chinese Classification of Mental Disorders, 2nd edition (Chinese Medical Association, 1995). For this study, two master-level research assistants translated and back-translated an English version (Paralikar, Sarmukaddam, Agashe, & Weiss, 2007). A diagnosis is given if at least three of five groups of symptoms (weakness, emotional, mental agitation, nervous pain, and sleep disturbance) are present over three months and one of three conditions (interference with daily activities, significant and persistent distress, and help or treatment sought) is met.

ASQ-14 was a new measure developed in the current study with the analytic procedure described for empirical modeling with the PCM. Its reliability was .87. Although ASQ-14 was not validated outside of the current study, it served as secondary culture-based construct with which to assess convergent validity considering that acculturative stress is a proximal

risk factor for depression (Hwang & Ting, 2008). The scale encompasses social, cultural, biological, and environmental variables associated with acculturative stress (Berry, Kim, Minde, & Mok, 1987). It inquires about: cultural differences, language, one's own work or study, family's work or study, housing and safety, transportation, finances, maintaining health, illness, health care, missing native country, missing family and friends, social support and racial discrimination. Response options range from 0 (*not at all*) to 3 (*always*).

The PHQ-9 has one item stem that inquires about symptom occurrence over the preceding two weeks. Items are scored from 0 (*not at all*) to 3 (*nearly every day*) (Kroenke et al., 2001). The cutoff score is 15 for a provisional diagnosis of MD. Severity ranges are: minimal symptoms (< 10), moderate symptoms or minor depression (10–14), moderately severe MD (15–19), and severe MD (≥ 20).

A participant questionnaire gathered sociodemographic, language, immigration, and mental illness information. The questionnaire assessed English proficiency with questions on reading, writing and speaking, with options from 0 (*very poor*) to 3 (*excellent*). It also asked for a self-evaluation of current depression symptoms, with options of *mild*, *moderate* and *severe*.

Procedures—Research assistants administered instruments in Cantonese, Mandarin, or Toishanese. The first was either CADS-47 or PHQ-9, which were switched randomly. The order of remaining measures was NTQ, ASQ-14 and the participant questionnaire. Research assistants did not explain CADS-47 and PHQ-9 items when asked, but encouraged participants to apply their own understanding. They noted difficult-to-understand items, which became the object of *think aloud* investigations in which later participants were asked to talk aloud about their thoughts while responding (American Institutes for Research, 2000). Participants received \$40 for 30–90 minute interviews. Data was collected from April 2008 to April 2009. It concluded once the sample size was sufficient for model estimation, which, for item response analysis, in general, is when every response option obtains a minimum count of five or six (Linacre, 1994).

Assessment of Validity—Concurrent validity was assessed with the point biserial correlation of CADS-42 scores with the study's criterion standard. Convergent validity was assessed with point biserial correlations of CADS-42 scores with diagnoses given by NTQ and PHQ-9 and Pearson product-moment correlations of CADS-42 scores with ASQ-14 scores. Raw scores were used for all correlations.

Results

Participant characteristics—Table 1 shows participants' characteristics. Participants were born in China (74.4%), Hong Kong (17.6%), Taiwan (4.8%) and other countries (2.7%). Primary dialects were Cantonese (70.9%), Mandarin (18.9%), and other dialects (10.1%). Cantonese speakers also spoke Mandarin (75.9%) and vice versa (41.8%). Those who spoke other dialects also spoke Mandarin (52.2%) or Cantonese (47.8%). Clinical participants' self-reported diagnoses were: MD (41.7%), dysthymia (8.7%), "don't know the name" (40.8%), and "no diagnosis" (8.7%). Participants with comorbid diagnoses (34.3%) reported anxiety disorder (13.7%), schizophrenia (12.7%), bipolar disorder (3.9%), "don't know the name" (2.0%), and neurasthenia (1.0%).

Estimated recruitment success rates were: 70.0% of community women, 55.0% of clinical women, 45.0% of community men and 25.0% of clinical men. The final two months of data collection were dedicated to recruiting clinical men. Dropout of recruited participants was low. Four missed interview appointments; one did not complete the interview due to poor concentration. No participants were excluded due to incomplete data. The sample had a low

representation of immigrants with higher acculturation and socioeconomic status due to failing to obtain a health maintenance organization as a study site.

Estimation of model—Beginning with CADS-47, four calibrations were conducted to arrive at a 42-item full model (see Figure 2 and Table 2). Five items were removed due to Infit values above 1.33. These items were: “You always want to sleep and don’t want to get out of bed”, “You have headaches”, “You drink or gamble to make yourself feel comfortable”, “You are severely bothered by family relationships or matters”, and “You have stomach pains or discomfort”. 38-*appetite* also had Infit (1.35) beyond the desired range, but was retained given that appetite disturbance is a basic indicator of emotional distress across cultures and a DSM-IV symptom of MD. The 42-item model had significantly better overall fit than the 47-item model ($\chi^2_{diff} = 2176.10$, $df = 15$, $p < 0.01$) and similar reliability (.98 both models). There was, also, no misuse of response options.

Substantive evaluation of model—A comparison of the Wright map (Figure 2) with the construct map (Figure 1) showed that CADS-42 represented the global construct adequately. First, in the Wright map, items were located in the upper half of the distribution of persons as expected. Second, estimated item locations matched theorized locations fairly well. Eighteen items matched on estimated and theorized severity levels. Sixteen items were nearly matched, with estimated severity levels one level from theorized levels. Only eight items were poorly matched, with estimated severity levels two levels from their theorized levels. The majority of poorly matched items, also, were social items whose severity levels were underestimated by experts. There was also an overall tendency of experts to underestimate the severity of items—the Wright map showed more items in the upper half of the continuum than theorized and fewer items in the lower half of the continuum than theorized.

With regard to the distribution of items by domain, eliminating five items with poor fit (3 somatic, 1 psychological, 1 social) changed the distribution of items very little. The 42 items were distributed as: 15 psychological, 11 somatic, 9 social, 5 psychological-social, 1 psychological-somatic, and 1 somatic-social. A pattern, also, was observed regarding each domain. First, psychological items occupied the entire continuum as theorized, yet there was only one strictly psychological item in the moderate-to-severe range. Second, social items were not present on the entire continuum as theorized, but occupied the moderate and moderate-to-severe levels exclusively. Third, the somatic domain generally occupied a moderate range as theorized, but this range was very broad, extending into the mild range.

Two items were object of think aloud investigations. First, for 40-*psychosis*, which 23.2% of participants affirmed, respondents expressed their mental confusion about whether an experience had occurred or had been imagined or dreamt; they also cited revisiting a dreamt experience when awake and imagination and fantasy. They thus used this item to communicate a problem of poor mental functioning. Second, for 25-*irritated*, difficulties with answering this item were due to poor understanding of the concept by some people with very little education.

Assessment of validity—The strong correlation of CADS-42 with the criterion standard ($r_{bp} = .61$) demonstrated concurrent validity. The correlations of CADS-42 with NTQ ($r_{bp} = .73$) and ASQ-14 ($r = .69$) demonstrated its convergence with emic constructs. The correlation of CADS-42 with PHQ-9 ($r_{bp} = .91$) showed its convergence with the Western construct of depression.

