Cultural relevance and equivalence in the NLAAS instrument: integrating etic and emic in the development of cross-cultural measures for a psychiatric epidemiology and services study of Latinos

MARGARITA ALEGRIA,[1] DORYLIZ VILA,[2] MEGHAN WOO,[1] GLORISA CANINO,[2] DAVID TAKEUCHI,[3] MILDRED VERA,[2] VIVIAN FEBO,[2] PETER GUARNACCIA,[4] SERGIO AGUILAR-GAXIOLA,[5] PATRICK SHROUT[6]

1 Center for Multicultural Mental Health Research, Cambridge Health Alliance/Harvard Medical School, Boston MA, USA
2 Medical Sciences Campus, University of Puerto Rico, San Juan PR, USA
3 Department of Social Work, University of Washington, Seattle WA, USA
4 Institute for Health, Health Care Policy and Aging Research, Rutgers University, New Brunswick NJ, USA
5 Department of Psychology, California State University, Fresno CA, USA
6 Department of Psychology, University of New York, New York NY, USA

Abstract

This paper describes the development, translation and adaptation of measures in the National Latino and Asian American Study (NLAAS). We summarize the techniques used to attain cultural relevance; semantic, content and technical equivalency; and internal consistency of the measures across languages and Latino sub-ethnic groups. We also discuss some of the difficulties and challenges encountered in doing this work. The following three main goals are addressed in this paper: (1) attaining cultural relevance by formulating the research problem with attention to the fundamental cultural and contextual differences of Latinos and Asians as compared to the mainstream population; (2) developing cultural equivalence in the standardized instruments to be used with these populations; and (3) assessing the generalizability of the measures – i.e., that the measures do not fluctuate according to culture or translation. We present details of the processes and steps used to achieve these three goals in developing measures for the Latino population. Additionally, the integration of both the etic and emic perspectives in the instrument adaptation model is presented.

Key words: National Latino and Asian American Study, NLAAS, Latinos, cultural relevance, cultural equivalence, instrument translation, ethnicity

Introduction

In 1990, the National Institutes of Health enacted a policy requiring the inclusion of minorities in study populations. However, specific guidelines have yet to be formulated on how to integrate the contextual and cultural life circumstances of minorities into the conceptualization of a research problem and the operationalization of constructs. To the contrary, the majority of studies of mental health issues to date continue to be based on assumptions of universality of the American mainstream experience. This assumption is justified on the basis of results from the standardized measures commonly employed in both psychiatric epidemiology and services research studies that themselves have not been developed to be culturally sensitive (Rogler, 1999). Over the last decade, many cross-cultural psychiatric studies used measures that were standardized on the mainstream population to assess correlates of psychiatric disorder and unmet need for mental health care without knowing for

certain whether the correlates of interest were relevant for, or that the standardized measures were applicable to, ethnic minority populations.

Research on ethnic minorities requires instrumentation that is sensitive to cultural and contextual variations (Canino and Bravo, 1994). Psychological research on minorities usually involves comparisons among different ethnic groups living in different locales and exposed to varied social, political and historical forces. These comparisons demand instruments capable of identifying similar psychological phenomena in dissimilar groups. A challenge to the researcher is to ensure that the assessment tools are equivalent across groups, that the questions capture the same constructs, and that the underlying explanations for the phenomena are included. Attaining cultural sensitivity in instrumentation requires translations and adaptations into languages other than English, as well as confirming that the complexity of language matches the literacy levels of the population of interest.

This paper describes challenges involved in the development, translation and adaptation of measures in the National Latino and Asian American Study (NLAAS). The process of development, translation and adaptation of instruments addressed three main goals:

- attaining *cultural relevance* by formulating the research problem with attention to the fundamental cultural and contextual differences of Latinos and Asians as compared to the mainstream population;
- developing *cultural equivalence* in the standardized instruments to be used with these populations;
- assessing the *generalizability* of the measures – that the measures do not fluctuate according to culture or translation.

We present details of the processes and steps used to achieve these three goals in developing measures for the Latino population. We summarize the techniques used to attain cultural relevance; semantic, content and technical equivalency; and internal consistency of the measure across languages and Latino subethnic groups. We also discuss some of the difficulties and challenges in doing this work. Another manuscript (forthcoming) focuses on the particular adaptations and translations of the measures for the Asian American population.

Background

The NLAAS, the National Co-morbidity Study – Replication (NCS-R) and the National Survey of American Life (NSAL) together comprise the national Collaborative Psychiatric Epidemiological Studies (CPES), designed to provide psychiatric epidemiological information on different populations. (See the guest editorial at the beginning of this issue for a description of the collaborative studies.) The NLAAS questionnaire consists of the NLAAS Core, the NLAAS Non-Core and the NLAAS Study Specific section. The NLAAS Core battery was designed to be comparable to the NCS-R and the NSAL, with identical measures of psychiatric illness, service use and impairment. The NLAAS Non-Core battery includes measures that correlate with one of the other two studies – some constructs are shared with NSAL and others with NCS-R. The Study Specific sections were those limited to just the NLAAS. In this paper we discuss changes, additions or deletions that were made to any part of the NLAAS instrument (Core, Non-Core or Study Specific sections) to provide an understanding of how to capture differences in Latino populations and develop comparable and culturally relevant instruments. To achieve this goal, we included in our model (as Alegria et al., 2004) alternative explanations of psychiatric disease, unmet need, and non-compliance with mental health services that take into account cultural and contextual features of Latino and Asian American groups.

A challenge to this process is that Latino and Asian culture is not homogeneous even within the 'Latino' or 'Asian' rubric, requiring an in-depth examination of how important concepts are understood across a wider range of Latino and Asian American cultures and in divergent contexts. Culture is defined as the product of group values, beliefs, norms, practices and experiences that pertain to a given ethnocultural group (Lopez and Guarnaccia, 2000). Although ethnic minorities share a common context with mainstream culture, each group has unique cultural characteristics that permeate their lives such as immigration history or loss of social ties with their home country. These characteristics are the product of the continued interaction of their culture of origin with the dominant or majority culture. Cultures and subcultures vary not only by national, regional, or ethnic background but also by age, gender, and social class, requiring that contextual differences

be also considered. Ethnic minorities also vary by whether the group includes mainly migrants or natives to the host country. For example, most Cubans included in our sample are non-US born and mainly live in the south.

Most investigators agree on the value of cross-ethnic and cross-cultural research findings and on the need to make research culturally sensitive (Canino, Lewis-Fernández and Bravo, 1997), as it will increase the scientific accuracy of the research rather than merely promote multicultural political correctness (Rogler, 1999; Beals et al., 2003). However, there is disagreement as to the extent of revisions that should be incorporated into research instruments in order to obtain cultural equivalency and cultural relevance. Cross-cultural studies can be approached from two different perspectives, which have been defined as the *emic-etic paradigm* (Brislin, Lonner and Thorndike, 1973). The *emic* perspective starts with concepts from within the culture and seeks to understand the meaning of that which is studied and its associations with other factors using that cultural framework. On the other hand, the *etic* perspective involves the evaluation of phenomena using more cultural neutral or 'objective' constructs.

Both perspectives have been critically evaluated in the literature (Canino et al., 1997; Beals et al., 2003). Cross-cultural research based on the emic approach suffers from the problem of observation bias and lack of generalizability. Although a thorough understanding of concepts relevant to one culture is obtained using the emic approach, these concepts are not necessarily comparable to those of other cultures. On the other hand, the etic approach emphasizes reliability by standardizing the measures at the expense of validity (measuring what is supposed to be measured). Validity may be compromised by imposing artifactual cross-cultural homogeneity due to a constricted conceptualization (omitting differences across groups) embedded in the instrumentation. This limitation has been called the 'category fallacy' (Kleinman and Good, 1985). Several investigators have devised strategies that attempt to integrate emic and etic perspectives into one overall research methodology that is both culturally valid and generalizable (see Canino et al., 1997, and Lopez and Guarnaccia, 2000 for examples from mental health research). Similarly, the instrument adaptation model presented in this paper integrates both the etic and emic perspectives. Its main purpose is to produce instruments that attain the equivalence of psychological phenomena across linguistically, culturally and contextually different populations, thus enabling comparisons inherent to the etic perspective. However, it aims to do so in a culturally sensitive way that makes possible the identification of culture specific characteristics within groups (the emic perspective).

