

Cultural Adaptation of Physical Activity Self-Report Instruments

Elva M. Arredondo, Tamar Mendelson, Christina Holub, Nancy Espinoza,
and Simon Marshall

Context: The validity of physical activity (PA) self-report measures can be a problem when using these measures with target populations that differ from the population for which the measures were originally developed. **Objectives:** Describe an approach to further tailor PA self-report measures to a target community, and report on focus group and cognitive interview findings. **Process:** Topics relevant to culturally tailoring measures are discussed, including translation, focus groups, and cognitive interviews. We describe examples from our own work, including focus groups and cognitive interviews conducted to assess Latinos' interpretations of PA questions derived from various epidemiological surveys that were developed in White communities. **Findings:** Findings from focus groups and cognitive interviews provide valuable information about the comprehension, interpretation, and cultural relevance of the PA questions to Latino communities. **Conclusions:** It is recommended that investigators collect formative data to better assess the equivalence of items being applied to a different cultural group. Guidelines for cultural attunement of self-report instruments are described to promote more uniform and rigorous processes of adaptation and facilitate cross-cultural investigations.

Keywords: culturally tailoring measures, cognitive testing, self report measures

Subjective and objective measures of physical activity (PA) suggest low rates of PA in the United States¹⁻² However, ethnic patterns of physical activity vary depending on the type of instrument used to measure physical activity. Nationally representative data collected through accelerometers suggest that Mexican American men and women engage in higher levels of physical activity than non-Hispanic white or non-Hispanic black men and women.² These findings contrast with those from studies that have collected physical activity patterns through self report instruments, showing lower rates of physical activity among Hispanics.³ This discrepancy between the accelerometer and self report survey may be due, in part, to challenges in recalling physical activity or the fact that accelerometers assess movement below the waist. Another contributing factor may be problems with the cross-cultural validity of the self-report instruments used to assess PA, including poor translation, lack of culturally appropriate items, and failure to assess the types of activities in which community members participate. Despite the limitations of self-report measures, they can provide valuable information about the types of physical activity (ie, occupational, leisure time) in which individuals engage, as well as other relevant constructs,

such as participation in specific physical activity program and seasonality of physical activity. Accurately measuring physical activity through self-report measures can provide detailed information on physical activity in underserved communities, which can inform interventions and programs.

One of the most common challenges investigators face when conducting cross-cultural research is developing measurement tools that elicit valid and reliable responses regarding the research construct of interest. The goal of this paper is to describe steps researchers can take when adapting self-report survey measures for cross-cultural use so as to increase the reliability and validity of those instruments. We outline key considerations regarding survey translation and adaptation, including ways to incorporate the target community's input into the adaptation process (Figure 1). Of note, this paper does not present an exhaustive discussion of all approaches relevant to adaptation of measures. For instance, analytic strategies for assessing measurement equivalence (eg, confirmatory factor analysis) are not covered here; the reader is referred to other resources for a discussion of those issues.^{4,5} In addition, when conducting research in a target community with dramatic cultural differences from the population in which measures were generated, the development of new measures is advised, involving topics that are beyond the scope of this paper. We focus instead on recommended practices for adapting self-report measures with sound psychometric properties for use with a new target population.

Arredondo, Holub, Espinoza, and Marshall are with the School of Public Health, San Diego State University, San Diego, CA. Mendelson is with the Dept of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

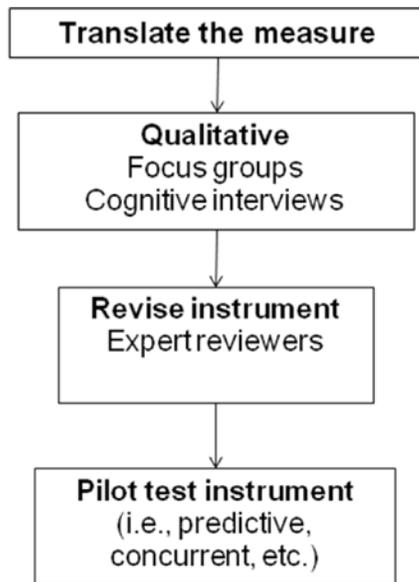


Figure 1 — General process for adapting measures to a cultural group.