Analysis of DIF—Subgroups by age (34.4% 40 years, 65.6% > 40 years), education (26.9% less than high school, 73.1% high school or more), and English (45.1% very poor,

54.9% fair or better) were based on five experts' evaluations of ability to acculturate. The younger, more educated and those with stronger English were expected to acculturate more quickly. The cutoff for years in U.S. (41.0% \leq 7 years, 59.0% $>$ 7 years) was established empirically. (Items with statistically significant DIF are referred to as *biased* items.)

The impact of DIF was broad. Thirty of 42 items (71.4%) showed statistically significant DIF by one or more variables (see Tables 3 and 4). Twenty-four items had small effect sizes; five items had intermediate effect sizes; one item had a large effect size. DIF, also, affected every domain. Fourteen of 20 (70.0%) psychological items were biased. Eleven of 13 (84.6%) somatic items were biased. Eleven of 15 (73.3%) social items were biased.

The influence of acculturation-related variables was very strong. By one or more of the acculturation-related variables, 27 items (64.3% of items) were biased, with 5 having intermediate or large effect sizes. The impact of gender and age was weaker. By gender, nine items (21.4% of items) were biased, with one item having a large effect size. By age, eight items (19.0% of items) were biased, with one item having a large effect size.

Experts' predictions of DIF by gender were correct for 39-*crying*, 30-*meaning* and 9-*memory*, which were easier for women and for 32-*respect* and 7-*bored*, which were easier for men. Their predictions were incorrect, however, for 33-*fault* and 16-*angry*, which were unbiased.

Discussion

A valid and reliable comprehensive measure was created based on a construct defined in the community. The measure revealed the role of Chinese and Western symptom concepts and expressions in the manifestation of depression. Culturally based social and somatic symptoms, as well as anxiety and irritation concepts, were a significant part of the construct. Core DSM-IV psychological concepts were also valid—but communicated with Chinese expressions. Also, while some social symptom concepts appear to overlap with Western symptoms, there are also social items that do not overlap. Whether the identified domains or certain culture-specific symptoms have utility for improving the cultural sensitivity and accuracy of assessment are questions for future research.

DIF results clarified the challenge and complexity of accurate detection for Chinese American immigrants. A short instrument containing two or three items with intermediate and large effect sizes—or containing many items with small effect sizes—would likely be sufficient to create potential bias in total scores and over- or under-detection. The less acculturated, furthermore, appeared to risk under-recognition with U.S. scales due to their stronger difficulty with endorsing DSM-IV psychological (42-*suicidal*, 41-*self-harm*, 28-*talk*, and 20-*interest*) and somatic (38-*appetite*, 1-*fatigued*, and 9-*memory*) concepts. DIF analysis thus highlighted the extreme care and empirical testing required for developing a sufficiently unbiased scale and rule out the need of separate instruments for subgroups.

Study 3 – Screening Scale

Method

Study 3 uses the methods of Study 2 regarding sample, data analysis, and instruments. The strategy for choosing items and scale validation are described below.

Strategy for choosing items—DIF findings in Study 2 influenced the strategy for choosing items for a short scale. Considering the difficulty of eliminating bias by acculturation-related variables, the main focus of CADS-9 was to minimize bias by gender and age—and by acculturation secondarily. The Wright Map was used to choose several sets

of nine items that covered the upper severity range relevant to clinical depression. For each set of items, the selection criteria for items were: not repetitive conceptually nor in location, no display of DIF by gender or age, minimization of DIF by acculturation-related variables, and, if possible, overlapping with DSM-IV symptoms of MD.

Seven experts independently selected the set of items with highest content validity for screening. This final set of items, CADS-9, was calibrated. DIF was also analyzed although its results were unreliable because CADS-9 was not administered as a stand-alone measure. For comparison, one of the alternate 9-item scales, CADS-9A, was also calibrated and evaluated.

Assessment of validity—Several methods were used to validate CADS-9 in addition to those applied in Study 2. First, concurrent validity was assessed with receiver operating characteristic (ROC) analysis (Swets, 1995). Gender-specific optimal cutoff scores were derived to calculate area under the ROC curve (AUC) and performance indicators. Second, also in support of concurrent validity, participants' self-evaluated depression severity levels were compared with exploratory severity levels based on CADS-9 scores. Third, content validity was based on experts' evaluations and information from research assistants who administered study instruments. CADS-9, CADS-9A and PHQ-9 were also compared.

Results

Choice of items—Experts considered the following set of items to have the highest acceptability for screening: 41-*self-harm*, 35-*socialize*, 29-*afraid*, 24-*hide*, 23-*angry*, 18-*unhappy*, 11-*health*, 8-*concentrate*, and 4-*worried*. Four of these items, 42-*self-harm*, 35-*socialize*, 18-*unhappy*, and 8-*concentrate*, overlapping with DSM-IV symptoms of MD. Two items had DIF by acculturation-related variables. 23-*angry* was chosen due to a lack of other available items in the moderately severe range and to its small DIF effect size. Experts, also, noted that men were willing to disclose anger. 41-*self-harm*, with an intermediate DIF effect size, was chosen because it had the smallest DIF effect size among the three items located at the highest severity. It was also more culturally acceptable for screening than 42-*suicidal*.

Estimation of model and DIF—All items had acceptable Infit (.77–1.24). Verifications of the proper usage of response options were met. CADS-9 showed strong reliability (.90). There was, also, no statistically significant DIF by gender and age. Four items (41-*self-harm*, 35-*socialize*, 23-*angry*, and 11-*health*) showed DIF by acculturation-related variables with small effect sizes (.216–.404).

Assessment of validity—The moderately strong correlation of CADS-9 with the criterion standard ($r_{bp} = .58$) supported concurrent validity. The strong correlations of CADS-9 with NTQ ($r_{bp} = 0.71$), ASQ-14 ($r = .69$), and PHQ-9 ($r = .89$) supported convergent validity.

ROC-derived optimal cutoff scores were nine points for men and ten points for women. With these cutoffs, CADS-9 showed moderately strong rates of correct classification (.77 overall, .77 men, .76 women), sensitivity (0.79 overall, .79 men, 0.79 women), specificity (.75 overall, .76 men, .74 women), positive predictive value (.72 overall, .65 men, .77 women), and negative predictive value (.81 overall, .86 men, .75 women). AUC was also moderately high (.84 overall, .84 men, .83 women). These performance indicators are likely biased downward due to the unknown numbers of clinical participants who were asymptomatic at time of participation and community participants who were symptomatic but undiagnosed.

Figure 3 shows the distribution of CADS-9 scores. Nearly one-quarter (24.3%) of the clinical sample scored below the cutoff scores; one-quarter (25.0%) of the community sample scored above the cutoff scores. Among clinical participants, men had lower mean scores ($M = 13.9$, $SD = 6.70$) than women ($M = 14.7$, $SD = 6.04$). Among community participants, men also had lower mean score ($M = 5.68$, $SD = 5.02$) than women ($M = 6.74$, $SD = 5.90$).