The rationale for the adaptations, additions and deletions of questions and measures in the NLAAS instrument was guided by evidence that research strategies and assessments are more effective when they are compatible with the client's cultural patterns and experiences (Mezzich et al., 1999). Failure to use culturally sensitive research strategies may compromise the validity and generalizability of the results of a study. Several investigators (Guarnaccia and Rogler, 1999; Kirmayer and Young, 1999; Alarcon et al., 2002) have recommended using alternative frameworks for incorporating local categories to the posed questions. There is a need to balance comparability to other epidemiological studies – for example, the National Comorbidity Study – Replication (NCS-R) and the National Survey of American Life (NSAL) – with the need to have cultural and contextual specificity of underlying beliefs about the phenomena (for example, psychiatric symptoms) and its causes. For this reason, the development of the NLAAS instrument focused on how context, social position, lifestyle and culture influence the expression of illness, the disparities in mental health status, the help-seeking pathways, the treatments selected and received, the disparities in service access and quality of mental health treatments.

A major consideration in our selection of constructs and measures was the ability to compare the results directly with the NCS-R and the NSAL, two other national epidemiological studies that were conducted during the same time period as the NLAAS. The main outcome variables in the NLAAS are core measures in common with the NCS-R and NSAL: psychiatric status (World Mental Health Composite International Diagnostic Interview (WMH-CIDI) (Kessler and Üstün, under review)), functional impairment (World Health Organization Disablement Assessment Schedule (WHO-DAS) (Rehm, Üstün, Saxena, Nelson, Chatterji, Ivis and Adlaf, 1999; World Health Organization, 1998)), and service

utilization (WMH-CIDI) (Kessler and Üstün, under review). Measures addressed in the NLAAS, NCS-R, and NSAL also include sociodemographics and measures of social position related to wealth, average income, persistent poverty, and employment (refer to Pennell et al., 2004 for a detailed description of all sections included in the three studies). Measures of environmental context and psychosocial factors included only in the NLAAS and NSAL but not in the NCS-R are neighbourhood safety, social cohesion, migration status and history, and discrimination. Some measures exclusive to the NLAAS are problem recognition and a verbatim description of the mental health problem, barriers to the receipt of primary health services, familism, acculturation, acculturative stress, language proficiency, language barriers in service use, therapeutic alliance with the mental health service provider, intergenerational conflict, gender roles, loss of social ties, income and wealth in country of origin, use of non-prescription drugs for mental health problems, context of exit, relativity of US social position in contrast to social position in country of origin; unmet expectation in moving to US, experiences of incarceration, homelessness and use of public assistance; and ataques de nervios. We denote these measures as NLAAS Specific, selected mainly because of their relevance in understanding psychiatric illness and service use in Latino and Asian American populations.

The constructs included in the NLAAS were the result of an exhaustive selection process based on the study's conceptual model that was adapted from a framework proposed by McKinlay and Marceau (1999). After selecting the essential constructs, a process for identifying available measures was undertaken. The selected measures were then translated and adapted for Latino and Asian populations and tested through in-depth interviews and a formal pre-test.

Methods

The NLAAS specific instrument was finalized through a five step process:

1. identification and prioritization of the constructs represented in our conceptual model;
2. identification of available NLAAS measures;
3. adaptation and translation of the identified NLAAS measures;
4. pre-testing of the NLAAS measures; and
5. testing the internal consistency of the final NLAAS measures (see Figure 1).

These steps allowed the researchers to identify the constructs and measures relevant to the study population as well as ensure the content, semantic, and technical equivalence of the measures and their internal consistency. A detailed description of the process follows in this section.

Step 1: identification and prioritization of constructs

A team of experienced researchers from diverse professional and ethnic backgrounds composed the expert panel that was responsible for the operationalization of the constructs included in the NLAAS battery. The collaborative team included the authors of this paper as well as Javier Escobar, William Vega, Steve Lopez, Jeanne Miranda, Maritza Rubio-Stipec, Juan Ramos (NIMH representative), Pauline Agbayani-Siewert and Fong Gong. During the initial phase of the project, these researchers met to prioritize the questionnaire content and develop a preliminary list of essential constructs that correspond to the conceptual framework regarding potential mechanisms linked to psychiatric disorders and mental health service disparities (see Figure 2). These content areas were included in the NLAAS instrument (see Alegria et al. 2004 for a brief description of the model).

Step 2: identification of available measures for essential constructs

Three major approaches were used to identify and develop available measures for each construct in the conceptual model. First, investigators submitted measures that represented the predetermined content areas they had previously used in other studies. Second, an exhaustive literature search was conducted to identify additional measures of these content areas that had good psychometric properties. However, for some of the constructs, appropriate measures could not be identified, and investigators therefore developed new items or measures, which was the third approach employed. Once the compilation of measures was completed, the team of investigators and a panel of experts reviewed them using the following criteria: cultural relevance, psychometric properties of the measures, length of administration, required level of comprehension of the questions, and face validity of the measures.

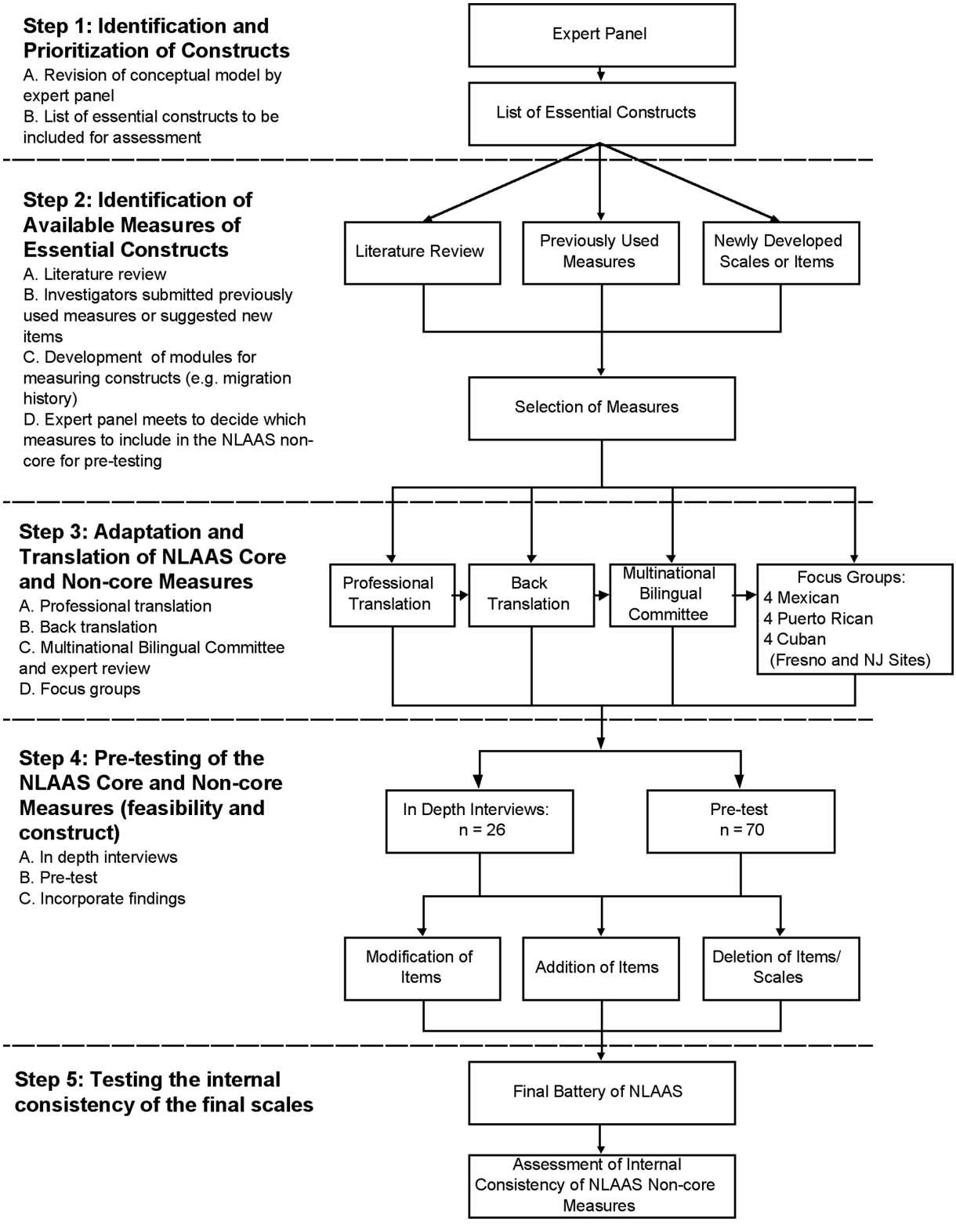

Figure 1. Five-step process to attain cultural relevance and cultural equivalence of the NLAAS measures.