Construct Validity and Measurement Error

Construct validity refers to the degree to which a measure accurately assesses the construct (conceptual definition of a variable) of interest.⁵ A measure that does not accurately measure the intended construct is problematic (eg, by having excessive amounts of random error, or measurement specific error, or method bias) and can lead to erroneous conclusions. For instance, let us suppose that an investigator is interested in comparing the leisure time physical activity of Latinas and non-Latina White women. The investigator translates a physical activity questionnaire from English to Spanish without verifying that her translation of “leisure time physical activity” has the same meaning to the target population that it has to her. If the findings suggest lower leisure time physical activity among Latinas compared with non-Latina White women, it is difficult to know whether Latinas truly engaged in lower levels of leisure physical activity than non-Latina White women or whether the Latina participants simply did not understand the concept of leisure time physical activity as intended. In this example, measurement error arises from a discrepancy between what an investigator is trying to assess and what he or she *actually* assesses. Erroneous assumptions about health behavior patterns can be made in such cases.

Measurement error can lead to other negative consequences. For example, error can mask relationships between measures. In other words, a physical activity survey that contains measurement error may not correlate accurately with other key measured constructs, such as body mass index and other physiological fitness

indicators or psychological variables. In addition, participants who are confused about what is being asked in a survey are likely to provide inaccurate responses or skip items altogether. Missing data compromise statistical power, which in turn reduces the likelihood that true associations between variables will be detected.⁶

Cultural Competence

Understanding the target community and its culture is critical for reducing investigator biases and increasing the likelihood of accurate measurement. Cross et al⁷ note that cultural competence is a complex framework and emphasizes that the process of achieving cultural competency occurs along a continuum that includes (1) “cultural destructiveness,” to (6) “cultural proficiency.” Culturally competent and proficient researchers rigorously evaluate the appropriateness of any survey that will be used with the target community. As a culturally competent researcher, one may consider including experts on the target community or individuals from the target community in a research team or an advisory board and taking other steps to understand the target population through spending time in that community and reading relevant publications and materials. Most available measures have been developed in middle-class White communities, which presents a challenge to investigators who wish to administer them to other racial/ethnic or socioeconomic groups. The steps outlined below provide a guide to researchers on ways to culturally adopt instruments to a target population. Many of these steps have been outlined in detail by other researchers.⁸⁻¹¹

Translation

Linguistic translation is often necessary when adapting a measure for use in cross-cultural research. A number of considerations are relevant for minimizing measurement error as a result of translation. As idiomatic expressions (ie, colloquialisms and metaphors) are culturally bound, investigators may want to avoid including them in the translated version. In addition, translated items should preferably be written at a third-grade reading level and should not exceed the sixth-grade level.¹² Hilton and Skrutkowski¹³ provide additional suggestions, such as ensuring that items are written in the active versus passive voice to facilitate comprehension, keeping items simple and short (fewer than 16 words), and repeating nouns rather than using pronouns.

Increasing Linguistic Equivalence

There are a number of different approaches to translating a measure from one language to another. One-way translation, the most straightforward and frequently-used approach, involves translation by a single bilingual person from the original to the target language. Although one-way translation has the advantage of requiring minimal resources, it is limited by reliance on a single

individual's linguistic skills and interpretive abilities, which often leads to low validity and reliability in the translated measure.¹⁴⁻¹⁷ The second most common form of translation is the backtranslation method, in which one bilingual translator translates the survey from the original to the target language and a second bilingual translator translates it back to the original language. Differences between the translated items are reconciled by the translators. Although this method requires more resources, it is generally more accurate than one-way translation because it incorporates input from 2 translators. A committee approach involves 3 or more bilingual individuals who translate the survey items independently and then compare their translated versions and together agree on the final translated version. A potential drawback is the social pressure committee members may feel to agree to the translated items. However, the committee approach is generally the strongest of the 3 in that it involves members who are knowledgeable about questionnaire design and the measurement objectives; thus, the reviewers will likely keep in mind the intent of the question. Following the translation of the instruments, it is recommended that a team of experts review the translated measure to check for the omission of words and correct grammatical/spelling errors.

