



Published in final edited form as:

J Aging Health. 2012 September ; 24(6): 992–1017. doi:10.1177/0898264312440321.

A Framework for Understanding Modifications to Measures for Diverse Populations

Anita L Stewart, PhD,

Institute for Health & Aging, University of California San Francisco, 3333 California St. Suite 340, San Francisco, CA 94118, Phone: 415 502-5207, anita.stewart@ucsf.edu

Angela D Thrasher, PhD, MPH,

Department of Health Behavior and Health Education, University of North Carolina Gillings School of Global Public Health, 315 Rosenau Hall, CB #7440, Chapel Hill, NC 27599-7440, Phone: 919-843-9293, angela.thrasher@unc.edu

Jack Goldberg, PhD, and

Vietnam Era Twin Registry, Seattle VA and the University of Washington School of Public Health, Box 359780, 1730 Minor Avenue, Suite 1760, Seattle, WA 98105-1597, Phone: 206 543-4667, jack.goldberg@va.gov

Judy A. Shea, Ph.D.

University of Pennsylvania, School of Medicine, Division of General Internal Medicine, 1223 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, Phone: 215 573-5111, sheaja@mail.med.upenn.edu

Abstract

Objectives—Research on health disparities and determinants of health disparities among ethnic minorities and vulnerable older populations necessitates use of self-report measures. Most established instruments were developed on mainstream populations and may need adaptation for research with diverse populations. Although information is increasingly available on various problems using these measures in diverse groups, there is little guidance on how to modify the measures. We provide a framework of issues to consider when modifying measures for diverse populations.

Methods—We describe reasons for considering modifications, the types of information that can be used as a basis for making modifications, and the types of modifications researchers have made. We recommend testing modified measures to assure they are appropriate. Suggestions are made on reporting modifications in publications using the measures.

Discussion—The issues open a dialogue about what appropriate guidelines would be for researchers adapting measures in studies of ethnically diverse populations.

Keywords

Measurement; modifying measures; adapting measures; minority aging; health disparities

Eliminating health disparities among ethnic minorities and vulnerable older populations is a national priority (Smedley, Stith, & Nelson, 2003; U.S. Department of Health and Human Services, 2000). To accomplish this goal, it is essential to understand factors that contribute to these disparities (National Institute on Aging, 2011). Research on health disparities and its

determinants among ethnic minorities and vulnerable older populations necessitates use of self-report measures. However, many of these widely-used measures were developed and tested on mainly white, young and middle-aged, well-educated samples. These measures may have limitations when used in studies of minority or lower-socioeconomic status (SES) older adults included in health disparities and minority aging research.

A substantial amount of research has attempted to address these issues over the past 15 years, including systematic efforts by the Resource Centers for Minority Aging Research (RCMARs) (Stahl & Hahn, 2006); also see preface for an overview of RCMAR contributions in measurement in diverse older populations (Teresi, Stewart, & Stahl, 2012). There are published guidelines on methods for examining the conceptual and psychometric adequacy of measures in ethnically diverse populations (Collins, 2003; Hahn & Cella, 2003; Johnson, 2006; A.M. Nápoles-Springer, Santoyo-Olsson, O'Brien, & Stewart, 2006; A.M. Nápoles-Springer & Stewart, 2006; A.L. Stewart & Nápoles-Springer, 2003; Teresi, Stewart, Morales, & Stahl, 2006). There also are systematic reviews examining the conceptual or psychometric adequacy of particular concepts and measures in diverse populations (Coates & Monteilh, 1997; A.L. Stewart & Nápoles-Springer, 2000) including older adults (Mui, Burnette, & Chen, 2001; Mutran, Reed, & Sudha, 2001).

However, when problems are found with measures in diverse population groups, there are no guidelines on what to do next. Once a measure has been determined to be inappropriate for minority or lower-SES participants, researchers have three options. One is to use the measure “as is” without modification and articulate the limitations. This abides by the tenet of administering measures as published which some believe preserves score reliability and validity (Juniper, 2009). However, if a measure is not suitable for a population, scientific inferences derived from it may be compromised.

A second option is to create a new measure *de novo*. Developing a new measure is fraught with challenges, including having the expertise, time, and resources to develop and test a new measure, which may not be practical for most health disparities researchers. Methods for developing new measures to be culturally sensitive involve a mixed-methods approach that includes concept development, writing items, pretesting, revising, field testing, and conducting psychometric analysis to derive final measures. Although there are no general guidelines for developing new measures in minority populations, several publications provide detailed examples of these steps (Jackson, 1996; Krause, 2006; A. L. Stewart, Napoles-Springer, Gregorich, & Santoyo-Olsson, 2007).

The third option is to modify or adapt an existing measure. There is a delicate balance between trying to retain the strength of an existing measure, which may have undergone extensive development and testing, but clearly will be problematic, and making modifications which may or may not work. Furthermore, there are very few practical guidelines on how to go about making those modifications. In multi-national research, measures developed in English must be translated for use in non-English speaking countries. In guidelines for multi-national studies, adaptations are an integral part of the translation process, e.g., items may be modified to achieve “semantic” equivalence (Aaronson et al., 1992; Beaton, Bombardier, Guillemin, & Ferraz, 2000; Bullinger et al., 1998). However, these guidelines give little attention to issues in modifying measures when no language translation is needed.

This paper attempts to address this gap by providing a framework for modifying measures to improve their reliability and validity in health disparities research involving ethnically and racially diverse populations. The goal of such modifications would be to increase the likelihood that the modified measure has comparable meaning, reliability, and validity as the

original measure, but in a new population group. The issues raised here pertain primarily to addressing differences from mainstream populations (on which original measures were tested) in socioeconomic status, race/ethnicity, language, and literacy. Issues of modifications based on these types of group differences are relevant to health disparities studies of adults of all ages. Occasionally, modifications are made to adapt a measure specifically to be more appropriate for older adults (Thiamwong, Stewart, & Warahut, 2009; Vanderplas & Vanderplas, 1981). We describe three issues: 1) reasons for considering modifications – why a modification would be needed; 2) the basis for modifications – information that can be used to make the modifications; and 3) possible types of modifications. We conclude with recommendations for assessing modified measures and for reporting results of these assessments.