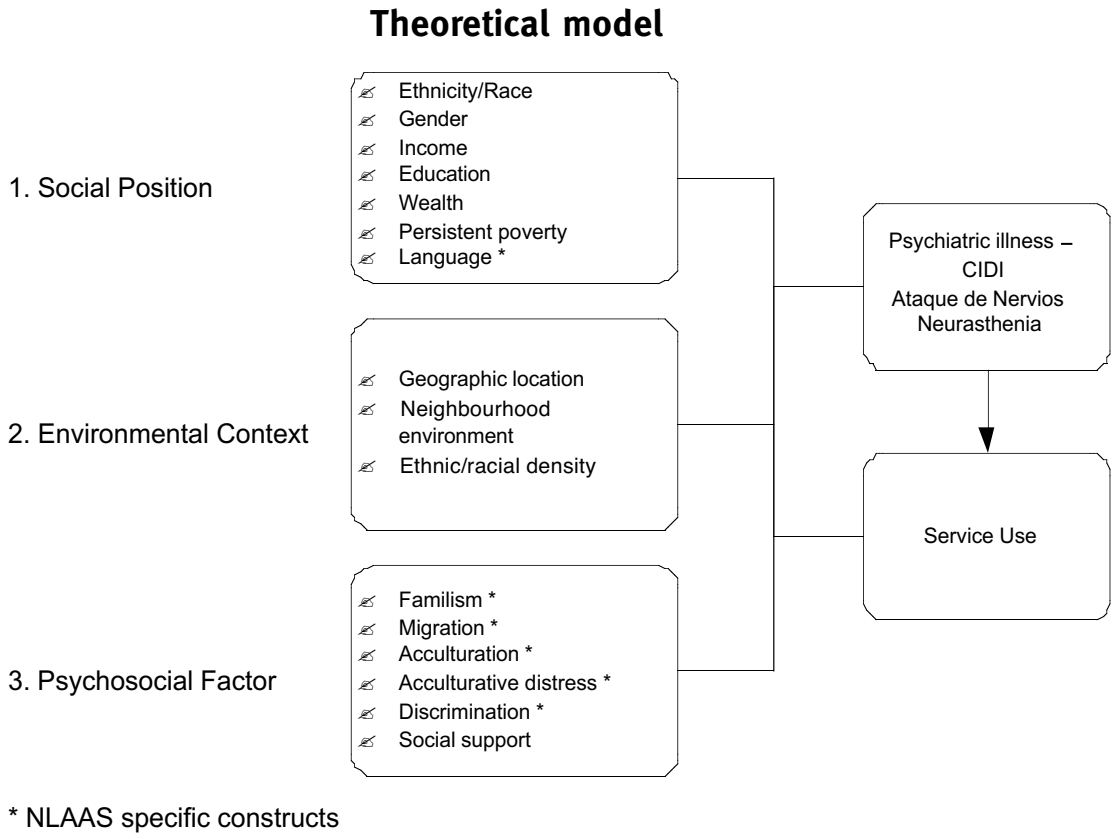

Figure 2. Constructs in the NLAAS conceptual model.

Step 3: translation and adaptation of the NLAAS measures

The translation and adaptation of the NLAAS measures followed the same steps described by Bravo, Canino, Rubio-Stipec and Woodbury (1991) in their study of methodological challenges in cross cultural mental health research. Semantic, content and technical equivalence were assessed during the process of translation and adaptation (Matias et al., 2000). All new content areas were sent for professional translation and back translation, and underwent a review by a multinational bilingual committee. This multinational bilingual committee evaluated whether the Spanish translation was culturally relevant for different Latino groups.

After finishing the professional translation, 12 focus groups were conducted with Spanish-speaking respondents (four each with Mexican, Cuban and Puerto Rican respondents) to ensure the instrument was adequately translated and adapted as well as culturally relevant for the targeted subgroups included in the study. After the focus groups were completed and analysed, investigators met to determine which suggestions or changes to incorporate into the NLAAS Specific instrument. Changes were made to the Spanish, and occasionally to the English, version of the questionnaire. This process of instrument translation and adaptation was completed during the first year of the study.

Step 4: pre-testing of the NLAAS measures (testing for feasibility and comprehension)

During the second year of the study, we examined the feasibility of using, and the comprehension of, questions and scale items for the NLAAS sections for different Latino subgroups. Using in-depth interviews with our three main Latino study groups: Mexicans, Puerto Ricans, Cubans, and other Latinos, we conducted 25 interviews in total: eight in Fresno County, California, six in Rutgers, New Jersey and eleven in Puerto Rico.

Study investigators discussed the findings and reached consensus before implementing modifications. Once changes were incorporated into the NLAAS measures, the revised measures were then programmed as a part of the computer assisted interview (CAI) of the NLAAS.

The initial NLAAS CAI was then pre-tested with 70 Latino respondents (50 English-speaking, and 20 Spanish-speaking). Six bilingual interviewers were trained to participate in three separate pre-tests of the NLAAS instrument. The pre-test assessed interview length, comprehension, consent, cultural issues, as well as specific questionnaire sections, the respondent's manual, and the pre-test manual. Measures that were either cognitively difficult, did not appear relevant, or did not adequately capture the construct of interest were modified through additions, changes or deletion of questions.

Step 5: testing the internal consistency of the final scales

Once the final NLAAS instrument was configured, testing of the internal consistency of the selected scales was conducted. Reliability analyses determined whether certain items lacked cultural equivalence across the Latino subgroups so that these items could be revised or removed to provide more comparable measurement. High reliability is a requirement in construct validation (Crocker and Algina, 1986). The internal consistency of the scales was calculated using the most common estimator of reliability, Cronbach's coefficient alpha (Cronbach, 1951) by Latino subgroup (Puerto Rican, Cuban, Mexican, and other Latino) and by language of interview (English and Spanish). Reliability was computed through the use of Cronbach's alphas for the following scales: family pride, family cohesion, everyday experiences of discrimination, perceived discrimination, family cultural conflict, acculturative distress, language proficiency (in English and Spanish), social cohesion, and neighbourhood safety.

Results

Step 1: identification and prioritization of constructs

The first step in developing the NLAAS battery involved the development and revision of appropriate constructs. Theoretical memos on topics such as acculturation and artifactual explanations were developed to assist investigators in prioritizing important constructs related to the conceptual model. A preliminary list of constructs was categorized into high and low (included mainly for comparative reasons or for the particular interest of one of the investigators) priority areas.

The essential content areas were: migration history, familism, acculturative distress, intergenerational conflict, exposure to toxic and violent environments, discrimination, and relativity of social position. Other constructs designated as having lower priority given time (Non-core and Study-specific battery was limited to 60 minutes) and budget constraints, or lack of applicability for Latino or Asian American groups were family burden, powerlessness, 'loss of face' and stressful life events.

Step 2: identification of available measures of essential constructs

Once the investigators identified the constructs necessary for inclusion in the instrument, appropriate scales and measures that encompassed these constructs were identified. A considerable number of scales were evaluated during this process, yet not many were included due to constraints on the length of the interview. For example, a subscale of the Cultural Identity Scales for Latino Adolescents (Félix-Ortiz, Newcomb and Meyers, 1994) was used to assess language preference and proficiency rather than the longer Cultural Life Style Inventory (Takeuchi et al., 1998). In other cases, instead of selecting the original scale, only a specific subscale was included (for example, nine items from the Hispanic Stress Inventory (Cervantes, Padilla and Salgado de Snyder, 1991) were chosen to create the Acculturative Distress Scale used in the NLAAS). For each construct, one or more available measure was selected. Table 1 provides a description of the final measures that were included as part of the NLAAS battery. These measures are the result of comprehensive testing and include additions and modifications to the original measures. Descriptions of these additions and modifications can be found in steps 3 and 4 of this section.