The most significant limitation in all the above translation methods is the involvement of bilingual translators because these translators often have different world views and may interpret survey items differently than the monolingual speakers in the target population who will be surveyed.¹⁸⁻¹⁹ Irrespective of the translation method, simply translating a survey from one language to another does not lead to linguistic and conceptual equivalence.^{16,20} Incorporating the input of monolingual representatives through one-on-one interviews or focus groups may help address this limitation. Qualitative methods can be collected in the form of focus groups and/or cognitive interviews.

Conceptual Equivalence

Conceptual equivalence between measures is accomplished when survey items have the same meaning in 2 or more cultures. Qualitative methods can help achieve conceptual equivalence between measures and minimize measurement error or bias. Qualitative approaches involve gathering input from the target community about the content of the measure either individually or in a group format or both. Qualitative methods capture the target community's point of view, ideas, thoughts, and beliefs, which can be used to shape the measure appropriately.²¹

Focus Groups

Focus groups are a commonly-used method by which investigators attempt to achieve the conceptual equivalence of measures. The goal of conducting focus groups is to elicit information from the target community regarding

their understanding and interpretation of the survey instructions and items. Focus groups usually involve 1 or more facilitators and several participants (6–10) who provide input on the wording and content of the measure. Focus groups usually last approximately 60 to 90 minutes, and the discussions are generally tape-recorded, transcribed, and coded by the investigative team. Following data collection, the investigative team incorporates the input collected from the community and modifies the measure accordingly.

Linguistic or cultural adaptation of measures may be necessary even among groups that share many cultural characteristics. For instance, although Latinos share many cultural traits, there is a great deal of linguistic and cultural variation across Latino subgroups.^{20,22,23} Thus, one cannot assume that measures that have been pilot tested with Mexican Americans are appropriate for use with Puerto Ricans. There is also diversity *within* Latino subgroups largely due to acculturation and income influences.

The following is an example of how focus groups can be used that also illustrates the inclusion of diverse Latino subgroups. Our group conducted formative research with 4 Latino subgroups (Cuban Americans, Mexican Americans, Central Americans, and Puerto Ricans) for a 4-site study (Miami representing Cuban Americans; The Bronx representing Puerto Ricans; Chicago representing Central Americans; and San Diego representing Mexican Americans) that aimed to develop a single survey appropriate for all these groups.²⁴ The majority of the surveys were developed in White communities. A professional translator translated all the survey items before our collection of focus group data. Following the translation process, a committee of investigators representing each of the Latino subgroups evaluated the appropriateness of the translated survey items. The committee identified items that they believed required pilot testing in focus groups involving members of each of the target communities in each group. The goal of the focus groups was to solicit input from each of the 4 Latino community subgroups on the wording of the items and appropriateness of the response categories. In San Diego, we recruited members of one Latino subgroup (Mexican Americans) to participate in a focus group. During the 90 minute focus group, 2 bilingual Research Assistants gave participants the list of statements, questions and responses to be discussed in the focus groups. Participants were asked to describe their thoughts about the clarity and purpose of the question. Each focus group lasted approximately 60 minutes. Findings from the focus groups suggested that there were some variations in interpretation across the different groups and that a number of words in the items and response categories required modification (Table 1). For example, participants indicated that "football" (*futbol*), which means soccer, needed to be changed to "American football" ("*football americano*"). Other suggestions included changing the wording to reflect more formal Spanish and revising the language to be appropriate for individuals with a lower literacy level.