Reasons for Considering Modifications to a Measure

There are a number of reasons why investigators might consider modifying a measure. In health disparities research, the most common reason is that the population group(s) being studied differs substantially from the one in which the original measure was developed. Essentially, the motivation for modifying measures is the concern that racial/ethnic or generational differences might adversely affect the meaning, reliability, or validity of the original measure (information on these basic measurement concepts is available elsewhere, e.g., McDowell & Newell, 2006; Nunnally & Bernstein, 1994). Some key reasons why measures developed in one group may not be suitable include: 1) a concept or dimension may be missing from a measure; 2) the meaning of the concepts or items may differ; 3) items/phrases may not be interpreted as intended; 4) the process of responding to the questions is complex or difficult; and 5) the study context or mode of administration may differ from the original. Any of these reasons could result in findings that measures do not meet minimal psychometric criteria in these new groups or that the measures do not demonstrate invariance across groups (Gregorich, 2006; A.L. Stewart & Nápoles-Springer, 2000; A.L. Stewart & Nápoles-Springer, 2003).

Concept or dimension is missing

Many concepts used in health disparities research are complex and multidimensional. As researchers delve into existing measures, it often is apparent that key dimensions relevant to a diverse group may be missing. For example, in focus groups exploring social support concepts for older Chinese and Korean immigrants, getting help with English in conducting business or in health care visits and aspects of financial support were identified as important but were not included in standard measures of social support (Wong, Yoo, & Stewart, 2005).

The meaning or appropriateness of concepts may differ

The meaning of the existing concepts that are the target of a measure may differ by race or ethnicity or by generation. Published measures reflect the population, place, and time in which they were developed. To the extent that a new population group may define or perceive the concept differently, the original measure may lack content validity for the new group. One example is the use of food frequency questionnaires to assess intake of foods and nutrients. The validity of food frequency instruments is partly dependent on the list of foods, thus use of a food frequency questionnaire developed for a general population in a study of a minority population may result in underestimates of intake unless it is adapted for the new population (Tucker, Bianchi, Maras, & Bermudez, 1998). Investigators aiming to determine energy and nutrient sources of older Hispanics modified a questionnaire developed by the National Cancer Institute to add southwestern regional foods such as chile rellenos and tamales (Mayer-Davis et al., 1999). Without these modifications, researchers

would not have learned that chile sauces were major sources of vitamins A and C for Hispanic elderly.

The new group may not interpret item terms/phrases as intended

A measure might include terms and phrases that may be misinterpreted due to unfamiliar language, idioms, or colloquialisms. Terms may be so general as to generate a variety of interpretations, and words may be confusing. There is a greater likelihood of unfamiliar terminology when working with low-literate or non-native English-speaking populations. For example, several items in the Beck Depression Inventory (Beck, Steer, & Garbin, 1988) were misunderstood in a sample of older and low-literacy patients, leading to changes to make the items clearer (Sentell & Ratcliff-Baird, 2003). In a pretest of diabetes knowledge in low-literate Spanish-speaking patients with diabetes, the concept “blood sugar drop” was not understood; similarly in the same study the phrase “bothered by” in the CES-D item “I was bothered by things that usually don’t bother me” was interpreted as having a physical complaint (Rosal, Carbone, & Goins, 2003).

The new group may use different styles of responding

Systematic differences may exist across racial/ethnic groups in the way they respond to survey question items. For example, Blacks and Latinos are more likely than whites to choose the extreme response options rather than middle options of Likert response scales (Bachman & O'Malley, 1984; Hui & Triandis, 1989; Warnecke et al., 1997). In contrast, Asian Americans appear to be more likely than whites to avoid extremely positive options in surveys of health services (Murray-Garcia, Selby, Schmittiel, Grumbach, & Quesenberry, 2000; Ngo-Metzger, Legedza, & Phillips, 2004; Taira et al., 1997).

Process of responding is complex or difficult

For older adults who might have limited vision, or for those with limited English proficiency or low literacy, question formatting and the response task may impact the validity of a measure. Complex instructions and formatting may work with younger white populations but could be challenging for older minority groups or persons with low-literacy (Mullin, Lohr, Bresnahan, & McNulty, 2000; Rosal et al., 2003). Scientific studies have examined ways to present questions to maximize respondent ability to answer the questions accurately (Stone et al., 2000).

The study context or mode of administration may differ from that in which the original measure was developed

Studies need to be sensitive to the appropriateness of different modes of data collection. For example, pencil and paper administration may pose no difficulties in middle aged populations but might be inappropriate for older populations with low vision where oral administration might work better. The growth in use of technologies to accommodate low-literacy populations such as audio-visual computer-based platforms (Hahn & Cella, 2003; Hahn et al., 2004) as well as other electronic modes of administration (web-based surveys, touch-screen computers, handheld computers, interactive voice response, and automated telephone assessment) has led to a call for extensive testing to assure the comparability of electronic and paper-based measures (Coons et al., 2009).

Basis for Modifications: Source of Information Used to Decide How to Modify

Once a decision is made to modify an existing measure, investigators need information on which to base those modifications. Such information is often available in the same studies

that identified the need for modifications. Three sources of information can be used as a basis for modifications: 1) qualitative research on a concept or measure, 2) literature reviews of the adequacy of a measure, and 3) investigator experience or judgment. We review these three information sources and provide examples of how the information source was used to make modifications.

Qualitative research on concept or measure

The two most common qualitative methods for exploring concepts and measures in diverse populations include cognitive interview pretests (Collins, 2003; Drennan, 2003; A.M. Nápoles-Springer et al., 2006; Rosal et al., 2003; Willis, 2005) and focus groups (Fuller, Edwards, Vorakitphokatorn, & Sermsri, 1993; Hughes & DuMont, 1993; A. M. Nápoles-Springer, Santoyo, Houston, Perez-Stable, & Stewart, 2005; Vogt, King, & King, 2004). In addition to identifying specific problems, transcripts often include dialogue that can be used to help make modifications such as alternative phrasing. Qualitative research on the adequacy of a specific measure in a new population group often identifies items that are not culturally appropriate or relevant. The following are some examples of how qualitative research provided information by which to modify a measure:

- Cognitive interviewing to test several diabetes-related measures for low-literate, Spanish-speaking older adults with diabetes resulted in several modifications (Rosal et al., 2003). Responses to probes regarding the suitability of specific words, clarity of response options, and clarity of instructions provided suggestions for modifications.
- To develop new dimensions of language support and financial support for older Chinese and Korean immigrants, transcripts of the focus group dialogue provided specific ideas and phrasing for writing new items (Wong, Yoo, & Stewart, 2007).
- To explore the concept of cultural sensitivity in health care, Nápoles-Springer and colleagues conducted focus groups with African-Americans, Latinos and non-Latino Whites (A. M. Nápoles-Springer et al., 2005) to explore the meaning of culture and what cultural factors influenced the quality of their medical encounters. Analysis of transcripts identified several dimensions of cultural sensitivity, forming the basis for a new multi-dimensional measure based in part on participants' phrases and comments (Nápoles et al., 2011).