The comparison of participants' self-evaluated severity levels (see Table 1) and severity levels based on CADS-9 scores showed high concordance, providing additional support for concurrent validity. Exploratory severity levels were created for CADS-9 based on the severity levels of PHQ-9, which considers every five points as one level of severity. The levels for men and women, respectively, on CADS-9 were thus: Insignificant symptoms (< 4 and < 5), minimal symptoms (4–8 and 5–9), minor depression (9–13 and 10–14), moderately severe MD (14–18 and 15–19), and severe MD (19 and 20). Applying these levels, only small proportions of self-evaluations (13.9% clinical, 11.7% community) did not match their CADS-9 severity groups within one level. Also, only small proportions of participants (8.9% clinical, 10.0% community) were detected as more severe by CADS-9 than self-predictions.

Comparison with PHQ-9—Overall, the reliability and validity indicators between CADS-9 and PHQ-9 were very similar. The reliability of PHQ-9 (.91) was slightly higher than that of CADS-9. PHQ-9 also correlated a little more strongly with the criterion standard ($r = .62$) than CADS-9. PHQ-9, however, converged more weakly than CADS-9 with NTQ ($r_{bp} = .68$) and ASQ-14 ($r = .59$).

With nine points as the ROC-derived optimal cutoff score for men and women, PHQ-9 showed slightly better performance for men but equal performance for women compared with CADS-9. Two differences in scoring patterns between PHQ-9 and CADS-9 were observed. First, the difference in mean scores on the PHQ-9 between men in the clinical sample ($M = 15.2$, $SD = 7.02$) and the community sample ($M = 5.15$, $SD = 4.72$) was larger than on CADS-9. Second, contrary to CADS-9 results in the clinical sample, men's mean score was higher than women's mean score ($M = 14.7$, $SD = 6.72$) on the PHQ-9.

Comparison with CADS-9A—The alternative scale was comprised of items that assessed the higher severity range more closely than CADS-9. Its items were: 41-*self-harm*, 37-*blaming*, 36-*burden*, 34-*lostface*, 33-*ownfault*, 31-*heart*, 26-*useless*, 12-*fate*, and 5-*ability-loss*. All items had acceptable fit statistics and response option verifications were met.

CADS-9A showed slightly higher performance indicators than CADS-9 and PHQ-9. CADS-9, however, converged more strongly with NTQ and ASQ-14 than CADS-9A. CADS-9A, in comparison with PHQ-9, converged more weakly with NTQ, but more strongly with ASQ-14.

Content Validity—Experts evaluated CADS-9 as having had higher content validity than CADS-9A and PHQ-9. CADS-9 covered topics commonly shared in daily social life, required easily accessible levels of self-awareness, and excluded more shameful and stigmatized experiences. The items of CADS-9 were also easier to understand and more appropriate due to emphasizing a mild and moderate range of depression severity. CADS-9A was inappropriate for screening due to its emphasis on critical cognitive evaluations about oneself and one's social relationships, which requires a deeper level of reflection and willingness to disclose.

The PHQ-9 contained four problematic items: "Feeling down, depressed, or hopeless", "Trouble falling asleep or staying asleep, or sleeping too much", "Poor appetite or

overeating”, and “Moving or speaking so slowly that other people could have noticed. Or, the opposite—being so fidgety or restless...” Regarding the first three items, some participants, especially those with little education, were unsure how to answer when they experienced only one of two or three concepts contained in these items. From an item response perspective, also, is not possible to evaluate the validity of these items because they contain more than one concept—since it is unknown which one the respondent affirmed. Regarding the fourth question, participants commonly asked, “How do I know what others are thinking?” Experts, also, viewed overeating and agitation as invalid concepts—because Chinese do not consider these as problems—and the term “depressed” as inappropriate for people with little education.

Discussion

A reliable and valid screening tool was developed from a comprehensive measure of the culture-based construct of depression. The stronger convergence of CADS-9 than PHQ-9 with constructs of neurasthenia and acculturative stress show that constructs operationalized in these two scales diverge. CADS-9, also, with only culturally relevant concepts and expressions, had stronger content validity than PHQ-9. Whether CADS-9 will provide higher accuracy and cultural sensitivity than the PHQ-9 requires testing in the field. CADS-9, however, is expected to be more culturally sensitive even though its content overlapped with DSM-IV symptoms of MD.

Health care providers may require training to understand the meaning of the culture-specific concepts. 29-*afraid* and 4-*worry* are anxiety symptoms, but, in the Chinese lens of depression, can signify a state of interpersonal distress that is intimately associated with depression. Talking to a patient about his or her objects of fear and worry may lead to understanding the interpersonal context that surrounds the person’s depressed state. The use of anxiety symptoms to assess depression, also, is opportune given that the presentation of anxiety is little stigmatized and not biased by method of report for Asian Americans (Okazaki, 2000). 23-*angry* signifies, for Chinese, distress associated with the loss of control of one’s mind and emotions and with one’s poor fit with the social environment.

In the process of developing and validating CADS-9, important observations were gained about screening instruments and the task of screening. First, many sets of nine items could have been drawn from the comprehensive measure to develop a scale that performs comparably with CADS-9 and PHQ-9. CADS-9A, for instance, with mostly culture-specific content, performed slightly better than CADS-9 and PHQ-9. Second, the clinical status of individuals who are moderately or severely depressed, also, can easily be detected given that it is very easy for this group to score beyond the cutoff. A screening measure, thus, must perform well in the mild range, near the cutoff value. Third, the validity indicators found for CADS-9 and PHQ-9 may not be as favorable as they appear. A small group of clinical participants (16.8%) scored below the CADS-9 cutoff but self-evaluated as having mild or moderate symptoms. With these observations, a next of research is to explore assessment needs and forms of manifestation in the low severity range.

Concluding Discussion

The study demonstrated the utility of the item response approach, coupled with ethnographic methods, for generating detailed knowledge of a culturally based mental illness construct and for identifying the sociocultural factors that influence its expression. The study also showed the practical utility of item response techniques for creating unbiased and efficient measures with a small sample size. With these methods, it was possible to confirm the cultural nuances of the Chinese American manifestation of depression and to provide an empirical model with which to further examine how assessment can be improved.

Acknowledgments

This study was funded by the National Institute of Mental Health (R36MH080607), Chinese Community Health Care Association and Fahs-Beck Fund for Research and Experimentation. We thank Donaldina Cameron House, Chinatown Child Development Center, Richmond Area Multi-Services (RAMS), Inc., Chinatown Northbeach Mental Health Services, Asian Health Services and Asian Community Mental Health Services for their participation as study sites.