Table 1 Measurement Items That Were Adapted

Type of changes	Former question	Modified questions
Meaning	“Futbol” (soccer)	“Futbol americano” (football)
Literacy	“expendedora” (vending machine)	“maquinas que venden alimentos” (machines that sell food)
	“Un dia normal” (a normal day)	“Un dia tipico” (a typical day)
	“Convive con su pareja” (cohabitate with your partner)	“Vive con su pareja” (live with your partner)
	“intensidad vigorosa” (vigorous)	“mucho esfuerzo fisico” (a lot of effort physically)
Inclusion of a variety of words	“cirugia” (surgery)	“operacion” (operation)
	“lugar de veneracion o culto” (place of worship or sacred place)	“lugar de veneracion, oracion, o culto” (place of worship, sacred place, or prayer place)
Cultural relevance: What activities do you like to do?	Remove hunting, fishing, and rock-climbing	The following activities were deleted: limpieza (cleaning), construccion (construction), fabricas (factories), cajero (cashiers), carniceria (meat sellers)
Metric	“dos pies” (two feet)	“24 pulgadas” (24 inches)
Feminine words	“nervioso” (nervous)	“nervios(a)” (nervous)
Tu vs. Usted (Informal vs. formal)	“describe come te sientes” (tell us how you feel)	“describa come se siente” (tell us how you feel)

One advantage to conducting a focus group compared with conducting individual interviews is that the investigator can usually obtain general information about the measure’s equivalence with 1 or 2 focus groups. Thus, implementing focus groups is faster, requiring fewer resources than cognitive interviews. However, a potential limitation is that participants may provide socially desirable answers and/or may not feel comfortable responding to sensitive questions in a group format. One way to help reduce socially desirable answers is by conducting one-on-one interviews or cognitive interviews.

Cognitive Interviews

Cognitive interviews can augment the information collected in focus groups, and it is advisable to conduct cognitive interviews as a follow up to focus groups. Data collected from focus groups informs an evaluation of the conceptual equivalence of items and helps to identify sources of measurement error.²⁵ Cognitive interviews provide more detailed and specific feedback, which can be used in the modification of available instruments or the development of new ones.^{26,27} Cognitive interviews involve interviewing 1 participant from the target community at a time. During the interview, the investigator delves deeply into a specific topic or question to better

understand how the target community thinks about what is being asked. The questions are asked in a standardized manner so as to be equivalent across participants; they include open-ended probes that ask respondents to think aloud, highlight problems, express their opinions, and evaluate survey items in detail. Cognitive interviews help reduce or minimize measurement error by providing information that helps the investigator to modify a measure so that it can be better understood by the target community.

In cognitive interviews, several individuals from the target community are invited to participate in one-on-one intensive interviews conducted by a trained interviewer. The one-on-one interviews generally last 45 to 90 minutes and are tape recorded, transcribed, and coded. The cognitive interviews begin with a carefully crafted introduction to help minimize difficulties in understanding and responding to cognitive probes.¹⁰ The information gathered in these interviews can improve specific survey items. For instance, an interviewer may ask the participant to explain what she thinks the following question is asking: “How often do you exercise?” The interviewer may ask the participant to articulate any difficulty she might have in responding to that question, whether there are any words she would change, and how she would ask the question to a friend. The interviewer may also ask the

participant to discuss the meaning of specific words (eg, “exercise”) and the difficulty in recalling the information needed to answer the question correctly. It is important for the interviewer to read the item to the participant exactly how it is written in the measure and to ask participants how they are interpreting key words in the item. The advantage of conducting cognitive interviews is that the investigator can explore the participant’s reactions to the measure in considerable depth. A notable disadvantage is that it often requires more time and expense because the interviews are conducted one-on-one by trained interviewers and are also transcribed and coded, a time-consuming process.