Literature reviews

Literature reviews of the adequacy or appropriateness of measures of particular concepts for racial/ethnic groups can be done by individual investigators or by convening an expert panel (Mutran et al., 2001). For example, literature on measures of park and recreation environments was reviewed for adequacy in studies of how parks and recreation settings contribute to physical activity in low-income communities of color. The review provided suggestions for improving measures, for example, the need to reflect the concerns and preferences of residents in these communities and to “explicitly reflect inequality in the availability and quality of parks and recreation areas” (Floyd, Taylor, & Whitt-Glover, 2009). Another example is from a review of physical activity questionnaires for appropriateness for middle-aged and older minority women (Masse et al., 1998). The authors suggested a number of alterations: clarify physical activity-related words like exercise and physical activity; improve definition of phrases like moderate physical activity; modify instructions about walking to capture walking in multiple contexts; modify unsuitable phrases such as “leisure-time” activities, and; substitute culturally relevant activities such as dancing for items such as playing golf or tennis. Other examples of how literature reviews of measures provided specific information by which to modify measures

include reviews of measures of socioeconomic status for minority aging studies (Rudkin & Markides, 2001), measures of acculturation (Salant & Lauderdale, 2003), and food frequency questionnaires for minority populations (Coates & Monteilh, 1997).

Investigator experience

Investigator experience and knowledge can provide ideas for modifications, based on long-term programs of research with diverse populations. For example, based on their experience conducting research with persons with disabilities, Meyers and Andresen (2000) recommended shortening the recall period in health-related quality of life instruments for patients with disabilities. For instance, changes may occur over very short periods of time in multiple sclerosis patients, and in patients with stroke, recall periods need to be compressed due to short-term memory loss.

Types of Modifications

The types of modifications that can be made range from simple format changes such as improving the contrast or increasing the font size to extensive changes such as adding new subscales or changing item wording. To facilitate thinking about the various types of modifications, we have classified them into three broad categories: content, context, and format. For each of these, in Table 1, we define the various specific modifications and provide examples.

Content modifications can be made at the level of dimensions, item stems, or response options, all of which can be added, dropped, modified, or replaced. Adding dimensions or items may be indicated when additional components are found to be needed. Dropping dimensions or items might be done when either is found to be unsuitable for a particular group. Replacing items might be done when an item is unsuitable and a comparable alternative has been suggested by respondents during cognitive interviews.

Context modifications are those made primarily because of study-specific differences, for example, due to a different referent such as nurses instead of doctors. This can include changes to instructions for a self-administered measure to verbal or web-based administration. For example, to modify the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey for an American Indian health service setting, items were modified to substitute the term “doctor or nurse” for “doctor,” and to substitute “health professional” for “health provider” (Weidmer-Ocampo et al., 2009).

Format and presentation modifications include changes in appearance or the way of responding to reduce respondent burden or enhance readability. For older adults this might include the use of audio or touch screens and the addition of response aids; it might also include simplifying instructions and increasing the font size. Mullin and colleagues (2000) have summarized “state-of-the-art” formatting methods for self-administered questionnaires that are known to improve data quality, based on an extensive literature review. They describe a variety of formatting and presentation ideas that can reduce errors in responding, reduce burden, and enhance motivation to respondents to complete questionnaires. For example, a study focused on improving self-reported questionnaires for older people with multiple sclerosis, inconsistent formatting was distracting, thus questionnaires were reformatted for consistency (e.g., response choices printed below each question) (Ploughman, Austin, Stefanelli, & Godwin, 2010).

Recommendations for Assessing Modified Measures

Once a measure is modified it is essential that the new measurement be assessed for its reliability and validity. For some minor modifications such as modifying font typeface, it may not be necessary to conduct a formal assessment of the psychometric properties of the instrument. However, for changes such as adding or deleting a number of question items or modifying response categories, investigators should seriously consider more formal assessment of the new measure.

One classification system for determining the magnitude or extensiveness of a modification by Coons and colleagues (2009) is based on defining potential effects of modification on "...the content, meaning, or interpretation of the measure's items and/or scales" (Coons et al., p. 422–423). The system was originally conceived for electronic data collection methods but it is likely equally applicable to paper/pencil or orally administered instruments. The 3-level classification system includes:

- *Minor modifications* that are not expected to change content or meaning. This would include changing from paper and pencil format into a screen text or using touch response on screen instead of circling a response on paper.
- *Moderate modifications* may change the meaning of the items but in small, subtle ways. There are many examples such as splitting a single item into two, changes in item wording, changing the order of item presentation, and changing the mode of administration from self-administration to interactive voice.
- *Substantial modifications* are more extensive and almost always change the content or meaning of the measure. These more aggressive modifications might include dropping items and changes in item wording or response options.

One of the biggest challenges faced by health disparities researchers is the time and cost that is involved in a full-scale psychometric assessment of a measure that was extensively modified. As Coons and colleagues note (Coons et al., 2009), the rigor of equivalence testing probably should vary with the extent of modification. However, regardless of the level of modification, it is judicious to test the new measure for psychometric adequacy as well as equivalence with the original measure. For minor modifications, a small-scale pretest would suffice to assess that the changes are working as expected. For moderate modifications, a more thorough assessment of the psychometric adequacy of the measure or the extent to which its properties are similar to the original measure should be undertaken. For substantial modifications, where there will be little in the way of *prima fascia* evidence for measurement adequacy or equivalence, a full-scale psychometric assessment is probably needed. In many cases it is impractical to conduct a detailed assessment and the investigators must balance the requirements of a time-limited research program and having a modified measure without fully assessing its reliability and validity. We offer some thoughts on this tricky situation from our own and others' experiences.

One approach we recommend is to conduct an in-depth pretest on a small sample to determine if the modifications are "working" prior to going into the field with the main study. While likely not definitive about measurement performance, this kind of information can be invaluable in determining if there are major problems with the modified measure prior to mounting a major study that includes it. Once the investigative team decides to move forward with pretesting the measure, it is possible to assess its psychometric properties. Typically these assessments would include item-scale correlations and internal consistency reliability. For example, Fongwa and colleagues conducted a field test of a modified patient satisfaction questionnaire in a sample of African Americans and whites and reported results of extensive psychometric analysis (Fongwa, Hays, Gutierrez, & Stewart,

2006). Ideally, one would evaluate the adequacy of the original and the modified measure in the new sample (Hays, Hahn, & Marshall, 2002), but this requires administering the original measure and the modified items. To allow for this possibility, we strongly suggest that researchers do not drop any items and instead add new or modified items to the established measure (Aroian, Hough, Templin, & Kaskiri, 2008; Hays et al., 2002); only in this way will it be possible to compare the original and modified measures. For example, Gonzalez and colleagues analyzed the construct validity of their modified Visual Analogue Pain scale in relation to the original Visual Analogue Pain scale for Hispanics recruited from several U.S. communities; the correlation of the modified scale and the original was 0.72, and the modified scale had less missing data (6% compared to 24%) (González, Stewart, Ritter, & Lorig, 1995). There are several other examples of the benefits of comparing the original and modified measures in the same study (Aroian et al., 2008; Kazis et al., 2004; Tucker et al., 1998). However, including both old and new versions in the same study may not be practical and can introduce context effects.