References

- Abe-Kim J, Takeuchi DT, Hong S, Zane N, Sue S, Spender MS, et al. Use of mental health-related services among immigrant and US-born Asian Americans: Results of the National Latino and Asian American Study. *American Journal of Public Health*. 2007; 97(1):91–98. [PubMed: 17138905]
- Beck AT. An inventory for measuring depression. *Archives of General Psychiatry*. 1961; 4:53–61.
- Berry JW, Kim U, Minde T, Mok D. Comparative studies of acculturative stress. *International Migration Review*. 1987; 21:491–511.
- Boey KW. Cross-validation of a short form of the CES-D in Chinese elderly. *International Journal of Geriatric Psychiatry*. 1999; 14:608–617. [PubMed: 10489651]
- Bond, TG.; Fox, CM. Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
- Chan B, Parker G, Tully L, Eisenbruch M. Cross cultural validation of the DMI-10 measure of state depression: The development of a Chinese language version. *J Nerv Ment Dis*. 2007; 195(1):20–25. [PubMed: 17220735]
- Chan DW. The Beck Depression Inventory: What difference does the Chinese version make? *Psychological Assessment: A Journal of Consulting and Clinical Psychology*. 1991; 3(4):616–622.
- Chen H, Guarnaccia PJ, Chung H. Self-attention as a mediator of cultural influences on depression. *International Journal of Social Psychiatry*. 2003; 49(3):192–203. [PubMed: 14626362]
- Chen S, Chiu H, Xu B, Ma Y, Jin T, Wu M, et al. Reliability and validity of the PHQ-9 for screening late-life depression in Chinese primary care. *International Journal of Geriatric Psychiatry*. 2009; 25:1127–1133. [PubMed: 20029795]
- Chen TM, Huang FY, Chang C, Chung H. Using the PHQ-9 for depression screening and treatment monitoring for Chinese Americans in primary care. *Psychiatric Services*. 2006; 57(7):976–981. [PubMed: 16816282]
- Cheng T-A, Williams P. The design and development of a screening questionnaire (CHQ) for use in community studies of mental disorders in Taiwan. *Psychological Medicine*. 1986; 16:415–422. [PubMed: 3726013]
- Cheng TA, Wu JT, Chong MY, Williams P. Internal consistency and factor structure of the Chinese Health Questionnaire. *Acta Psychiatrica Scandinavica*. 1990; 82(4):304–308. [PubMed: 2260484]
- Cheung C-K, Bagley C. Validating an American scale in Hong Kong: the Center for Epidemiologic Studies Depression Scale (CES-D). *The Journal of Psychology*. 1998; 132(2):169–186. [PubMed: 9529665]
- Chung H, Teresi J, Guarnaccia PJ, Meyers BS, Holmes D, Bobrowitz T, et al. Depressive symptoms and psychiatric distress in low income Asian and Latino primary care patients: Prevalence and Recognition. *Community Mental Health Journal*. 2003; 39(1):33–46. [PubMed: 12650554]
- Diener E, Diener M. Cross-cultural correlates of life satisfaction and self-esteem. *Journal of Personality and Social Psychology*. 1995; 68(4):653–663. [PubMed: 7738768]
- Embretson, SE.; Reise, SP. Item response theory for psychologists. Mahwah: Lawrence Erlbaum Associates; 2000.
- Gupta R, Yick A. Validation of CES-D scale for older Chinese immigrants. *Journal of Mental Health and Aging*. 2001; 7(2):257–273.
- Hong GK, Lee BS, Lorenzo MK. Somatization in Chinese American clients: Implications for psychotherapeutic services. *Journal of Contemporary Psychotherapy*. 1995; 25(2):105–118.

- Huang RY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *Journal General and Internal Medicine*. 2006; 21:547–552.
- Hwang W-C, Ting JY. Disaggregating the effects of acculturation and acculturative stress on the mental health of Asian Americans. *Cultural Diversity and Ethnic Minority Psychology*. 2008; 14(2):147–154. [PubMed: 18426287]
- Kleinman, A. *Social origins of distress and disease: Depression, neurasthenia and pain in modern China*. New Haven: Yale University Press; 1986.
- Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *Journal of General and Internal Medicine*. 2001; 16:606–613.
- Kung W. Cultural and practical barriers to seeking mental health treatment for Chinese Americans. *Journal of Community Psychology*. 2004; 32:24–37.
- Kuo WH. Prevalence of depression among Asian-Americans. *Journal of Nervous and Mental Disease*. 1984; 172:449–457. [PubMed: 6747614]
- Kwan VSY, Bond MH, Singelis TM. Pancultural explanations for life satisfaction: Adding relationship harmony to self-esteem. *Journal of Personality and Social Psychology*. 1997; 73(5):1038–1051. [PubMed: 9364759]
- Lai BPY, Tang AKL, Lee DTS, Yip ASK, Chung TKH. Detecting postnatal depression in Chinese men: A comparison of three instruments. *Psychiatry Research*. 2010; 180:80–85. [PubMed: 20493548]
- Lin N. Measuring depressive symptomatology in China. *Journal of Nervous and Mental Disease*. 1989; 177(3):121–131. [PubMed: 2918295]
- Linacre JM. What do infit and outfit mean-square and standardized mean? *Rasch Measurement Transactions*. 2002; 16(2):878.
- Linacre JM, Wright BD. Reasonable mean-square fit values. *Rasch Measurement Transactions*. 1994; 8(3):370.
- Liu S-I, Yeh Z-T, Huang H-C, Sun F-J, Tjung J-J, Hwang L-C, et al. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. *Comprehensive Psychiatry*. 2011; 52:96–101. [PubMed: 21111406]
- Longford, NT.; Holland, PW.; Thayer, DT. Stability of the MH D-DIF statistics across populations. In: Holland, PW.; Wainer, H., editors. *Differential item functioning: Theory and practice*. Hillsdale: Erlbaum; 1993.
- Mak W, Zane N. The phenomenon of somatization among community Chinese Americans. *Social Psychiatry and Psychiatric Epidemiology*. 2004; 39:967–975. [PubMed: 15583904]
- Mak WWS, Law RW, Teng Y. Cultural model of vulnerability to distress: The role of self-construal and sociotropy on anxiety and depression among Asian Americans and European Americans. *Journal of Cross-Cultural Psychology*. 2011; 41(1):75–88.
- Masters, GN.; Wright, BD. The partial credit model. In: Van der Linden, WJ.; Hambleton, RK., editors. *Handbook of modern item response theory*. New York: Springer; 1996. p. 101-122.
- Norasakkunkit V, Kalick SM. Culture, ethnicity, and emotional distress measures: The role of self-construal and self-enhancement. *Journal of Cross-Cultural Psychology*. 2002; 33(1):56–70.
- North East Medical Services. *Short Depression Scale (unpublished instrument)*. San Francisco: Author; 1999.
- Okazaki S. Sources of ethnic differences between Asian-American and White American college students on measures of depression and social anxiety. *Journal of Abnormal Psychology*. 1997; 106:52–60. [PubMed: 9103717]
- Okazaki S. Asian American and White American differences on affective distress symptoms: Do symptom reports differ across reporting methods? *Journal of Cross-Cultural Psychology*. 2000; 31(5):603–625.
- Paek, I. Unpublished Dissertation. University of California; Berkeley: 2002. Investigations of differential item functioning: comparisons among approaches, and extension to a multidimensional context.