Cognitive Interview Methods and Findings

Members of a Latino subgroup (Mexican American) were recruited from a community near the border region to participate in cognitive interviews. A trained bilingual graduate research assistant scheduled individual appointments with 10 participants in their homes. Following an introduction of what was going to be asked of each participant and role-play exercises to practice cognitive interviewing skills, the research assistant asked participants to answer the questions read to them. Cognitive probes administered were either prescribed or, in some cases, were devised spontaneously during the interview. All interviews were audio taped, transcribed by the research assistant, coded by 2 independent bilingual/bicultural reviewers (the Principal Investigator and research assistant), and compared.

Findings from the cognitive interviews suggested it would be advisable to rephrase some of the original questions. For example, participants were asked, “During the past month, other than your regular job, did you participate in any physical activity or exercises such as running, calisthenics, golf, gardening, or walking for exercise (BRFSS)?” They were subsequently probed, “Are there any examples of physical activity given relevant to you?” Many participants indicated that they were not aware of the meaning of “calisthenics.” Furthermore, several stated that “golf” was not an activity they engaged in, thereby not applicable to them. Participants were also presented with the following item: “First, think about all the vigorous activities which take hard physical effort that you do as part of your work. Vigorous activities make you breathe much harder than normal. These may include things like heavy lifting, digging, heavy construction work, or climbing up stairs. Think about only these vigorous physical activities that you did for at least 10 minutes at a time. During the last 7 days, on how many days did you do vigorous physical activities as part of your work?” When probed, “How easy or difficult did you find this question (IPAQ)?” one participant noted that “I don’t understand why asked about the things I do for 10 minutes . . . that’s confusing . . . I don’t know.” Other participants explained that it was difficult to think about physical activity in terms of 10 minutes. About half of

the participants noted that the questions were rather long and required for individuals to consider many factors (duration, number of days, and intensity, and setting) when thinking about an answer, which was challenging.

Participants were also asked about different types (ie, traveling, occupational, etc.) of physical activity: “Now, think about how you traveled from place to place, including to places like work, stores, movies and so on. How much time did you usually spend on one of those days traveling in a car, bus, train or other kind of motor vehicle?” Later, participants were probed, “Is there anything you would change about this question?” One participant noted, “I would change that word ‘traveling’ (*viajar*) . . . it made me think . . . like traveling for a trip.” Participants suggested to replace it with “to go” (*irse*). Moreover, individuals were asked about what they considered “occupational activity.” Almost all indicated that they believed that they were being asked questions about paid work. This is in contrast to the instructions that are given in the GPAQ (considers paid and unpaid work like volunteering when considering physical activity). When asked to consider physical activity, combining paid and unpaid work can be a challenge for many individuals, particularly those who volunteer for church ministries as many would not consider this effort as “work.” These examples illustrate that members of our target population had several interpretations of the survey items that we had not intended, highlighting the need to modify the original wording to enhance cultural equivalence.

We asked participants to define, in their own words, different physical activity intensities. Several participants explained that they had not heard the term “vigorous activity.” Those that had heard this term, thought it had a negative connotation (“working too hard”). When asked to define “moderate activity,” one 38-year-old man indicated that he equated to breathing a little more than normal, validating the definition that is given in the GPAQ. In addition, when asked to define physical activity, men associated it with weights and equipment whereas women related it to going to the gym and participating in sports, underscoring the importance of including all of these activities in a PA survey. Consistent with the intended definition of leisure-time PA, participants interpreted leisure time PA as “pleasurable,” “without other commitments,” and “their own time.”

Discussion

The goal of this paper is to recommend steps that would help minimize measurement error and maximize measurement equivalence when conducting cross-cultural research. To illustrate these steps, we described our collection of formative data to improve physical activity related questions for the Latino (Mexican) community in San Diego, CA. We asked participants to interpret specific questions and key words from various health surveys by conducting focus groups and cognitive interviews. Information from these interviews guided the development of specific recommendations for our target population.