Investigators should also consider how they might assess the validity of the modified measure. One approach is to conduct validity tests to parallel those done with the original measure, thus administration of the same indicators of validity used with the original measure is necessary. The expectation is that the modified measure is an improvement over the original measure in the new context or population (Hays et al., 2002). For example, in a study to adapt a patient satisfaction instrument for lower literacy population, two versions of a 'new' instrument, one with cartoon/pictorial enhancements and one a computer-assisted telephone delivery, the original self-report text instrument and new versions were compared head-to-head in a randomized trial, with each arm using the same validity indicators (Shea et al., 2008). Another approach is to use new indicators of validity to assess the performance of the modified measure. While this approach may produce strong evidence of validity, it makes it difficult to compare the validity assessment with previous validity results using the original measure.

Recommendations for Reporting Results of Modifications

If assessment studies of measure modifications were routinely published, we would gain a tremendous amount of information on how various modifications affect the reliability and validity of measures in new populations, as well as point to new strategies and methods for test and measurement assessment. Currently, reporting appears to be missing, e.g., in a review of measures of spirituality for use in palliative care, the authors noted that changes in the content of instruments after cultural adaptation were poorly reported (Selman, Harding, Gysels, Speck, & Higginson, 2011).

One approach would be to provide details of the modification and its assessment in a separate methods paper. Investigators reported on an adaptation of the CAHPS Clinician and Group Survey for use in the American Indian Health Service, including methods for deciding to modify the measure, the types of modifications, and the psychometric characteristics (Weidmer-Ocampo et al., 2009). Similarly, in reporting modifications to the CHAMPS Physical Activity Questionnaire (A.L. Stewart et al., 2001) for a church-based lifestyle intervention for African Americans, Resnicow and colleagues (Resnicow et al., 2003) clearly described the entire process of modifications, including tests of validity. Other examples include a study adapting a patient satisfaction instrument for low literacy individuals by modifying the format (Shea et al., 2005) and one adapting a patient satisfaction survey for low literacy VA patients by adding illustrations (Weiner et al., 2004).

Another approach is to report details of the modification and assessment process within the methods section of a substantive paper. For example, Nápoles and colleagues (Napoles,

Ortiz, O'Brien, Sereno, & Kaplan, 2011) reported their modifications of the Cancer Behavior Inventory in the methods section of a paper on coping resources and self-rated health among Latina breast cancer survivors, including results of psychometric tests of the modified measure.

When publishing papers that include modified measures, we recommend at a minimum reporting: 1) features of the original measure that required modification; 2) source of information on the basis for modifications; 3) specific type of modification made; and 4) how the modified measure was tested for psychometric adequacy and results. It might also be necessary to report on whether permission was obtained to modify it or permission was granted on measure or website. Incorporating this information adds to accumulated knowledge on the particular measure as applied in diverse groups, and helps our understanding of broader issues involved in adapting measures for diverse populations. These reporting guidelines may not always be practical, though, given space limitations in published papers and typically brief treatment of measurement issues. Especially when this information would be disproportional to the substantive portion of a paper, we encourage taking advantage of options to provide supplemental methodological information on the websites of journal. This is becoming more common in some journals that discourage extensive methodological information in articles (Tobin, 2000).

Discussion

We have addressed a gap in the literature pertaining to use of measures in health disparities and minority aging studies that are appropriate for the population group. There are virtually no published papers that discuss methodological issues in modifying measures. Our paper becomes a first step toward understanding the issues. As the field of measurement in diverse populations evolves, we will increasingly find more information on such modifications. It was very difficult to find examples in the literature on modifications to existing measures for use in a diverse population group. Modifications were typically described in a brief paragraph in the measures section, but it was not possible to find these modifications through keyword searches. Thus, by calling attention to the importance of providing details, we hope to promote better reporting in publications that use modified measures.

Many modifications are designed to improve the ability of the measure to answer specific, contemporary research questions. Thus, often modifications are made to update terminology or to reflect historical changes (Krause, 2006). In this case, modifications are an integral part of the evolution of a concept or measure, with modifications improving an existing measure for use by other researchers. The SF-36 is a good example of evolution in measures of health-related quality of life. As noted by Ware (2000), the SF-36 version 2 includes several modifications to the original SF-36. The original SF-36 in turn was a modification or subset of several longer-form Medical Outcomes Study (MOS) measures of "functioning and well-being." The MOS measures in turn were adapted and developed based on measures of health status developed for the Health Insurance Experiment (HIE) (Stewart & Ware, 1992) and HIE measures were based on even earlier measures such as the General Psychological Well-Being Inventory (Dupuy et al., 1984). Thus, the SF-36 version 2 reflects numerous iterations of prior measurement instruments, each involving modifications and adaptations of earlier versions to improve the measures.

Other examples of how measures evolve can illustrate this process. One is the adaptation of the Mini-Mental State examination by Teng and Chui (Teng & Chui, 1987) to sample a broader variety of cognitive functions, include a broader range of difficulty levels, and improve reliability and validity. Chatters and colleagues reviewed the evolution of concepts and measures of religiosity for research on the role of religiosity in the lives of older African

Americans (Chatters, Taylor, & Lincoln, 2001). They noted a progression from single measures of church attendance to multi-dimensional concepts and measures including prayer, private religious practices, daily spiritual experiences, and organizational religiousness. As the concept evolved through research on how spirituality affects health and well-being, the measures also evolved to reflect the changing concepts.

Consistent with the concept of measurement evolution is the perspective that the validity of the measure is not a property of a test or measure, but of a measure tested under a particular set of conditions (Messick, 1995; Sechrest, 2005). This perspective holds that because validity pertains to understanding the meaning of scores, construct validity can only be established incrementally based on the accumulation of evidence on how the measure relates to other measures (Sechrest, 2005, p. 1596). Well-designed modifications thus contribute to enhancing the validity of measures in new and diverse populations.