- Paralikar V, Sarmukaddam S, Agashe M, Weiss MG. Diagnostic concordance of neurasthenia spectrum disorders in Pune, India. *Social Psychiatry and Psychiatric Epidemiology*. 2007; 42:561–572. [PubMed: 17476439]
- Parker G, Chan B, Tully L, Eisenbruch M. Depression in the Chinese: The impact of acculturation. *Psychological Medicine*. 2005; 35:1475–1483. [PubMed: 16164771]
- Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1:385–401.
- Rankin SH, Galbraith ME, Johnson S. Reliability and validity data for a Chinese translation of the Center for Epidemiological Studies-Depression Scale. *Psychological Reports*. 1993; 73(3):1291–1298. [PubMed: 8115582]
- Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Paedagogiske Institute; 1960.
- Singelis TM, Bond MH, Sharkey WF, Lai CSY. Unpackaging culture's influence on self-esteem and embarrassment: The role of self-construals. *Journal of Cross-Cultural Psychology*. 1999; 30(3): 315–341.
- Smith RM. Fit analysis in latent trait measurement models. *Journal of Applied Measurement*. 2000; 1:199–218. [PubMed: 12029178]
- SPSS Inc. PASW Statistics 18. Chicago: Author; 2009.
- Traub, RE. Reliability for the social sciences. Thousand Oaks: Sage; 1994.
- Tung M. Symbolic meanings of the body in Chinese culture and “somatization”. *Culture, Medicine and Psychiatry*. 1994; 18(4):483–492.
- Wilson, M. Constructing measures: An item response modeling approach. Mahwah: Lawrence Erlbaum Associates; 2005.
- Wu, ML.; Adams, RJ.; Wilson, M.; Haldane, S. ACER ConQuest version 2.0: General item response modeling software. Hawthorne: Australian Council for Educational Research Press; 2007.
- Yen S, Robins CJ, Lin N. A cross-cultural comparison of depressive symptom manifestation: China and the United States. *Journal of Consulting and Clinical Psychology*. 2000; 68:993–999. [PubMed: 11142551]
- Yeung A, Chang D, Gresham RL, Nierenberg AA, Fava M. Illness beliefs of depressed Chinese American patients in primary care. *Journal of Nervous and Mental Disease*. 2004; 192(4):324–327. [PubMed: 15060408]
- Yeung A, Fung F, Yu S, Vorono S, Ly M, Wu S, et al. Validation of the Patient Health Questionnaire-9 for depression screening among Chinese Americans. *Comprehensive Psychiatry*. 2008; 49(2):211–217. [PubMed: 18243896]
- Yeung A, Howarth S, Chan R, Sonawalla S, Nierenberg AA, Fava M. Use of the Chinese version of the Beck Depression Inventory for screening depression in primary care. *Journal of Nervous and Mental Disease*. 2002; 190(2):94–99. [PubMed: 11889362]
- Yeung A, Neault N, Sonawalla S, Howarth S, Nierenberg AA, Fava M. Screening for major depression in Asian-Americans: A comparison of the Beck and the Chinese Depression Inventory. *Acta Psychiatrica Scandinavica*. 2002; 105:252–257. [PubMed: 11942928]
- Ying Y-W. Depressive symptomatology among Chinese-Americans as measured by the CES-D. *Journal of Clinical Psychology*. 1988; 44:739–746. [PubMed: 3192712]
- Ying Y-W. Nonresponse on the Center for Epidemiologic Studies Depression Scale among Chinese Americans. *International Journal of Social Psychiatry*. 1989; 35(2):156–163. [PubMed: 2767921]
- Ying Y-W. Explanatory models of major depression and implications for help-seeking among immigrant Chinese-American women. *Culture, Medicine and Psychiatry*. 1990; 14:393–408.
- Ying Y, Lee PA, Tsai JL, Yeh Y, Huang JS. The conception of depression in Chinese American college students. *Cultural Diversity and Ethnic Minority Psychology*. 2000; 6(2):183–195. [PubMed: 10910531]
- Zheng Y-P, Lin KM, Takeuchi D, Kurasaki KS, Wang Y, Cheung F. An epidemiological study of neurasthenia in Chinese-Americans in Los Angeles. *Comprehensive Psychiatry*. 1997; 38(5):249–259. [PubMed: 9298316]

- Zheng Y, Lin K. Comparison of the Chinese Depression Inventory and the Chinese version of the Beck Depression Inventory. *Acta Psychiatrica Scandinavica*. 1991; 84(6):531–536.
- Zheng Y, Wei L, Goa L, Zhang G, Wong C. Applicability of the Chinese Beck Depression Inventory. *Comprehensive Psychiatry*. 1988; 29(5):484–489. [PubMed: 3180758]
- Zheng Y, Xu L, Shen Q. Styles of verbal expression of emotional and physical experiences: A study of depressed patients and normal controls in China. *Culture, Medicine and Psychiatry*. 1986; 10:231–243.

Appendix A

美國華人移民心理健康問卷
Chinese American Depression Scale (CADS-9)

姓名：NAME _____

日期：DATE _____

在過去兩星期，你是否被以下事情所困擾？如果是，被困擾多少天？請讀題並圈上最合適的答案。
Over the last two weeks were you bothered by the following problems? If so, how many days? Please read each statement carefully and circle the most appropriate answer.

| | 完全沒有/ 沒有一天 (0天) Not at all/ No days | 非常少/ 幾天 (1-3天) A little bit/ A few days | 有時/大約 一半時間 (4-10天) Quite a bit/ About half the days | 經常是/差 不多一天 (11-14天) Extremely/ Nearly every day |
|--|---|---|--|--|
| 1. 很多事情讓你感到很擔心。 Many things make you feel very worried. | 0 | 1 | 2 | 3 |
| 2. 你難以集中精神。 You are unable to concentrate well. | 0 | 1 | 2 | 3 |
| 3. 你非常害怕自己健康有問題。例如，患癌症或心臟病。 You are very afraid that you have health problems. For example, you might have cancer or heart disease. | 0 | 1 | 2 | 3 |
| 4. 你感到非常不開心。 You feel very unhappy. | 0 | 1 | 2 | 3 |
| 5. 你很容易發脾氣和發怒。 You have tantrums and get angry very easily. | 0 | 1 | 2 | 3 |
| 6. 你隱瞞自己生活有困難。 You hide your life difficulties from other people. | 0 | 1 | 2 | 3 |
| 7. 你感到很害怕。 You feel very afraid. | 0 | 1 | 2 | 3 |
| 8. 你完全不想和別人接觸、交往或外出。 You don't want to have contact with people, socialize, or go out at all. | 0 | 1 | 2 | 3 |
| 9. 你想過傷害自己。 You have thought about hurting yourself. | 0 | 1 | 2 | 3 |

合計：
ADD COLUMNS: [] + [] + []
總計：
TOTAL: []

CADS-9 may be used and duplicated freely. The development of CADS-9 was funded by the National Institute of Mental Health (R36MH080607), Chinese Community Health Care Association of San Francisco, and Fahn/Beck Fund for Research and Experimentation.

How to use CADS-9 Chinese American Depression Scale

- **Eligibility:** CADS-9 is for adults, ages 21–60, who speak and understand Chinese fluently. It is intended for the screening of symptoms associated with depression in medical clinics and social services agencies.