Step 3: adaptation and translation of measures

The process of adaptation and translation provided further guidance for modifying the selected measure, once the team of investigators completed the final list of selected measures. One of the most important aspects in translating the battery was using words that were understandable by different Latino subgroups under study.

Table 1. Description of final measures included as part of the NLAAS battery.

Measure/Scale	Description
Language proficiency in English and Spanish	Description: These measures assess the respondent's English or Spanish language proficiency by asking the respondent to rank his or her ability to speak, read and write in the specified language. The items are first administered in Spanish and are then repeated in English. The Spanish language proficiency scale originated from the Cultural Identity Scales for Latino Adolescents (Félix-Ortiz, Newcomb, and Meyers, 1994), while the English language proficiency scale was created specifically for the NLAAS using the parallel version of the Spanish language questions. The original 35-item instrument can be used to gauge a Latino adolescent's cultural identity through language, familiarity with culture and cultural values. The scales can be used as a group or subset of scales to assess particular aspects of Latino culture. The instrument may also be used to assess Latino cultural identity across many subgroups because no specific knowledge of any particular Latino culture is necessary. Reliability and validity of the instrument have been demonstrated (Félix-Ortiz, Newcomb, and Meyers, 1994). In the NLAAS, three items assess English proficiency and three items assess Spanish proficiency. Lower scores indicate a lower level of proficiency in either English or Spanish while higher scores indicate a higher level of proficiency. Representative items: 'How well do you speak (language)?' 'How well do you read (language)?' Mean: 8.71 (Spanish) 7.38 (English), Standard deviation: 2.64 (Spanish) 3.58 (English). Response categories: The 4 response categories range from poor (1) to excellent (4) with minimum and maximum scores of 3 and 12 respectively.
Language preference	Description: These items measure the language preference of the respondent. The items evaluate the respondent's level of preference for using Spanish or English when speaking with friends, family, and when thinking. The items were adapted from the Cultural Identity Scales for Latino Adolescents (Félix-Ortiz, Newcomb, and Meyers, 1994). The three items used in the NLAAS originated from the Spanish Language Preference scale. Higher scores indicate a greater preference for speaking English rather than Spanish. Representative items: 'What language do you speak with most of your friends?' 'In what language do you think?' Response categories: The response categories range from Spanish All the Time (1) to English All the Time (5) with minimum and maximum scores of 3 and 15 respectively.
Neighbourhood safety	The Neighbourhood Safety scale uses three items to measure the respondent's perceived level of neighbourhood safety and neighbourhood violence. Respondents are asked to rate various statements regarding their neighbourhood (e.g., whether people use drugs or if people are often mugged or attacked). The items were modified from the National Longitudinal Study of Adolescent Health or 'Add Health' (Bearman, Jones, and Udry, 1997) and the NIMH multi-site project, 'Cooperative Agreement for a Multi-Site Study of Mental Health Service Use, Need, Outcomes, and Costs in Child and Adolescent Populations' (NIMH, 1994). These items have also been adapted for use in the National Survey of American Life by James Jackson and colleagues (see article by Jackson et al. in this issue). Higher scores indicate a greater degree of neighbourhood safety than lower scores. Representative items: 'I feel safe being out alone in my neighbourhood during the night.' 'People often get mugged, robbed or attacked in my neighbourhood.' Mean: 9.39, standard deviation: 2.45 Response categories: The 4 response categories range on a Likert scale from very true (1) to not at all true (4). Minimum and maximum scores are 4 and 16 respectively.
Family pride	Description: Measurement of family pride is obtained through the sum of a seven-item subscale from the Family Environment Scale of Olson and colleagues (Olson, 1986; 1989). Respondents indicate their opinions on statements regarding their family that cover an array of about shared familial cultural values such as trust between family members, loyalty to the family, family pride, and a general orientation toward family. The scale is congruent with values of familism reported in Latino cultures as demonstrated by Vega (1990) and Ortiz (1995) and has been used in many studies with various Latino subgroups (Vega, Gil et al., 1993a; Gil, Vega, and Dimas, 1994; Gil and Vega, 1996). Higher scores represent lesser levels of family pride as compared to lower scores.

(contd)

Table 1. (contd)

Measure/Scale	Description
	Representative items: 'Family members respect one another.' 'We share similar values and beliefs as a family.' Mean: 9.68, standard deviation: 3.78 Response categories: The four response categories are rated on a Likert scale ranging from (1) strongly agree to (4) strongly disagree. The minimum and maximum scores are 7 and 28 respectively.
Family cohesion	Description: Measurement of family cohesion is obtained through the sum of a three-item subscale from the Family Cohesion scale of Olson and colleagues (Olson, 1986; 1989). Respondents indicate how strongly they agree with a variety of statements regarding their family. The items focus on elements of family closeness and communication such as whether the respondent considers family togetherness to be very important. The scale has been used in many studies with various Latino subgroups (Vega et al., 1993a; Gil et al., 1994; Gil and Vega, 1996). Higher scores represent lesser levels of family cohesion as compared to lower scores. Representative items: 'Family members like to spend free time with each other.' 'Family members feel very close to each other.' Mean: 4.09, standard deviation: 1.66 Response categories: The four response categories are on a Likert scale and range from strongly agree (1) to strongly disagree (4). The minimum and maximum scores are 7 and 28 respectively.
Family cultural conflict	Description: The Family Cultural Conflict scale addresses issues of cultural and intergenerational conflict between the respondents and their families. Respondents are asked to indicate frequency of incidents of cultural conflict with their families such as interference with personal goals, arguments with family members due to different belief systems, and the breakdown of family unity. The items were drawn from a subscale of the Hispanic Stress Inventory (HSI) (Cervantes et al., 1991). Cultural specificity of the HSI has been proven for those of Hispanic culture through extensive psychometric testing (Cervantes et al., 1991). In the NLAAS, five items from the Family/Culture Stress subscale of the HSI were selected to develop the Family Cultural Conflict scale. Higher scores represent greater levels of family cultural conflict as compared to lower scores. Representative Items include: 'Because of the lack of family unity, you have felt lonely and isolated.' 'Your personal goals have been in conflict with your family.' Mean: 6.34, standard deviation: 1.96 Response categories: The response categories are on a Likert scale and range from 'hardly ever' (1) to 'often' (3) with minimum and maximum scores of 5 and 15 respectively.
Migration status and history	Description: The migration status and history questions were adapted from the Mexican American Prevalence and Services Survey (MAPPS), a psychiatric epidemiology and services study of Mexican origin resident and migrant samples in Fresno County, California (Vega et al., 1998 and Alderete et al., 2000). Four questions were developed to assess the number of times the respondents returned to their own or their parents' country of origin, the length of stay during their last visit, and their country of primary residence. These items give insight to the respondent's ties to his or her country of origin. Representative items: 'How often have you returned to your country of origin?' and 'Approximately for how long did you stay in your country of origin last year?' Response categories: Response categories for this section are open ended.
Context of exit questions	Description: The context of exit questions were adapted from a community-based study of low income Mexican American women, ages 35–50, in San Diego County, California (Vega et al., 1986). Fifteen questions were selected from this study's original questionnaire to evaluate reasons for migration (voluntary or imposed), extent of preparation/anticipation preceding migration, and occupational and/or economic opportunities in the recipient nation. Representative items: 'Did you move from your country of origin because you or your family wanted to or because you had to?' 'If you had to make the decision today, would you still move to the United States?'