The scientific study of the adequacy of measures in diverse populations is still emerging, but research promoting understanding of how to modify measures and test those modifications is even newer. By laying out some of the issues in understanding how to modify and test measures and by describing some possible solutions, we hope to guide future research efforts to advance the field of measurement in diverse populations. By highlighting the value of modifications in the ongoing evolution of concepts and measures in studies of minority aging, we also hope to encourage authors of published measures to be flexible in imposing limitations on use of their measures.

References

- Aaronson NK, Acquadro C, Alonso J, Apolone G, Bucquet D, Bullinger M, et al. International Quality of Life Assessment (IQOLA) Project. *Quality of Life Research*. 1992; 1:349–351. [PubMed: 1299467]
- Aroian KJ, Hough ES, Templin TN, Kaskiri EA. Development and psychometric evaluation of an Arab version of the Family Peer Relationship Questionnaire. *Research in Nursing and Health*. 2008; 31:402–416. [PubMed: 18389498]
- Bachman, Jerald G.; O'Malley, Patrick M. Yea-saying, nay-saying, and going to extremes: Black-white differences in response style. *Public Opinion Quarterly*. 1984; 48:491–509.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000; 25:3186–3191. [PubMed: 11124735]
- Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*. 1988; 8:77–100.
- Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Tracy RP. Multi-ethnic study of atherosclerosis: Objectives and design. *American Journal of Epidemiology*. 2002; 156:871–881. [PubMed: 12397006]
- Bird ST, Bogart LM. Perceived race-based and socioeconomic status(SES)-based discrimination in interactions with health care providers. *Ethnicity and Disease*. 2001; 11:554–563. [PubMed: 11572421]
- Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S, Ware JE Jr. Translating health status questionnaires and evaluating their quality: The IQOLA Project approach. *Journal of Clinical Epidemiology*. 1998; 51:913–923. [PubMed: 9817108]
- Chapleski EE, Lamphere JK, Kaczynski R, Lichtenberg PA, Dwyer JW. Structure of a depression measure among American Indian elders: confirmatory factor analysis of the CES-D scale. *Research on Aging*. 1997; 19:462–485.
- Chatters LM, Taylor RJ, Lincoln KD. Advances in the measurement of religiosity among older African Americans: implications for health and mental health researchers. *Journal of Mental Health and Aging*. 2001; 7:180–200.
- Coates RJ, Monteilh CP. Assessments of food-frequency questionnaires in minority populations. *American Journal of Clinical Nutrition*. 1997; 65:1108S–1115S. [PubMed: 9094906]

- Collins D. Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*. 2003; 12:229–238. [PubMed: 12769135]
- Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, Basch E. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value Health*. 2009; 12:419–429. [PubMed: 19900250]
- Dixon JS, Bird HA. Reproducibility along a 10 cm vertical visual analogue scale. *Annals of the Rheumatic Diseases*. 1981; 40:87–89. [PubMed: 7469530]
- Drennan J. Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing*. 2003; 42:57–63. [PubMed: 12641812]
- Dupuy, HJ. The psychological general well-being (PGWB) index. In: Wenger, NK.; Mattson, MD.; Furberg, CD.; Elinson, J., editors. *Assessment of quality of life in clinical trials of cardiovascular therapies*. New York: Le Jacq Publishing, Inc; 1984. p. 170-183.
- Floyd MF, Taylor WC, Whitt-Glover M. Measurement of park and recreation environments that support physical activity in low-income communities of color: Highlights of challenges and recommendations. *American Journal of Preventive Medicine*. 2009; 36:S156–S160. [PubMed: 19285207]
- Fongwa MN, Hays RD, Gutierrez PR, Stewart AL. Psychometric characteristics of a patient satisfaction instrument tailored to the concerns of African Americans. *Ethnicity and Disease*. 2006; 16:948–955. [PubMed: 17061752]
- Fuller, TD.; Edwards, JN.; Vorakitphokatorn, S.; Sermisri, S., editors. *Using focus groups to adapt survey instruments to new populations*. Newbury Park, CA: Sage Publications, Inc; 1993.
- González VM, Stewart A, Ritter PL, Lorig K. Translation and validation of arthritis outcome measures into Spanish. *Arthritis and Rheumatism*. 1995; 38:1429–1446. [PubMed: 7575693]
- Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*. 2006; 44:S78–S94. [PubMed: 17060839]
- Hahn EA, Cella D. Health outcomes assessment in vulnerable populations: Measurement challenges and recommendations. *Archives of Physical Medicine and Rehabilitation*. 2003; 84:S35–S42. [PubMed: 12692770]
- Hahn EA, Cella D, Dobrez D, Shiomoto G, Marcus E, Taylor SG, Webster K. The talking touchscreen: A new approach to outcomes assessment in low literacy. *Psycho-Oncology*. 2004; 13:86–95. [PubMed: 14872527]
- Hausmann LR, Kressin NR, Hanusa BH, Ibrahim SA. Perceived racial discrimination in health care and its association with patients' healthcare experiences: Does the measure matter? *Ethnicity and Disease*. 2010; 20:40–47. [PubMed: 20178181]
- Hays RD, Hahn H, Marshall G. Use of the SF-36 and other health-related quality of life measures to assess persons with disabilities. *Archives of Physical Medicine and Rehabilitation*. 2002; 83:S4–S9. [PubMed: 12474166]
- Hughes D, DuMont K. Using focus groups to facilitate culturally anchored research. *American Journal of Community Psychology*. 1993; 21:775–806.
- Hui CH, Triandis HC. Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*. 1989; 20:296–309.
- Jackson, JS. Concepts and measures in the National Survey of Black Americans. In: Jones, RL., editor. *Handbook of tests and measurements for black populations*. Vol. 2. Hampton, VA: Cobb & Henry Publishers; 1996. p. 535-540.
- Johnson TP. Methods and frameworks for crosscultural measurement. *Medical Care*. 2006; 44:S17–S20. [PubMed: 17060823]
- Juniper EF. Validated questionnaires should not be modified. *European Respiratory Journal*. 2009; 34:1015–1017. [PubMed: 19880615]
- Kazis LE, Miller DR, Clark JA, Skinner KM, Lee A, Ren XS, Ware JE Jr. Improving the response choices on the veterans SF-36 health survey role functioning scales: Results from the Veterans Health Study. *Journal of Ambulatory Care Management*. 2004; 27:263–280. [PubMed: 15287216]