- **Directions:** CADS-9 is a self-report instrument that may be administered by a healthcare or social services provider, or self-administered. Read each item as written. For example, do not change “veryworried” to “extremely worried”. Make sure that the respondent understands the meaning of the respective answers (0, 1, 2 and 3) in terms of the number of days over the past two weeks. Encourage the respondent to answer each item with his or her own understanding and judgment.
- **Cutoff Score:** A total score of 10 or more points for women and 9 or more points for men indicate risk of major depression. Use CADS-9 as an initial screen, rather than a means of clinical diagnosis.
- **Severity Levels:** Higher total scores indicate more severe depression.
 - Minimal symptoms: 5–9 *women*, 4–8 *men*.
 - Mild: 10–14 *women*, 9–13 *men*.
 - Moderate: 15–19 *women*, 14–18 *men*.
 - Severe: 20+ *women*, 19+ *men*.
 - Provide education and follow-up to people who have “minimal symptoms”. Provide referral to mental health assessment and treatment to people who score “mild”, “moderate”, and “severe”.
- **Other Information:** The first items indicate milder depression, and the last items indicate more severe depression. For example:
 - Mild: 1-*worry*, 2-*poor concentration*.
 - Moderate: 3-*health concerns*, 4-*unhappiness*, 5-*anger*, 6-*hiding difficulties*.
 - Severe: 7-*fear*, 8-*social avoidance*, 9-*self-harm*.

An individual may be at risk of major depression even if his or her total score is very low or below the cutoff score. Consider providing further assessment and referral to anyone who answers:

- 2 (quite a bit) or 3 (extremely) to several of the nine items, or
 - 1 (a little bit), 2 (quite a bit), or 3 (extremely) to 8-*social avoidance* or 9-*self-harm*.
- **Scale Development:** CADS-9 was developed with a sample of 227 Chinese immigrant adults in the San Francisco metropolitan area during 2008–09. Study participants were immigrants diagnosed with major depression or dysthymia, and community members. As a new scale, CADS-9 will need to be researched with more samples to confirm its validity. The National Institute of Mental Health (R36MH080607), Chinese Community Health Care Association of San Francisco, and Fahs-Beck Fund for Research and Experimentation provided research grants for the development of CADS-9. Donaldina Cameron House was the principal community study site.

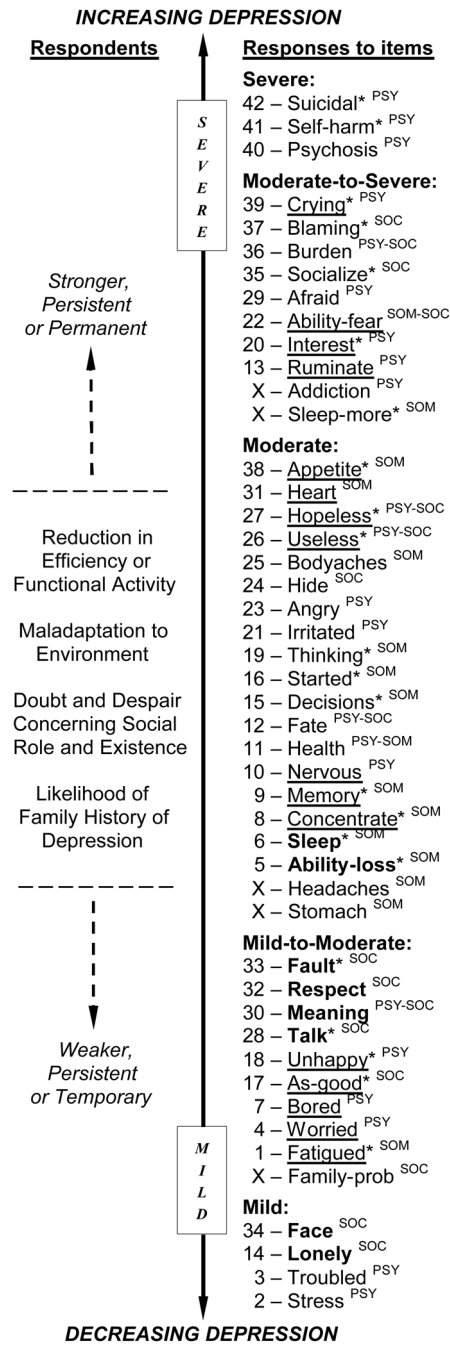


Figure 1. Construct map with theoretical description of respondents and 47 pilot items. Item numbers come from results of modeling (see Study 2, Table 2). Higher numbers indicate higher severity and ‘X’ indicates items that were removed due to poor fit. Underlined items were estimated as one level away from theorized levels. Bolded items were estimated as more than one level away from theorized levels. *Items that overlap with DSM-IV symptoms of major depression. PSY Psychological item, SOM Somatic item, SOC Social item.

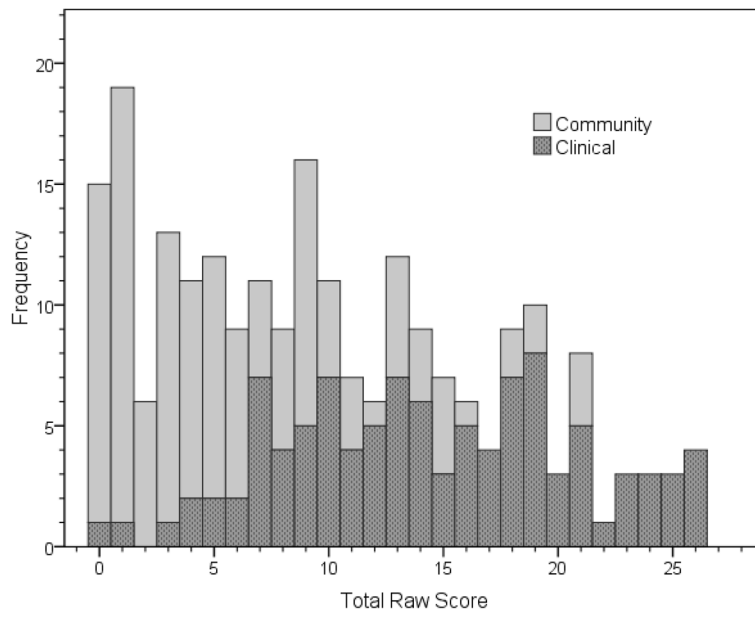


Figure 3. Distribution of CADS-9 scores by clinical and community samples.

Table 1

Socio-demographic, recruitment and clinical characteristics of participants

| | Total | Clinical | Community |
|---|-------------|-------------------|-------------------|
| Sample <i>n</i> (%) | 227 (100.0) | 103 (45.4) | 124 (54.6) |
| Men | 96 (42.3) | 38 (36.9) | 58 (46.8) |
| Women | 131 (57.7) | 65 (63.1) | 66 (53.2) |
| Age, mean (SD) | 44.6 (10.9) | 45.3 (10.6) | 44.0 (11.2) |
| Immigration, mean (SD) | | | |
| Age ^{**} | 32.7 (12.5) | 30.2 (13.0) | 34.8 (11.7) |
| Years in US ^{***} | 11.9 (9.7) | 15.1 (10.6) | 9.2 (8.0) |
| Personal income [*] % | | | |
| < US\$20000 | 69.8 | 80.6 | 61.3 |
| US\$20000–US\$39999 | 20.3 | 12.2 | 26.6 |
| US\$40000 | 9.9 | 7.2 | 12.1 |
| Education [*] % | | | |
| Elementary | 9.3 | 12.6 | 6.5 |
| Middle school | 17.6 | 20.4 | 15.3 |
| High school | 49.3 | 42.7 | 54.8 |
| College (≥ 2 years) | 23.3 | 23.3 | 23.4 |
| English (self-report) ^{a*} % | | | |
| Very Poor | 37.9 | 37.6 | 38.2 |
| Fair | 44.2 | 38.6 | 48.8 |
| Very good/Excellent | 17.8 | 23.7 | 13.0 |
| Very Poor | 37.9 | 37.6 | 38.2 |
| Recruitment sites % | | | |
| Mental health ^b | — | 53.4 | 14.5 ^c |
| Non-profit medical | — | 14.6 | 1.6 |
| Non-profit community | — | 8.7 | 75.0 |
| Private TCM ^d | — | 6.8 | 4.8 |
| Private psychotherapy | — | 1.9 | — |
| Other | — | 14.6 ^e | 4.0 |
| Scores on study instruments | | | |
| ACC-US-11 mean (SD) | 15.5 (5.3) | 15.5 (6.1) | 15.5 (4.7) |
| ASQ-14 ^{***} mean (SD) | 22.2 (9.3) | 25.7 (8.5) | 19.3 (9.0) |
| NTQ % diagnosed | 48.9 | 80.4 | 26.8 |
| PHQ-9 % diagnosed | 49.8 | 80.6 | 24.2 |
| Self-evaluated depression severity ^f % | | | |
| Never ^g | 34.8 | — | 64.2 |
| None | 9.0 | 7.9 | 10.0 |
| Mild | 36.7 | 51.5 | 24.2 |