Table 1. (contd)

Measure/Scale	Description
	Response categories: Response categories in this section vary, but most responses range from 'very important' (1) to 'not important at all' (3), with a minimum and a maximum score of 9 and 27, respectively.
Acculturative distress	Description: The eight items of the Acculturative Distress scale were taken from the Mexican American Prevalence and Services Survey (MAPSS) (Vega et al., 1998). This scale measures the stress of culture change that results from immigrating to the United States (Cervantes, Padilla et al., 1990; Cervantes, Padilla and Salgado de Snyder, 1991) and has been tested mostly with Mexican American samples. A series of questions focus on respondent's experiences upon immigration: discrimination, fear of deportation, limited contact with family and friends, and difficulties with the English language. The MAPSS originally adapted these items from the Occupational/Emotional Stress subscale of the Hispanic Stress Inventory (HSI) (Cervantes, Padilla et al., 1991). The HSI was developed to measure the experiences of psychosocial stress for Hispanics. Cultural specificity of the HSI has been proven for those of Hispanic culture through extensive psychometric testing (Cervantes et al., 1991). Representative items: 'Do you feel guilty for leaving family or friends in your country of origin?' 'Do you avoid seeking health services due to fear of immigration officials?' Mean: 1.99, standard deviation: 1.81 Response categories: This scale has dichotomous response categories of 'Yes' (1) or 'No' (5).
Everyday discrimination	Description: This discrimination scale measures the frequency of routine experiences of unfair treatment. Respondents indicate how often they experience situations such as being treated with less respect than other people, having people act afraid of them, and being called names or insulted. These nine items were taken from the Detroit Area Study (DAS) (Williams, Yu et al., 1997). The scale has been used extensively in the mental health field (see Schulz, Israel et al., 2000; Mays and Cochran, 2001; and Boardman, Finch et al., 2001). Higher scores represent fewer incidences of everyday discrimination as compared to lower scores. Representative items: 'You are treated with less courtesy than other people.' 'People act as if they think you are not smart.' Mean: 47.41, standard deviation: 7.28 Response categories: The six response categories range from 'almost everyday' (1) to 'never' (6). The minimum and maximum scores for the scale are 9 and 54, respectively.
Perceived discrimination	Description: The perceived discrimination scale was originally developed as a measure of acculturative strain for use in a longitudinal study of adolescent drug use by Vega and colleagues (1993b). The items ask the respondents how often they or their friends are disliked or treated unfairly because of their race/ethnicity. Higher scores indicate fewer incidences of perceived discrimination as compared to lower scores. Representative Items: 'How often do people dislike you because you are [ethnic/racial group of respondent]?' 'How often do people treat you unfairly because you are [ethnic/racial group of respondent]?' Mean: 9.68, standard deviation: 2.31 Response categories: The scale consists of three items with response options ranging from 'often' (1) to 'never' (4), with a minimum and a maximum score of 3 and 12, respectively.
Social Cohesion	Description: The Social Cohesion scale is a measure of social cohesion and trust assessed by four questions obtained from three different instruments. The scale evaluates the cohesiveness of the respondent's neighbourhood by enquiring if neighbours get along with each other and if they can count on each other in emergency situations. The items are adapted from a scale used by Sampson, Raudenbush, and Earls (1997), the UNOCCAP questionnaire (NIMH, 1994), and a questionnaire used in the National Longitudinal Study of Adolescent Health, entitled 'Add Health' (Bearman, Jones, and Udry, 1997). Scale items: Two of the items ('People in this neighbourhood can be trusted' and 'People in this neighbourhood generally get along with each other') were adapted from the

(contd)

Table 1. (contd)

Measure/Scale	Description
	Social Cohesion and Trust subscale by Sampson, Raudenbush, and Earls (1997), one item ('I have neighbors who would help me if I had an emergency') was adapted from UNOCCAP, and the final item ('People in my neighbourhood look out for each other.') was adapted from the Neighbourhood subscale of Add Health. Higher scores indicate a lesser degree of social cohesion than lower scores. Mean: 7.83, standard deviation: 2.89 Response categories: The 4 response categories range from 'very true' (1) to 'not at all true' (4). Minimum and maximum scores are 4 and 16 respectively.
Ataque de nervios	Description: The 'ataque de nervios' items were specifically developed for use in the NLAAS. This is the first time this idiom of distress, prominent in many Latino groups, has been studied among Latinos outside Puerto Rico. Using the extensive research of Guarnaccia and colleagues (1993, 1996), 18 questions were developed to assess this cultural syndrome. A screening question, 'Have you ever had an episode or nervous attack where you felt totally out of control?' was first added to the WMH-CIDI in the screening section of the instrument. If respondents answered positively to the ataque screener, they were then asked if they had experienced a range of fifteen different symptoms common to an ataque de nervios episode. Examples are, becoming hysterical, having a period of amnesia, and falling to the floor with a 'seizure'. Representative items: 'During that episode did you shout a lot?' 'During that episode did you have crying attacks?' Response categories: Respondents answered with a dichotomous response category of yes (1) or no (5) to each item.

For more information, please see http://www. multiculturalmentalhealth.org

The Spanish language varies greatly depending on context and in some instances a word has different meanings depending upon the country of origin of the respondent. For example, in some cases it was impossible to determine a unique word to describe an event or object that was understood by different Spanish speakers. To deal with this issue, a list of 'non-consensus' words was developed (see Table 2 for examples). This list included the appropriate word for each Latino subgroup. For example, the word 'pool' is translated as 'alberca' for Mexicans, as 'piscina' for Puerto Ricans and Other Latinos, and Cubans would understand either of these translations. The interviewer used the list as a tool in administering the interview according to the Latino subgroup to which the respondent belonged.

Other changes incorporated in the NLAAS battery were related to the format and applicability of the measures for certain Latino subgroups. Focus groups revealed a great deal of information regarding these aspects of the battery. For example, when respondents were asked to mention stressful situations that arise from being a member of an ethnic minority in the US, Puerto Ricans found the questions associated with deportation or illegal status to be irrelevant due to the fact that they are US citizens. For this reason, the Acculturative Distress Scale was modified to include a 'does not apply' response category. In the employment section, several employment categories mentioned by ethnic minority respondents often did not fit those listed (for example, parking attendant, manicurist), and

Table 2. Use of Spanish vocabulary in the NLAAS questionnaire

Words	Mexicans	Puerto Ricans	Cubans	Other Latinos
Piscina / alberca (pool)	Alberca	Piscina	Either	Piscina
Armarios / 'closets' (closets)	Armarios	Closets	Either	Closets
Una fila / hacer cola (stand in line)	Una fila	Una fila	Hacer cola	Una fila
Lidiar / manejar (manage)	Manejar	Manejar	Lidiar	Manejar

therefore the alternative category, 'Other', and instructions to specify what this 'Other' category represented were included to allow for respondent's occupation not listed. A similar situation was identified in the migration status and history questions where the module was modified to include respondents whose parents made the decision to leave their country of origin rather than the respondents themselves. Furthermore, it came to our attention that some immigrants spent as much time in their country of origin as they did in the US, and therefore questions regarding primary residence were included in the appropriate module.

Focus groups examined measurement constructs and provided in-depth information about difficulties in understanding the questions due to translation. For example, one services question originally asked whether the respondent had used 'un trabajador social' (social worker) for mental health problems; however, participants in the focus group pointed out that 'un trabajador social' was generally understood to mean someone who helps with social problems such as finding a job, not someone who provides mental health services. The focus groups also helped determine whether constructs accurately measured the concepts they were intended to assess. For example, one question originally asked: 'cree usted que tiene el *estatus* / *respeto* que tenia en su país de origen?' ('do you believe that you have the same status/respect that you had in your country of origin?'). The focus group participants indicated that the two words capture somewhat different concepts and that these may best be asked using separate questions if the intent was to capture one concept or the other without confusion. Specifically, 'estatus' represents social status in terms of money; whereas 'respeto' represents respect.

Focus group findings also served to identify difficulties in understanding the translation of the measures and the purpose of some questions. For example, Latino respondents had a difficult time identifying themselves as belonging to one of the racial categories provided by the Census. When asked to specify their race, many said 'other' or gave their country of origin instead of choosing one of the listed categories. A second question on race was therefore developed to include additional categories previously suggested in the literature (Amaro and Zambrana, 2000) such as 'Mestizo', 'Criollo', 'Mulato', 'Latinegro', 'Indigena', and 'Caribbean'. This additional question allows us to examine how the race categorization changes when other alternative categories are provided. Another example was the translation of the term 'to pray' was expanded to include additional words (for example, 'rezar' and 'orar') as the word 'pray' differs according to religious affiliation. Specifically, Catholics use 'rezar' whereas Protestants use 'orar'.