- Kazis LE, Ren XS, Lee A, Skinner K, Rogers W, Clark J, Miller DR. Health status in VA patients: Results from the Veterans Health Study. *American Journal of Medical Quality*. 1999; 14:28–38. [PubMed: 10446661]
- Krause N. The use of qualitative methods to improve quantitative measures of health-related constructs. *Medical Care*. 2006; 44:S34–S38. [PubMed: 17060833]
- Krause N, Markides K. Measuring social support among older adults. *International Journal of Aging and Human Development*. 1990; 30:37–53. [PubMed: 2407665]
- Martinez NC, Sousa VD. Cross-cultural validation and psychometric evaluation of the Spanish Brief Religious Coping Scale (S-BRCS). *Journal of Transcultural Nursing*. 2011; 22:248–256. [PubMed: 21505097]
- Masse LC, Ainsworth BE, Tortolero S, Levin S, Fulton JE, Henderson KA, Mayo K. Measuring physical activity in midlife, older, and minority women: Issues from an expert panel. *Journal of Womens Health*. 1998; 7:57–67.
- Mattson-Prince J. A rational approach to long-term care: Comparing the independent living model with agency-based care for persons with high spinal cord injuries. *Spinal Cord*. 1997; 35:326–331. [PubMed: 9160459]
- Mayer-Davis EJ, Vitolins MZ, Carmichael SL, Hemphill S, Tsaroucha G, Rushing J, Levin S. Validity and reproducibility of a food frequency interview in a multi-cultural epidemiology study. *Annals of Epidemiology*. 1999; 9:314–324. [PubMed: 10976858]
- McDowell, IY.; Newell, C. *Measuring health: a guide to rating scales and questionnaires*. 3rd ed.. New York: Oxford University Press; 2006.
- Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. 1995; 50:741–749.
- Meyers AR, Andresen EM. Enabling our instruments: Accommodation, universal design, and access to participation in research. *Archives of Physical Medicine and Rehabilitation*. 2000; 81:S5–S9. [PubMed: 11128904]
- Moody-Ayers SY, Stewart AL, Covinsky KE, Inouye SK. Prevalence and correlates of perceived societal racism in older African American adults with type 2 diabetes mellitus. *Journal of the American Geriatrics Society*. 2005; 53:2202–2208. [PubMed: 16398910]
- Mui AC, Burnette D, Chen LM. Cross-cultural assessment of geriatric depression: a review of the CES-D and the GDS. *Journal of Mental Health and Aging*. 2001; 7:137–164.
- Mullin PA, Lohr KN, Bresnahan BW, McNulty P. Applying cognitive design principles to formatting HRQOL instruments. *Quality of Life Research*. 2000; 9:13–27. [PubMed: 10981203]
- Murray-Garcia JL, Selby JV, Schmittiel J, Grumbach K, Quesenberry CP Jr. Racial and ethnic differences in a patient survey: Patients' values, ratings, and reports regarding physician primary care performance in a large health maintenance organization. *Medical Care*. 2000; 38:300–310. [PubMed: 10718355]
- Mutran EJ, Reed PS, Sudha S. Social support: clarifying the construct with applications for minority populations. *Journal of Mental Health and Aging*. 2001; 7:67–78.
- Nápoles-Springer AM, Santoyo J, Houston K, Perez-Stable EJ, Stewart AL. Patients' perceptions of cultural factors affecting the quality of their medical encounters. *Health Expectations*. 2005; 8:4–17. [PubMed: 15713166]
- Nápoles-Springer AM, Santoyo-Olsson J, O'Brien H, Stewart AL. Using cognitive interviews to develop surveys in diverse populations. *Medical Care*. 2006; 44:S21–S30. [PubMed: 17060830]
- Nápoles-Springer AM, Stewart AL. Overview of qualitative methods in research with diverse populations: Making research reflect the population. *Medical Care*. 2006; 44(Suppl. 3):S5–S9. [PubMed: 17060835]
- Napoles AM, Ortiz C, O'Brien H, Sereno AB, Kaplan CP. Coping resources and self-rated health among latina breast cancer survivors. *Oncology Nursing Forum*. 2011; 38:523–531. [PubMed: 21875840]
- Nápoles AM, Santoyo-Olsson J, Farren G, Olmstead J, Cabral R, Ross B, Stewart AL. The patient-reported Clinicians' Cultural Sensitivity Survey: A field test among older Latino primary care patients. *Health Expectations*. 2011
- National Institute on Aging. *Health Disparities Strategic Plan: Fiscal Years 2009–2013*. 2011