| | Total | Clinical | Community |
|----------|-------|----------|-----------|
| Moderate | 14.5 | 29.7 | 1.7 |
| Severe | 5.0 | 10.9 | 0.0 |

Note.

^a Mean of questions on speaking, understanding and reading English.

^b Public and non-profit.

^c 11.3% were parents of children receiving care in a developmental center.

^d Traditional Chinese medicine.

^e 6.8% were community members who disclosed clinical status. ACC-US-11 = Acculturation to US Scale; ASQ-14 = Acculturative Stress Questionnaire; NTQ = Neurasthenia Questionnaire; PHQ-9 = Patient Health Questionnaire Depression Module.

^f Severity of symptoms at time of interview.

^g Never had depression symptoms.

* $p < .05$;

** $p < .01$;

*** $p < .001$.

Table 2

The Chinese American Depression Scale (CADS-42) item labels and texts, dimensional and cultural classifications, and item parameters obtained with a Partial Credit Model

| No. | Label | Dimension ^a | Culture ^b | English and Chinese Text ^c | Item Parameters | | |
|------------------------|-----------|------------------------|----------------------|---|-----------------|-------|--------|
| | | | | | Estimate | Infit | Outfit |
| Severe | | | | | | | |
| 42 | suicidal | Psy | W | You think about killing yourself. 你想過自殺。 | 1.985 | 1.16 | 1.04 |
| 41 | self-harm | Psy | W | You have thoughts about hurting yourself. 你想過傷害自己。 | 1.860 | 0.92 | 0.71 |
| 40 | psychosis | Psy | C | You have auditory or visual hallucinations. 你有幻聽和幻覺。 | 1.639 | 1.24 | 1.74 |
| 39 | crying | Psy | W | You cry. 你哭。 | 0.872 | 1.06 | 4.21 |
| Moderate-Severe | | | | | | | |
| 38 | appetite | Som | W | You have poor appetite. 你感到沒有胃口。 | 0.652 | 1.35 | 1.33 |
| 37 | blaming | Soc | W | You strongly blame your family members or partner for your life difficulties. 當你生活有困難,你強烈地埋怨家人或伴侶。 | 0.582 | 1.12 | 1.09 |
| 36 | burden | Psy-Soc | C | You feel you are a burden to your family and society. 你覺得自己是家人和社會的負擔。 | 0.516 | 1.10 | 0.88 |
| 35 | socialize | Soc | W | You don't want to have contact with people, socialize, or go out at all. 你完全不想和別人接觸,交往,或外出。 | 0.381 | 0.97 | 0.83 |
| 34 | face | Soc | C | You think you made your family lose face. 你感到讓親人丟臉。 | 0.333 | 1.00 | 0.88 |
| 33 | fault | Soc | W | You feel everything is your fault. 你感到所有問題都是自己的錯。 | 0.325 | 0.89 | 0.73 |
| 32 | respect | Soc | C | You feel you don't have the kind of respect from work and family that you should have. | 0.303 | 1.26 | 1.43 |

| No. | Label | Dimension ^a | Culture ^b | English and Chinese Text ^c | Item Parameters | | |
|-----------------|--------------|------------------------|----------------------|---|-----------------|-------|--------|
| | | | | | Estimate | Infit | Outfit |
| 31 | heart | Som | C | 你覺得在家中和工作的地方得不到應該有的尊重。 You have heart palpitations or chest discomfort. 你感到心跳加速或胸悶。 | 0.286 | 1.14 | 1.22 |
| 30 | meaning | Psy-Soc | C | You feel life is meaningless. 你感到做人沒有意思。 | 0.283 | 0.73 | 0.58 |
| 29 | afraid | Psy | C | You feel very afraid. 你感到非常害怕。 | 0.208 | 0.72 | 0.64 |
| 28 | talk | Soc | W | You don't want to talk. 你不想說話。 | 0.202 | 1.21 | 1.17 |
| 27 | hopeless | Psy-Soc | W | You feel hopeless. 你感到沒有希望。 | 0.184 | 0.72 | 0.58 |
| 26 | useless | Psy-Soc | W | You feel useless. 你覺得自己沒用。 | 0.051 | 0.82 | 0.74 |
| Moderate | | | | | | | |
| 25 | bodyaches | Som | C | You have bodily aches and pains. 你感到渾身疼痛。 | 0.032 | 1.25 | 1.20 |
| 24 | hide | Soc | C | You hide your life difficulties from other people. 你隱瞞自己生活有困難。 | 0.015 | 1.18 | 1.14 |
| 23 | angry | Psy | C | You have tantrums and get angry very easily. 你很容易發脾氣和發怒。 | -0.063 | 1.31 | 1.34 |
| 22 | ability-fear | Som-Soc | C | You are afraid of losing your working ability completely one day. 你害怕有一天會完全失去做事能力。 | -0.139 | 1.00 | 1.17 |
| 21 | irritated | Psy | C | You feel emotionally irritated. 你感到心煩氣燥。 | -0.169 | 0.78 | 0.76 |
| 20 | interest | Psy | W | You have no interest in many things. For example, you used to like to exercise, go shopping or go out for entertainment, but not anymore. 你對很多事情失去興趣。例如： | -0.208 | 1.09 | 0.99 |