Step 4: pre-testing of the NLAAS measures

Additional changes were also incorporated into the NLAAS battery as a result of in-depth interviews and pre-testing. Interviewer candidates were first screened over the telephone and subsequently interviewed in person before being offered employment. Those whose native language was not English were also required to successfully pass a language assessment evaluation. Aspects related to interview length, comprehension, consent, cultural issues, specific questionnaire sections, respondent's manual, and pretest manual were assessed. Conference calls were scheduled between pretest interviewers, Institute for Social Research staff (the organization responsible for data collection) and study investigators. Findings about the interview length revealed that the pre-test version of the questionnaire needed to be reduced from nearly 3 hours to approximately 2 hours and 30 minutes. This required that several sections be eliminated or shortened. Additional modifications were made to the questionnaire to address issues raised by the pretest, such as eliminating questions that had double negatives. Other sections required simplifying some sentences or questions so that respondents with low literacy levels could understand them.

We found that elderly Latino people tended to have difficulties understanding numeric scales and that a graphic representation of a range of responses was easier for them to understand; therefore, some items were modified to substitute Likert scales with a graphical format. A similar approach was developed for explaining percentages. Rather than asking people for percentages, we provided alternative percentages and an example of what the percentage means (for example, 25% means that one in every four persons would get better with mental health services). In addition, the use of a Respondent Booklet (a tool for facilitating the interview process) to answer scale questions was helpful to the respondents. All scale item response categories were included in the Respondent Booklet, so that the respondent could refer to them when answering questions.

The pre-tests also showed that some items required clarification of definitions. In the familism scales, there was some confusion about the definition of family. Respondents did not know if the questions referred to their immediate family, their extended family, their family in the US, or their family in their country of origin. To clarify this confusion, the team of investigators decided to define family as 'whatever it meant to the respondent'. A similar situation occurred in the finances section where respondents were asked to indicate how much money they would have remaining after paying their debts and selling all of their assets. Many respondents inquired if the question referred to the assets they possessed in the US or in their country of origin and therefore a separate question was included about having money left after selling all assets in their country of origin.

The NLAAS battery was also tested and modified according to pre-test findings. In the services section, for example, questions were added to assess therapeutic alliance, availability of a Spanish-speaking mental health provider, and additional barriers to care. Table 3 includes examples of these items. Due to time constraints, only one item was included to assess therapeutic alliance (aspects of the therapeutic relation that make it possible for providers and patients to work together to accomplish therapeutic goals). Four items were included to measure linguistic barriers to care and barriers related to the respondents' race/ethnicity. One additional item was included to explore the availability of health providers that spoke the respondent's language. Similarly, ten items on barriers to primary care services were included as part of the NLAAS Non-Core.

In the pharmacoepidemiology section, which asks about the use of prescription and non-prescription medications in the past 12 months for 'problems with emotions, nerves, mental health, substance use, energy, concentration, sleep, or ability to cope with

Table 3. Questions added to sections of the NLAAS Instrument

Construct	Example of questions added to the NLAAS battery
Therapeutic alliance Barriers to care	• Did this professional accept you and make you feel understood? • You felt treated with less attention because of your race or ethnic background. • You could not communicate with your therapist. • You were concerned that you could be treated unfairly because of your race or ethnic back ground. • You thought you would not be able to communicate because of linguistic barriers.
Availability of health provider that speaks respondent's language	• When you went to see a professional about your emotions in the past year, were you able to communicate in your language of preference – nearly all the time, most of the time, sometimes, only rarely, or never?
Problem recognition	• In the past year, was there any time when you thought you had a nervous, emotional, drug or alcohol problem?
Gender roles	• When it comes to household chores – like cleaning, cooking, childcare, grocery shopping – who has the most responsibility - only you, mostly you and sometimes your spouse/partner, you and your spouse/partner about the same, mostly your spouse/partner and sometimes you, or only your spouse/partner? • When it comes to paying for household expenses – like mortgage or rent, electricity bills – who has the most responsibility - only you, mostly you and sometimes your spouse/partner, you and your spouse/partner about the same, mostly your spouse/partner and sometimes you, or only your spouse/partner?

Barriers to primary care	• In the last year, have you ever experienced any of the following situations with your primary doctor?
	  • Difficulty getting an appointment over the phone.
	  • Difficulty getting referrals to a specialist.
	  • The provider spends limited time with you.
	  • Long waits in waiting room (more than 1 hour).
	  • Difficulty getting information or advice by phone

stress', we found that some respondents in the pre-test interview did not know whether the medication they were taking was for a mental health condition. We therefore added a question on whether the respondent was taking medications (for example, Prozac or Zoloft), independent of whether they identified that they were taking the medication for a mental health disorder. An exhaustive list of prescription medications was given to the respondent as a visual aid in answering this question. In the case of gender roles, the in-depth interviews indicated that the gender role questions were evaluated as more reflective of attitudes than of the actual behaviour patterns which the constructs were intended to measure. Therefore, more specific questions were substituted for the gender role probes; questions that captured gender role behaviours such as identifying who manages the family finances or who does household chores.

The personality section was found to be cognitively difficult for Latino respondents due to items with double negatives and the finding that antisocial behaviours appeared to vary across cultures (Alarcón et al, 2002). 'I never met a person that I didn't like' is an example of one of the items that was very difficult for the Latino population to answer. The entire module was eliminated and limited only to 10 questions assessing the respondent's level of social desirability.

Other sections required simplifying sentences or modifying questions so that respondents with low literacy levels could understand them. In the finances section, for example, an original question read:

> Suppose you (and your spouse or partner) cashed in all your checking and savings accounts, stocks and bonds, real estate, sold your home, your vehicles, and all of your valuable possessions. Then suppose you put that money toward paying off your entire mortgage and all your other loans, debts, and credit cards. Would you have any money left over after paying your debts or would you still owe money?

This question had to be simplified because respondents could not understand it. The modified item was shorter and seemed to work better for the Latino and Asian populations. The modified item read:

> Suppose you (and your spouse or partner) sell everything you own (cars, houses, land, etc.) and use the money to pay all your debts (loans, credit cards, mortgage, etc.). Would you still have any money left over after paying your debts?

Step 5: Testing the internal consistency of the final scales

The results of the test to evaluate the internal consistency of the NLAAS study specific scales are presented by Latino subgroup in Tables 4 (Puerto Rican, Cuban, Mexican, and Other Latino) and by language of interview in Table 5 (English and Spanish). The Cronbach coefficients indicate high internal consistency of most of the NLAAS study specific scales by Latino subgroups as well as by language of interview. The only two exceptions were the Acculturative Distress scale for Puerto Ricans, Cubans and Other Latinos (Cronbach's alpha = 0.59, 0.65, 0.61, respectively) and the Neighbourhood Safety scale for Cubans (Cronbach's alpha = 0.61).

As demonstrated in Table 5, Cronbach's alphas calculated for the Latino subgroups show a high degree of internal consistency with very similar alpha coefficients for the different Latino subsamples. Almost all scales evaluated for internal consistency achieved a good to excellent level of internal consistency ranging from 0.70 to 0.98. Regarding specific subgroups, the Mexican subsample had the highest degree of internal consistency with scores ranging from 0.70 to 0.98, while the Puerto Rican subgroup had the lowest with scores ranging from 0.59 to 0.96. The English language proficiency scale reports the highest internal consistency across the subgroups (0.96 to 0.98). The acculturative distress scale reports the lowest internal consistency (0.59 to 0.76). However, the majority of coefficients are above 0.80, indicating that the questions included under the scales do appear to have cross-cultural comparability.

Table 5 presents Cronbach's alphas for the scales by language of interview, English or Spanish. Similar to the internal consistency results observed for the Latino subgroups, a high degree of internal consistency for both languages was obtained. Cronbach's alphas for scales administered in English ranged from 0.61 to 0.91 while scores for scales administered in Spanish range from 0.80 to 0.96. The coefficients obtained for the individual scales are nearly identical to each other. In particular, the language proficiency scale has the highest reliability with a score of 0.90 in English and 0.96 in Spanish. Once again, the one exception is the acculturative distress measure that shows Cronbach's alpha coefficients of 0.62 in English and 0.71 in Spanish. The computation of alpha coefficients indicated good to excellent reliability and comparable levels of inter-item consistency in both languages.