We found that certain categories of physical activity from the original survey were not understood as intended. One resulting recommendation was to more clearly define “work-related” and “leisure-time” physically activity. The use of culturally relevant examples (eg, walking, jogging, aerobics, soccer, etc.) can help in making the appropriate distinctions. Our data also highlighted the importance of using terms that the population is familiar with and that have the intended meaning. These included “difficult to do” versus “vigorous,” and “going from place to place” versus “traveling.” Taking into account these recommendations will help to minimize measurement error and maximize equivalence when adapting physical activity-related questions for the Latino community in San Diego, CA.

Reducing measurement error improves the degree to which true health behavior patterns can be observed. When the community does not interpret questions the way they were intended, error in the measure is introduced. Differences in health behaviors, for example, may be due to measurement error and inaccurate interpretation of the items rather than true differences. There is often heterogeneity within cultures, as is observed among Latinos.²⁸ Potential differences within subgroups of a culture are important to capture and distinguish to minimize measurement error and avoid misleading conclusions about health behavior patterns.

There are several methods for best capturing how a target community understands the underlying concepts presented in survey questions. Qualitative methods, such as focus groups and cognitive interviewing, allow participants to respond in a way that accurately represents their point of view, ideas, thoughts, and beliefs.²¹ The use of qualitative methods can help maximize conceptual equivalence between measures, thereby minimizing measurement error. Qualitative approaches solicit valuable feedback and information from the target community and can be done individually, in a group, or both. This strategy provides a critical component of measurement adaption, allowing researchers to assess the interpretation of specific questions and relevance of the concepts to the target community.

We found similarities in the information collected in cognitive interviews and focus groups. For instance, participants in cognitive interviews and focus groups suggested we replace “vigorous activity” with “working intensely.” There are advantages and disadvantages to conducting focus groups compared with cognitive interviews. As mentioned above, focus groups take less time and require fewer resources, but provide more general information on how the community interprets the questions. By contrast, a researcher can understand in greater detail how people think about activity when collecting data from cognitive interviews; this information can improve item construction. For example, when asked “How much time do you spend walking each day?” one participant indicated that she walks to the bus station every day, which takes her 15 minutes. She recommended

we add “walking to the bus” as a prompt in the GPAQ questionnaire.

Many of the strategies outlined in this paper and described in more detail in previous research⁸⁻¹¹ can apply to the adaptation of measures for use with target communities that vary by culture, gender, age, or other characteristics, and to various health topics. The steps outlined here include several options for the translation process, with the optimal approach involving multiple translators, followed by review by an expert committee. Next, survey items are presented to the target community, ideally including monolingual members, and culturally specific feedback is gathered regarding interpretation and relevance of the questions. Lastly, the iterative process of adjusting the questions and pretesting the survey for conceptual equivalence is an essential step. Figure 1 is a graphical representation of the process for maximizing conceptual equivalence. The adaptation of an instrument can take a large amount of resources, and where possible the most rigorous strategies should be used. However, in practice, many studies lack adequate resources. In such cases, researchers must strategize and select methods based on their knowledge of the target community and available resources. A key point of this paper is that cultural competency or proficiency on part of the investigators can help identify the biases that contribute to measurement error. In cross-cultural research, this competency informs the use of rigorous formative research to adapt measures.