| No. | Label | Dimension ^a | Culture ^b | English and Chinese Text ^c | Item Parameters | | |
|----------------------|-------------|------------------------|----------------------|--|-----------------|--------|--------|
| | | | | | Estimate | Infitt | Outfit |
| 19 | thinking | Som | W | 你從前很喜歡做運動, 逛街或出外娛樂, 但現在已經失去興。 你從前很喜歡做運動, 逛街或出外娛樂, 但現在已經失去興。 You feel that you are thinking very slowly. 你覺得自己思想很緩慢。 | -0.231 | 0.90 | 0.85 |
| 18 | unhappy | Psy | W | You feel very unhappy. 你感到非常不開心。 | -0.266 | 0.67 | 0.64 |
| 17 | as-good | Soc | W | You feel like no matter how hard you try, you can't be as good as other people. 你感到不論你多努力都比不上別人。 | -0.270 | 0.83 | 0.75 |
| 16 | started | Som | W | It feels very hard to get started on doing things. 你感到做事總是提不起勁。 | -0.272 | 0.88 | 0.80 |
| 15 | decisions | Som | W | You have trouble making decisions. 你常常拿不定主意。 | -0.285 | 1.15 | 1.15 |
| 14 | lonely | Soc | C | You feel very lonely. 你感到非常寂寞。 | -0.294 | 0.94 | 0.89 |
| 13 | ruminate | Psy | W | You think about unpleasant things the whole day and cannot stop. 你整天不停地想起不愉快的事情。 | -0.327 | 0.77 | 0.79 |
| 12 | fate | Psy-Soc | C | You feel very helpless. You feel you have no way to change your own fate. 你感到非常無助。你感到沒有辦法改變自己的命運。 | -0.365 | 0.80 | 0.67 |
| 11 | health | Psy-Som | C | You are very afraid of having health problems. For example, you might have cancer or heart disease. 你非常害怕自己健康有問題, 例如患癌症或心臟病。 | -0.373 | 1.16 | 1.43 |
| Mild-Moderate | | | | | | | |
| 10 | nervous | Psy | C | You are very nervous. 你感到非常緊張。 | -0.382 | 0.74 | 0.71 |
| 9 | memory | Som | W | Your ability to remember things has worsened very much. For example, you forget where you put your keys, or your appointments. 你記性比以前差很多, 例如忘記帶鑰匙或約會時間。 | -0.419 | 1.09 | 1.12 |
| 8 | concentrate | Som | W | You are unable to concentrate well. | -0.511 | 0.83 | 0.75 |

| No. | Label | Dimension ^a | Culture ^b | English and Chinese Text ^c | Item Parameters | | |
|-------------|--------------|------------------------|----------------------|--|-----------------|-------|--------|
| | | | | | Estimate | Infit | Outfit |
| Mild | | | | | | | |
| 7 | bored | Psy | C | You are very bored. 你感到很沉悶。 | -0.589 | 1.10 | 1.11 |
| 6 | sleep | Som | W | You are unable to sleep well. 你睡得不好。 | -0.598 | 1.31 | 1.32 |
| 5 | ability-loss | Som | W | You feel your working & learning ability has decreased significantly. 你感到做事和學習能力明顯下降。 | -0.705 | 1.03 | 1.01 |
| 4 | worried | Psy | C | Many things make you feel very worried. 很多事情讓你感到很擔心。 | -0.919 | 0.77 | 0.76 |
| 3 | troubled | Psy | C | Many things make you feel very troubled or bothered. 很多事情讓你覺得很煩惱。 | -0.940 | 0.81 | 0.81 |
| 2 | stress | Psy | C | You feel heavy stress living in the US. 你在美國生活感到重大的壓力。 | -0.969 | 1.31 | 1.32 |
| 1 | fatigued | Som | W | You feel very fatigued. 你感到很疲累。 | -1.415 | 1.08 | 1.12 |

Note. Psy = Psychological; Som = Somatic; Soc = Social. C = Chinese culture-specific; W = overlapping with a DSM-IV symptom of major depression.

Table 3

Analysis of Differential Item Functioning of the Chinese American Depression Scale (CADS-42) by Gender and Age: Ease of Endorsability and Effect Size

| No. | Item | Gender | Age |
|---|-----------------------------|--------------|--------------|
| Easier to endorse for males or younger adults | | | |
| 7 | bored ^{PSY} | 0.376 | 0.214 |
| 28 | talk ^{SOC} | 0.288 | |
| 32 | respect ^{SOC} | 0.252 | |
| 38 | appetite ^{SOM} | 0.252 | |
| 6 | sleep ^{SOM} | 0.242 | |
| 14 | lonely ^{SOC} | 0.220 | 0.216 |
| 1 | fatigued ^{SOM} | | 0.236 |
| 16 | started ^{SOM} | | 0.224 |
| Easier to endorse for females or older adults | | | |
| 39 | crying ^{PSY} | 0.724 | |
| 9 | memory ^{SOM} | 0.260 | |
| 30 | meaning ^{SOC/PSY} | 0.222 | |
| 40 | psychosis ^{PSY} | | 0.792 |
| 25 | bodyaches ^{SOM} | | 0.354 |
| 5 | ability-loss ^{SOM} | | 0.268 |
| 12 | fate ^{SOC/PSY} | | 0.224 |

Note. Effect sizes are in logits. Bolded text indicates a large effect size greater than .638 logits. PSY = Psychological; SOM = Somatic; SOC = Social.

Table 4

Differential Item Functioning of the Chinese American Depression Scale (CADS-47) by Education, English, and Years in U.S.: Ease of Endorsability and Effect Size

| No. | Item | Education | English | Years in U.S. |
|---|---------------------------------|--------------|--------------------|---------------|
| Easier to endorse for more acculturated | | | | |
| 32 | respect ^{SOC} | 0.420 | 0.412 | |
| 2 | stress ^{PSY} | 0.318 | | |
| 20 | interest ^{PSY} | 0.300 | 0.216 | |
| 31 | heart ^{SOM} | 0.238 | | |
| 38 | appetite ^{SOM} | 0.224 | | |
| 1 | fatigued ^{SOM} | 0.222 | | |
| 28 | talk ^{SOC} | 0.220 | 0.246 ^a | |
| 42 | suicidal ^{PSY} | | 0.616 | |
| 41 | self-harm ^{PSY} | | 0.452 | 0.270 |
| 40 | psychosis ^{PSY} | | | 0.526 |
| 34 | face ^{SOC} | | | 0.416 |
| 9 | memory ^{SOM} | | | 0.216 |
| Easier to endorse for less acculturated | | | | |
| 36 | burden ^{SOC/PSY} | 0.552 | | |
| 40 | psychosis ^{PSY} | 0.418 | | |
| 19 | thinking ^{SOM} | 0.322 | 0.318 | |
| 22 | ability-fear ^{SOM/SOC} | 0.254 | 0.250 | |
| 13 | ruminant ^{PSY} | 0.252 | | |
| 17 | as-good ^{SOC} | 0.242 | | |
| 25 | bodyaches ^{SOM} | 0.236 | | |
| 30 | meaning ^{SOC/PSY} | 0.232 | | |
| 5 | ability-loss ^{SOM} | | 0.570 | |
| 26 | useless ^{SOC/PSY} | | 0.288 | |
| 12 | fate ^{SOC/PSY} | | 0.230 | |
| 27 | hopeless ^{SOC/PSY} | | 0.222 | |
| 15 | decisions ^{SOM} | | | 0.418 |
| 7 | bored ^{PSY} | | | 0.366 |
| 2 | stress ^{PSY} | | | 0.312 |
| 14 | lonely ^{SOC} | | | 0.284 |
| 23 | angry ^{PSY} | | | 0.282 |

Note. Effect sizes are in logits. Bolded text indicates a large effect size that is greater than .638 logits. Bolded and italicized text indicates an intermediate effect size that is .426–.638 logits. PSY = Psychological; SOM = Somatic; SOC = Social.