Table 4. Internal consistency (Cronbach's alpha) of NLAAS scales by Latino subgroup

Scale	Total Latino Sample N = 2,554	Puerto Rican N = 495	Cuban N = 577	Mexican N = 868	Other Latino N = 614
Family pride	0.916	0.929	0.900	0.902	0.918
Family cohesion	0.826	0.845	0.784	0.833	0.813
Discrimination: everyday experiences	0.789	0.775	0.827	0.785	0.775
Family culture conflict	0.908	0.895	0.895	0.912	0.906
Acculturative distress	0.703	0.593	0.652	0.757	0.608
Language proficiency – Spanish	0.902	0.912	0.901	0.885	0.904
Language proficiency – English	0.975	0.962	0.979	0.974	0.975
Social cohesion	0.809	0.807	0.819	0.801	0.799
Neighbourhood safety	0.720	0.752	0.675	0.701	0.725
Discrimination: perceived discrimination	0.820	0.837	0.832	0.807	0.810

Table 5. Internal consistency (Cronbach's alpha) of NLAAS scales by language of interview (English and Spanish)

Scale	Language of interview	
	English N = 1,056	Spanish N = 1,498
Family pride	0.909	0.917
Family cohesion	0.817	0.826
Discrimination: everyday experiences	0.780	0.795
Family culture conflict	0.895	0.906
Acculturative distress	0.616	0.705
Language proficiency – Spanish	0.906	0.884
Language proficiency – English	0.901	0.963
Social cohesion	0.806	0.813
Neighbourhood safety	0.759	0.706
Discrimination: perceived discrimination	0.816	0.824

The internal consistencies of the NLAAS scales, as reported by Cronbach's alphas, demonstrate reliabilities that are comparable to or better than those of the original scales. The Everyday Experiences of Discrimination scale, where response categories were modified for inclusion in the NLAAS, exemplifies this finding with Cronbach's alphas of 0.78 for the English language interview and 0.80 for the Spanish language interview. These scores are comparable to the findings of Williams et al. (1997) in which Cronbach's alphas of 0.88 were reported. The Family Cultural Conflict scale is another example: although the subscale taken from the Hispanic Stress Inventory underwent modifications to the response categories and wording of items, the internal consistency remained the same. For the NLAAS, the five-item scale was found to have standardized Cronbach's alphas of 0.90 for the English language version and 0.91 for the Spanish language version, comparable to other reports of acceptable levels of reliability with coefficients ranging from 0.61 to 0.86 (Cervantes et al., 1990). The final example is the Language Proficiency scale where the measure of English language proficiency was created specifically for the NLAAS based on the Spanish language proficiency measure. Standardized Cronbach's alphas of 0.90 for the English language interview and 0.96 for the Spanish language interview were reported. These

scores are comparable to the findings of the original Spanish language proficiency scale by Félix-Ortiz, Newcomb and Meyers (1994) in which a Cronbach's alpha of 0.88 was reported.

Discussion

When standardized instruments from a mainstream culture are used with divergent cultural and ethnic groups, there is always a risk that the instruments may lack cultural relevance and equivalence, leading to measurement errors and misinterpretation of findings. For example, we discovered if an additional question inquiring how much money the respondent sent home to his or her country of origin was not added, household income was overestimated by 15%. Our results emphasize that it is critical to use both qualitative and quantitative methods to examine questions and explore response patterns in order to generate insight into methods for obtaining cultural equivalence and relevance across Latino groups and languages.

Our qualitative findings show that cross-cultural conceptual validity of measures can occur mostly when indigenous categories of experience are incorporated in the assessment. Particular attention was devoted to the responses across Latino groups in order to evaluate whether they were equally reliable and had semantic, content and conceptual equivalence to the construct that we intended to measure. Research indicates that when cross cultural assessments have no prior testing with different cultural, ethnic, racial, linguistic or economic groups, the questions and response categories to these questions may be processed differently (Barofsky, 2000).

One of the greatest challenges the NLAAS investigators faced was striking a balance between the emic perspective (seeking equivalence within the culture) and the etic perspective (maintaining comparability). In collaborative studies of this nature, comparability with the other ethnic or racial groups (for example, the NSAL) and particularly with the mainstream group (for example, the NCS-R) is essential. It is in comparing mental illness and service utilization rates of the minority group with the mainstream group that mental health and service disparities can be documented. But in order to maintain this comparability the same core constructs and assessment instruments need to be used. The NLAAS investigators made every effort to maintain comparability to the original versions of instruments to ensure a valid comparison to the other studies of the general US populations and African-Americans. However, several measures needed to be modified or expanded in both Spanish and English because of the inequality in the meaning of phenomena for the different Latino and Asian respondents.

Incorporating the emic perspective was a much more difficult task. New constructs, instruments or items needed to be added to the protocol. This constituted a problem particularly when the core protocol was long, because in order to avoid subject fatigue and low response rates, as well as keeping the study costs down, the time allotted for the new instruments needed to be limited. The investigators were then faced with the task of striking a balance between the number of cultural constructs that needed to be added and the length of the instruments used for measuring each construct. This problem is compounded when studying different subethnic groups of Latinos and Asians that differ in rates of psychiatric disorder and service use rates, as well as the risks associated with these rates. We opted for adding more culturally relevant constructs to the protocol but consequently had to shorten most of the scales that measured these constructs.

Striking a balance between the etic and emic perspective in collaborative studies is always the greatest challenge faced by investigators. The task is made easier if the group making decisions about the core protocol in a collaborative study is composed of diverse researchers. This diverse group should include minority researchers who can advocate for an equitable balance between time allotted for the main core protocol and the non-core protocol of the different ethnic groups represented in the study.

Although modifications and additions were made to some of the measures in the NLAAS battery, the investigators made every effort to maintain comparability to the original versions to ensure a valid comparison to the other studies of the general US populations and African-Americans. However, several measures needed to be modified or expanded in both Spanish and English because of the inequality in the meaning of the given phenomena for the different Latino respondents.

Testing the accuracy of mainstream assumptions that standardized instruments would work equally well,

independent of cultural or linguistic background, proved to be important in the development of the NLAAS battery. This testing showed the need to include new questions, rephrase items or add new responses to guarantee appropriate language and content choices for respondents. The multiple measures required several evaluations to assess if the words had similar meanings when translated and adapted. We also confronted challenges in that some words in English and in Spanish exist in one language but cannot be translated into the other language, because meaning is experiential and not replicable or exportable across cultures. (for example, words such as 'ataque' or 'nervios' do not have identical words in English).

The reliance on measures developed in our cultural context that do not take semantics and cultural variation into account fosters false assumptions, attributional errors, and misleading interpretations, mostly due to the absence of a solid understanding of how to incorporate cross-cultural variation in measurement. Using local cultural knowledge to facilitate new ways to interpret the measures and to make them relevant and conceptually equivalent proved essential. It is only by asking those from within the culture to help facilitate instrument development that we can do so – those within the culture understand traditions, challenges, stresses, and circumstances that allow them to be experts in measurement translation, adaptation and development.

References

Alarcón DR, Bell C, Kirmayer JL, Lin K-M, Üstün B, Wisner LK. Beyond the funhouse mirrors. In JD Kupfer, BM First, AD Regier (eds) A Research Agenda for DSM-V. Washington DC: American Psychiatric Association, 2002.

Alderete E, Vega WA, Kolody B, Aguilar-Gaxiola S. Lifetime prevalence of and risk factors for DSM-III-R disorders among Mexican migrant workers. American Journal of Public Health 2000; 90: 608–14.

Alegria M, Takeuchi D, Canino G, Duan N, Shrout P, Meng X, Vega W, Zane N, Vila D, Woo M, Vera M, Guarnaccia P, Aguilar-Gaxiola S, Sue S, Escobar J, Lin K, Gong F. Considering context, space and culture: The National Latino and Asian American Study. IJMPR 2004; 13(4): 208–20

Amaro H, Zambrana RE. Criollo, Mestizo, Mulato, LatiNegro, Indígena, White, or Black? The US Hispanic/Latino population and multiple responses in the 2000 census. American Journal of Public Health 2000; 90: 1724–7.

Barofsky I. The role of cognitive equivalence in studies of health-related quality-of-life assessments. Medical Care 2000; 38(9): suppl. II125–II129.