References

- Centers for Disease Control and Prevention. *Prevalence of self-reported physically active adults*. 2008.
- Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181-188.
- Ham SA, Yore MM, Kruger J, Heath GW, Moeti R. Physical activity patterns among Latinos in the United States: putting the pieces together. *Prev Chronic Dis*. 2007;4(4):A92.
- Devellis RF. *Scale development: theory and applications*. Thousand Oaks, CA: Sage Publications; 2003.
- Carmines EG, Zeller RA. *Reliability and validity assessment*. Newbury Park: Sage Publications; 1991.
- Allison PD. *Missing data: quantitative applications in the social sciences*. Thousand Oaks: Sage Publications, Inc.; 2002.
- Cross T, Bazron B, Dennis K, Isaacs M. *Towards a culturally competent system of care*. Washington, DC: Georgetown University Child Development Center; 1989.
- Goerman P. *Adapting cognitive interview techniques for use in pretesting Spanish Language*. *Research Report Series*. Washington DC: US Census Bureau; 2006.
- Willis G, Lawrence D, Hartman A, Stapleton Kudela M, Levin K, Forsyth B. Translation of a tobacco survey into Spanish and Asian languages: the Tobacco Use Supplement to the Current Population Survey. *Nicotine Tob Res*. 2008;10(6):1075-1084.

10. Levin K, Forsyth BH, Stark D, et al. Using cognitive interviews to evaluate the Spanish-Language translation of a dietary questionnaire. *Survey Research Methods*. 2009;3(1):13–25.
11. Willis GG, Stapleton Kudela M, Levin K, et al. Evaluation of a multistep survey translation process. In Harknewss JA., Braun M, Edwards B, et al (eds.). *Survey methods in multinational, multi-regional, and multicultural contexts*. Hoboken, NJ: Wiley; 2010.
12. Solomon FM, Eberl-Lefko AC, Michaels M, Macario E, Tesauro G, Rowland JH. Development of a linguistically and culturally appropriate booklet for Latino cancer survivors: lessons learned. *Health Promot Pract*. 2005;6(4):405–413.
13. Hilton A, Skrutkowski M. Translating instruments into other languages: development and testing processes. *Cancer Nurs*. 2002;25(1):1–7.
14. Sperber AD, Devellis RF, Boehlecke B. Cross-cultural translation: methodology and validation. *J Cross Cult Psychol*. 1994;25(4):501–524.
15. Erkut S, Alarcón O, Coll CG, Tropp LR, García HAV. The dual-focus approach to creating bilingual measures. *J Cross Cult Psychol*. 1999;30(2):206–218.
16. Geisinger KF. Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychol Assess*. 1994;6(4):304–312.
17. Bontempo R. Translation fidelity of psychological scales: an item response theory analysis of an individualism-collectivism scale. *J Cross Cult Psychol*. 1993;24(2):149–166.
18. Candell GL, Hulin CL. Cross-language and cross-cultural comparisons in scale translations: independent sources of information about item nonequivalence. *J Cross Cult Psychol*. 1986;17(4):417–440.
19. Drasgow F, Hulin CL. Cross-cultural measurement. *Rev Interam Psicol*. 1987;21(1-2):1–24.
20. Marin G, Marin B. *Research with Hispanic populations*. Newbury Park, CA: Sage Publications; 1991.
21. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res*. 1998;7(4):323–335.
22. Warnecke RB, Johnson TP, Chávez N, et al. Improving question wording in surveys of culturally diverse populations. *Ann Epidemiol*. 1997;7(5):334–342.
23. Ponce NA, Lavarreda SA, Yen W, Brown ER, DiSogra C, Satter DE. The California Health Interview Survey 2001: translation of a major survey for California's multiethnic population. *Public Health Reports (Washington, D.C.: 1974)*. 2004;119(4):388–395.
24. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20(8):642–649.
25. Wills G. *Cognitive interviewing: a tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications; 2005.
26. Friedenreich CM, Courneya KS, Bryant HE. The lifetime total physical activity questionnaire: development and reliability. *Med Sci Sports Exerc*. 1998;30(2):266–274.
27. Durante R, Ainsworth BE. The recall of physical activity: using a cognitive model of the question-answering process. *Med Sci Sports Exerc*. 1996;28(10):1282–1291.
28. Berrigan D, Forsyth BH, Helba C, Levin K, Norberg A, Willis GB. Cognitive testing of physical activity and acculturation questions in recent and long-term Latino immigrants. *BMC Public Health*. 2010;10:481.