- Ngo-Metzger Q, Legedza AT, Phillips RS. Asian Americans' reports of their health care experiences. Results of a national survey. *Journal of General Internal Medicine*. 2004; 19:111–119. [PubMed: 15009790]
- Nunnally, JC.; Bernstein, IH. *Psychometric Theory*, Third Edition. New York: McGraw-Hill, Inc; 1994.
- Ploughman M, Austin M, Stefanelli M, Godwin M. Applying cognitive debriefing to pre-test patient-reported outcomes in older people with multiple sclerosis. *Quality of Life Research*. 2010; 19:483–487. [PubMed: 20151208]
- Resnicow K, McCarty F, Blissett D, Wang T, Heitzler C, Lee RE. Validity of a modified CHAMPS physical activity questionnaire among African-Americans. *Medicine and Science in Sports and Exercise*. 2003; 35:1537–1545. [PubMed: 12972874]
- Rosal MC, Carbone ET, Goins KV. Use of cognitive interviewing to adapt measurement instruments for low-literate Hispanics. *Diabetes Educator*. 2003; 29:1006–1017. [PubMed: 14692374]
- Rudkin L, Markides KS. Measuring the socioeconomic status of elderly people in health studies with special focus on minority elderly. *Journal of Mental Health and Aging*. 2001; 7:53–66.
- Salant T, Lauderdale DS. Measuring culture: A critical review of acculturation and health in Asian immigrant populations. *Social Science and Medicine*. 2003; 57:71–90. [PubMed: 12753817]
- Santoyo-Olsson J, Cabrera J, Freyre R, Grossman M, Alvarez N, Mathur D, Stewart AL. An innovative multiphased strategy to recruit underserved adults into a randomized trial of a community-based diabetes risk reduction program. *Gerontologist*. 2011; 51(Suppl 1):S82–S93. [PubMed: 21565823]
- Sechrest L. Validity of measures is no simple matter. *Health Services Research*. 2005; 40:1584–1604. [PubMed: 16178997]
- Selman L, Harding R, Gysels M, Speck P, Higginson IJ. The measurement of spirituality in palliative care and the content of tools validated cross-culturally: A systematic review. *Journal of Pain and Symptom Management*. 2011
- Sentell TL, Ratcliff-Baird B. Literacy and comprehension of Beck Depression Inventory response alternatives. *Community Mental Health Journal*. 2003; 39:323–331. [PubMed: 12908646]
- Shea JA, Aguirre AC, Sabatini J, Weiner J, Schaffer M, Asch DA. Developing an illustrated version of the Consumer Assessment of Health Plans (CAHPS). *Joint Commission Journal on Quality and Patient Safety*. 2005; 31:32–42. [PubMed: 15691208]
- Shea JA, Guerra CE, Weiner J, Aguirre AC, Ravenell KL, Asch DA. Adapting a patient satisfaction instrument for low literate and Spanish-speaking populations: comparison of three formats. *Patient Education and Counseling*. 2008; 73:132–140. [PubMed: 18486414]
- Skinner JH. Acculturation: measures of ethnic accommodation to the dominant American culture. *Journal of Mental Health and Aging*. 2001; 7:41–52.
- Smedley, BD.; Stith, AY.; Nelson, AR., editors. *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: The National Academies Press; 2003.
- Stahl SM, Hahn AA. The National Institute on Aging's Resource Centers for Minority Aging Research. Contributions to measurement in research on ethnically and racially diverse populations. *Medical Care*. 2006; 44:S1–S2. [PubMed: 17060815]
- Stewart AL, Napoles-Springer AM, Gregorich SE, Santoyo-Olsson J. Interpersonal processes of care survey: patient-reported measures for diverse groups. *Health Services Research*. 2007; 42:1235–1256. [PubMed: 17489912]
- Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL. CHAMPS physical activity questionnaire for older adults: outcome for interventions. *Medicine and Science in Sports and Exercise*. 2001; 33:1126–1141. [PubMed: 11445760]
- Stewart AL, Nápoles-Springer A. Health-related quality of life assessments in diverse population groups in the United States. *Medical Care*. 2000; 38:II102–II124. [PubMed: 10982096]
- Stewart AL, Nápoles-Springer AM. Advancing health disparities research: Can we afford to ignore measurement issues? *Medical Care*. 2003; 41:1207–1220. [PubMed: 14583684]
- Stewart, AL.; Ware, JE., Jr, editors. *Measuring functioning and well-being: The Medical Outcomes Study approach*. Durham, NC: Duke University Press; 1992.

- Stone, AA.; Turkkan, JS.; Bachrach, CA.; Jobe, JB.; Kurtzman, HS.; Cain, VS. The science of self-report: Implications for research and practice. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers; 2000.
- Taira DA, Safran DG, Seto TB, Rogers WH, Kosinski M, Ware JE, Tarlov AR. Asian-American patient ratings of physician primary care performance. *Journal of General Internal Medicine*. 1997; 12:237–242. [PubMed: 9127228]
- Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *Journal of Clinical Psychiatry*. 1987; 48:314–318. [PubMed: 3611032]
- Teresi JA, Stewart AL, Morales L, Stahl S. Measurement in a multi-ethnic society: Overview to the special issue. *Medical Care*. 2006; 44(Suppl. 3):S2–S3.
- Teresi JA, Stewart AL, Stahl SM. Fifteen years of progress in measurement and methods at the resource centers for minority aging research. *Journal of Aging and Health*. 2012 Sep; 24(6):985–991. [PubMed: 22904184]
- Thiamwong L, Stewart AL, Warahut J. Development, reliability and validity of the Thai Healthy Aging Survey. *Walailak Journal of Science and Technology*. 2009; 6:167–188.
- Tobin MJ. Authors, authors, authors - follow instructions or expect delay. *American Journal of Respiratory and Critical Care Medicine*. 2000; 162:1193–1194. [PubMed: 11029316]
- Tucker KL, Bianchi LA, Maras J, Bermudez OI. Adaptation of a food frequency questionnaire to assess diets of Puerto Rican and non-Hispanic adults. *American Journal of Epidemiology*. 1998; 148:507–518. [PubMed: 9737563]
- U.S. Department of Health and Human Services. *Healthy People 2010 (Conference Edition, in Two Volumes)*. Washington, DC: 2000.
- Vanderplas JH, Vanderplas JM. Effects of legibility on verbal test performance of older adults. *Perceptual and Motor Skills*. 1981; 53:183–186. [PubMed: 7290867]
- Vogt DS, King DW, King LA. Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment*. 2004; 16:231–243. [PubMed: 15456379]
- Walter MJ, Castro M, Kunselman SJ, Chinchilli VM, Reno M, Ramkumar TP, Wechsler ME. Predicting worsening asthma control following the common cold. *European Respiratory Journal*. 2008; 32:1548–1554. [PubMed: 18768579]
- Ware JE Jr. SF-36 health survey update. *Spine*. 2000; 25:3130–3139. [PubMed: 11124729]
- Warnecke RB, Johnson TP, Chaavez N, Sudman S, O'Rourke DP, Lacey L, Horm J. Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology*. 1997; 7:334–342. [PubMed: 9250628]
- Weidmer-Ocampo B, Johansson P, Dalpoas D, Wharton D, Darby C, Hays RD. Adapting CAHPS for an American Indian population. *Journal of Health Care for the Poor and Underserved*. 2009; 20:695–712. [PubMed: 19648698]
- Weiner J, Aguirre A, Ravenell K, Kovath K, McDevit L, Murphy J, Shea JA. Designing an illustrated patient satisfaction instrument for low-literacy populations. *The American Journal of Managed Care*. 2004; 10:853–860. [PubMed: 15609739]
- Willis, GB. *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications; 2005.
- Wong ST, Yoo GJ, Stewart AL. Examining the types of social support and the actual sources of support in older Chinese and Korean immigrants. *International Journal of Aging and Human Development*. 2005; 61:105–121. [PubMed: 16161288]
- Wong ST, Yoo GJ, Stewart AL. An empirical evaluation of social support and psychological well-being in older Chinese and Korean immigrants. *Ethnicity & Health*. 2007; 12:43–67. [PubMed: 17132584]