Beals J, Manson SM, Mitchell, CM; Spicer, P AI-SUPERPFP Team (2003). Cultural specificity and comparison in psychiatric epidemiology: walking the tightrope in American Indian research. Culture, Medicine, and Psychiatry 27(3): 259–89.

Bearman J, Jones J, Udry R. The National Longitudinal Study of Adolescent Health. Chapel Hill NC: Carolina Population Center, 1997.

Boardman JD, Finch BK, Ellison CG, Williams DR, Jackson JS. Neighbourhood Disadvantage, Stress, and Drug Use Among Adults. Journal of Health and Social Behavior 2001; 42: 151–65.

Bravo M, Canino G, Rubio-Stipec M, Woodbury M. A cross-cultural adaptation of a diagnostic instrument: the DIS adaptation in Puerto Rico. Culture, Medicine, and Psychiatry 1991; 15: 1–18.

Brislin R, Looner W, Thorndike R. Cross-cultural Research Methods. New York: Wiley, 1973.

Canino G, Bravo M. The adaptation and testing of diagnostic and outcome measures for cross-cultural research. International Review of Psychiatry 1994; 6: 281–6.

Canino G, Lewis-Fernández R, Bravo M. Methodological challenges in cross-cultural mental health research. Transcultural Psychiatric Research Review 1997; 34(2): 163–84.

Cervantes RC, Padilla AM, Amado M, Salgado de Snyder N. Reliability and validity of the Hispanic Stress Inventory. Hispanic Journal of Behavioral Sciences 1990; 12: 76–82.

Cervantes RC, Padilla AM, Salgado de Snyder WN. The Hispanic Stress Inventory: a culturally relevant approach to psychological assessment. Psychological Assessment 1991; 3: 438–47.

Crocker L, Algina J. Introduction to Classical and Modern Test Theory. New York, NY: Holt, Reinhart & Winston, 1986.

Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951: 16; 297–334.

Felix-Ortiz M, Newcomb MD, Myers H. A multidimensional measure of cultural identity for Latino and Latina adolescents. Hispanic Journal of Behavioral Sciences 1994; 16: 99–115.

Gil AG, Vega WA. Two different worlds: acculturation stress and adaptation among Cuban and Nicaraguan families. Journal of Social and Personal Relationships 1996; 13: 435–56.

Gil AG, Vega WA, Dimas JM. Acculturation stress and personal adjustment among Hispanic adolescent boys. Journal of Community Psychology 1994; 22: 43–54.

Guarnaccia PJ, Canino G, Rubio-Stipec M, Bravo M. The prevalaence of ataques de nervios in the Puerto Rico

disaster study: the role of culture in psychiatric epidemiology. Journal of Nervous and Mental Diseases 1993; 181(3): 157–65.

Guarnaccia PJ, Rivera M, Franco F, Neighbors C. The experiences of ataques de nervios: towards an anthropology of emotions in Puerto Rico. Culture, Medicine, and Psychiatry 1996; 20(3): 343–67.

Guarnaccia PJ, Rogler LH. Research on culture-bound syndromes: New directions. American Journal of Psychiatry, 1999: 156(9): 1322–7.

Kessler RC, Üstün TB. The World Mental Health (WMH) survey initiative version of the World Health Organization Composite International Diagnostic Interview (CIDI). IJMPR 2004; 13(2): 93–121.

Kirmayer JL, Young A. Culture and context in the evolutionary concept of mental disorder. Journal of Abnormal Psychology 1999; 108: 446–52.

Kleinman A, Good B. Culture and Depression: Studies in the Anthropology and Cress Cultural Psychiatry of Affect and Disorder. Berkeley: California University Press, 1985.

Lopez SR, Guarnaccia PJJ. Cultural psychopathology: Uncovering the social world of mental illness. Annual Review of Psychology 2000; 51: 571–98.

Matias-Carrelo LE, Chavez LM, Negron G, Canino G, Aguilar-Gaxiola S, Hoppe S. The Spanish translation and cultural adaptation of five mental health outcome measures. Culture, Medicine and Psychiatry, 2003; 27(3), 291–313.

Mays VM, Cochran SD. Mental health correlates of perceived discrimination among lesbian, gay, and bisexual adults in the United States. American Journal of Public Health 2001; 91(11): 1869–76.

McKinlay JB, Marceau LD. A tale of three tails. American Journal of Public Health 1999; 89(3): 295–8.

Mezzich JE, Kirmayer LJ, Kleinman A, Fabrega Jr H, Parron DL, Good BJ, Lin K-M, Manson SM. The place of culture in DSM-IV. Journal of Nervous and Mental Disease 1999; 187(8): 457–64.

National Institute of Mental Health. Cooperative agreement for a multi-site study of mental health service use, need, outcomes, and costs in child and adolescent populations (UNO-CAP). In General Services Administration (eds) Catalog of Federal Domestic Assistance 93.242-Request for Applications. Rockville MD: NIMH, 1994.

Olson DH. Circumplex model VII: validation studies and FACES III. Family Process 1986; 25: 337–51.

Olson DH. Circumplex model of family systems VIII: family assessment and intervention. In DH Olson, CS Russell, DH Sprenkle (eds) Circumplex Model: Systemic Assessment and Treatment of Families. New York: Haworth Press, 1989.

Ortiz V. The diversity of Latino families. In RE Zambrana (eds) Understanding Latino Families: Scholarship, Policy, and Practice. Thousand Oaks CA: Sage, 1995.

Pennell BE, Gebler N, Chardoul S, Carr D, Pennell S, Torres M, Hupp A. Plan and Operation of the National Survey of Health and Stress, the National Survey of American Life, and the National Latino and Asian American Survey. IJMPR 2004; 13(2): 241–69

Rehm J, Üstün TB, Saxena S, Nelson CB, Chatterji S, Ivis F, Adlaf E. On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. IJMPR 1999; 8: 110–23.

Rogler LH. Methodological sources of cultural insensitivity in mental health research. American Psychologist 1999; 54; 424–33.

Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 1997; 277: 918–24.

Schulz A, Israel B, Williams D, Parker E, Becker A, James S. Social inequalities, stressors, and self reported health status among African American and white women in the Detroit metropolitan area. Social Science and Medicine 2000; 51(11): 1639–53.

Takeuchi DT, Chung RC, Lin K, Shen H, Kurasaki K, Chun CA, Sue S. Lifetime and twelve-month prevalence rates of major depressive episodes and dysthymia among Chinese Americans in Los Angeles. American Journal of Psychiatry 1998; 155; 1407–14.

Vega W, Zimmerman R, Gil A, Warheit G, Apospori E. Acculturation strain theory: its application in explaining drug use behavior among Cuban and other Hispanic youth. In M De La Rosa, JL Recio (eds) Drug Abuse Among Minority Youth. Rockville MD: National Institute on Drug Abuse, 1993.

Vega WA. Hispanic Families in the 1980s: a decade of research. Journal of Marriage and the Family 1990; 52: 1015–24.

Vega WA, Gil AG, Warheit GJ, Zimmermann RS, Apospori E. Acculturation and delinquent behavior among Cuban American adolescents: toward an empirical model. American Journal of Community Psychology 1993; 21; 113–25.

Vega WA, Kodoly B, Aguilar-Gaxiola S, Alderte E, Catalano R, Caraveo-Anduaga H. Lifetime prevalence of DSM-III-R psychiatric disorders among urban and rural Mexican Americans in California. Archives of General Psychiatry 1998; 55; 771–8.

Vega WA, Kolody B, Valle JR, Hough R. Depressive symptoms and their correlates among immigrant Mexican women in the United States. Social Science and Medicine 1986; 22(6): 642–5.

Williams DR, Yu Y, Jackson JS, Anderson NB. Racial differences in physical and mental health. Journal of Health and Psychology 1997; 2(3): 335–51.

World Health Organization. Disability Assessment Schedule II (WHO-DAS II). Geneva: World Health Organization, 1998.

Correspondence: Margarita Alegria Center for Multicultural Mental Health Research, Cambridge Health Alliance, 120 Beacon St. 4th floor Somerville, MA 02143, USA.
Telephone (+1) 617 5038440.
Fax (+1) 617 5038430.
Email malegria@charesearch.org.