Table 1

Organizing Framework of Types of Possible Modifications

CATEGORY/ Type of Modification	Definition	Examples
CONTENT MODIFICATIONS		
<u>Dimensions</u>		
Add dimension	Add a set of items hypothesized to measure a new dimension (subscale)	Two dimensions were added to an existing social support measure: "language support" and "financial support" (Wong et al., 2007) in response to qualitative research that these were important to older Chinese and Korean immigrants (Wong et al., 2005). Items were written based on the same qualitative studies. A new "discrimination/stereotyping" dimension was added to an existing patient satisfaction measure in response to qualitative studies indicating its importance to older African Americans (Fongwa et al., 2006). An "integration" dimension (providing various types of support to others in their social environment) was added to a social support measure to make it more appropriate for older adults (Krause & Markides, 1990).
Drop dimension	Omit all items from one dimension (subscale)	The SF-36 physical functioning scale was omitted in a study of spinal cord injury patients because patients considered the items demeaning, given that walking and climbing were prominent items (Mattson-Prince, 1997).
<u>Item Stems</u>		
Drop items	Drop items from an existing scale	Four items from the Cancer Behavior Inventory referring to treatment were dropped in a study of Latina breast cancer survivors who had completed treatment (Napoles et al., 2011). A few items from the McNeilly Perceived Racism Scale pertaining to racism in academic settings were dropped as not applicable to older African American patients (Moody-Ayers, Stewart, Covinsky, & Inouye, 2005). Several items from the CES-D were dropped as being inappropriate for older Asian groups (reported by Mui et al., 2001, page 14, p.141 (Mui et al., 2001). Eight items were omitted from the CES-D in a study of older American Indians because they were determined to be inappropriate (Chapleski, Lamphere, Kaczynski, Lichtenberg, & Dwyer, 1997).
Add items	Add items to an existing scale to fill in a missing element	The Insulin Resistance Atherosclerosis Study food frequency questionnaire was modified to add foods typically eaten by Chinese persons for a multi-ethnic study of atherosclerosis (Bild et al., 2002). The CHAMPS Physical Activity Questionnaire was modified for a church-based intervention for African Americans. Based on focus groups and pilot testing, "praise dancing" was added because it was common and contributed a substantial amount of energy expenditure (Resnicow et al., 2003).
Modify items	Substitute new term, modify wording without changing meaning, split complex item into two	Item stems in the Family Peer Relationship Questionnaire were modified to be appropriate for Arab Muslim women, based on information from five cultural informants (Arab immigrant women from the Middle East) with experience providing health and social services to Arab Muslim families (Aroian et al., 2008). Hays and colleagues (Hays et al., 2002) suggest modifying some SF-36 items for studies of patients with disabilities by changing "does your health now limit you in these activities" to "does your health or a disability now limit you in these activities." Several items in a patient satisfaction survey were reworded or split into two items based on problems identified with original items in qualitative research with older African Americans (Fongwa et al., 2006). To adapt CES-D items for verbal administration to low-literate respondents, based on cognitive pretesting, statements were modified to be questions and a time frame was added to each item (from only being in the initial instructions). For example, "I felt depressed" was modified to "During the past week, how many days did you feel depressed?" (Rosal et al., 2003).
<u>Response Options</u>		

CATEGORY/ Type of Modification	Definition	Examples
Replace response option set	Replace an entire set of response options with another.	To adapt a measure for older Thai adults, Thiamwong and colleagues (Thiamwong et al., 2009) replaced a bi-directional true/false response scale (1-definitely true to 5-definitely false) with a unidirectional response scale (1-not at all true to 5-definitely true). To improve variability of two role functioning scales from the SF-36 in the Veterans Health Study, the yes/no response options were replaced with a 5-level ordinal scale (ranging from “yes, all of the time to “no, none of the time”) (Kazis et al., 1999).
Modify response options	Split response options into two levels to reduce the number of options presented at any one time Add verbal labels to all response options rather than having endpoints only labeled Drop a response option, e.g., to reduce confusion Remove numbering from response scales	For a study of low-literate individuals, a 4-point Likert scale response option ranging from 4=very important to 1=not important was modified to be administered in a nested format. A 2-point scale was first administered (important, not important); for those responding “important,” 3 options were offered to indicate the strength of the choice (very important, important, somewhat important) (Rosal et al., 2003). For a study of low-literate individuals, response options asking how confident they are they can do something with endpoints anchored (1=not at all, 4=completely) were modified to provide a verbal anchor for each level (1=not confident, 2=a little confident, 3=somewhat confident, 4=completely confident) (Rosal et al., 2003). The SF-36 Version 1 had 6 response options for the mental health and vitality items (1=all of the time, 2=most of the time, 3=a good bit of the time, 4=some of the time, 5=a little of the time, 6=none of the time). In version 2, “a good bit of the time” was dropped because respondents consistently had difficulty understanding its meaning, leaving five options (McDowell & Newell, 2006; Ware, 2000). In adapting the Spanish Brief Religious Coping scale, the original items were asked on a 4-point response scale: 0=not at all to 3=a great deal. Because of difficulty with Likert scales, response options were modified to omit the numbers (leaving the phrasing: not at all, somewhat, quite a bit, a great deal) (Martinez & Sousa, 2011).
STUDY SPECIFIC CONTEXT MODIFICATIONS		
Modify recall period	Different time frame applies	The CHAMPS Physical Activity Questionnaire was modified for an intervention for African Americans; the one month recall period for activities was changed to 2 weeks, based on formative research indicating that respondents could more accurately recall activity during a shorter period (Resnicow et al., 2003).
Change referent	Change referent in instructions or item stems	To adapt the Everyday Discrimination Scale for use in a health care setting, instructions asking whether they “ever had any of the following things happen...” were modified to “when getting health care, have you ever had...” (Bird & Bogart, 2001). An item in the Experiences of Racial Discrimination scale referring to discrimination “while getting medical care” was modified to refer to getting care in a Veteran’s Administration (VA) healthcare setting (Hausmann, Kressin, Hanusa, & Ibrahim, 2010).
FORMAT OR PRESENTATION MODIFICATIONS		
Modify format for responding	Modify way in which respondents answer questions, e.g., check box instead of circle number	Instead of asking respondents to circle the number in front of each response, for each response choice, they were asked to check a box (Walter et al., 2008).
Modify appearance of page	Make skip patterns more obvious, reduce crowding, increase font, improve contrast	Because reading performance declines for older adults and persons with impaired vision when questionnaires use small fonts, questionnaires were designed using 14-point font (Vanderplas & Vanderplas, 1981).
Add visual aids for response options	Use a card with the response options to aid remembering the options Add visual aid to illustrate increasing intensity	An interviewer-administered survey in a study of lower-income, minority adults used a large card containing response options for each set of questions. Interviewers showed the appropriate card to respondents to see which aided memory of the options and allowed respondents to point to a response (Santoyo-Olsson et al., 2011). A Visual Analogue Pain Scale for Spanish-speaking individuals was modified from a 10cm line with endpoints labeled (Dixon & Bird, 1981) to a histogram; bars increased in size from 0 to 10 visually depicted increasing pain intensity and a number appeared below each bar (González et al., 1995).