2. CROSS-CULTURAL ADAPTATION OF SURVEY INSTRUMENTS: THE CAHPS®

EXPERIENCE

Background

Collecting accurate health data on the growing number of ethnic minorities in the United States has increased in policy relevance in recent years. Today, most general-population sample surveys require translation into at least one language (usually Spanish), and often other languages as well. However, cross-cultural research is threatened by the failure to produce culturally and linguistically appropriate survey instruments for minority populations. Guillemin, Bombardier and Beaton consider that cross-cultural adaptation of instruments is a "prerequisite for the investigation of cross-cultural differences" (1993, p.1425). A survey conducted with an inadequate instrument may lead to erroneous conclusions that are difficult to detect during analyses. Conclusions drawn from such research may be mistakenly attributed to differences between the source and target populations. These risks, and the increasing importance of cross-cultural research, have led to a re-examination of the prevalent techniques for developing survey instruments that will be used in different languages and for assessing the cultural appropriateness of survey instruments that are utilized for this type of research.

In this paper we define culturally appropriate translated survey instruments as conceptually and technically equivalent to the source language, culturally competent, and linguistically appropriate for the target population. This paper provides recommendations for the cross-cultural adaptation of survey instruments and illustrates this with
examples of what is being done in the Consumer Assessment of Health Plans Study (CAHPS®).

The CAHPS® Surveys

CAHPS® is a 5-year initiative that aims to produce a set of standardized survey instruments that can be used to collect reliable information from health plan enrollees about the care they have received. CAHPS® items include both evaluations (ratings) and reports of specific experiences with health plans. CAHPS® surveys are constructed from two pools of items: "core" items that apply across the spectrum of health plan enrollees and supplemental items that are used in conjunction with the core to address issues pertinent to specific populations, such as Medicaid fee-for-service and Medicare managed care. The results of these surveys are then used to prepare reports that provide information to consumers who are trying to select a health plan.

CAHPS® recognizes the need to translate its instruments into several languages in order for its users to adequately collect data on its consumers. The CAHPS® survey instruments were translated into Spanish because it is the second most widely used language in the U.S. (Weidmer, Brown, and Garcia, 1999). As CAHPS® has expanded, several states and users have expressed the need to translate the CAHPS® instruments into other languages as well. The principal goal of the translation process of the CAHPS® surveys and protocols is to produce instruments that are culturally appropriate for the different groups in the selected languages. The main challenge is to produce such
instruments while maintaining equivalency with the English-language version.

Cultural Adaptation of Survey Instruments

Guillemin et al. have described the process of cross-cultural adaptation of surveys as "oriented towards measuring a similar phenomenon in different cultures; it is essentially the production of an equivalent instrument adapted to another culture" (1993, p. 1425). We define culturally appropriate translated survey instruments as conceptually and technically equivalent to the source language, culturally competent, and linguistically appropriate for the target population.

In translating, it is important to distinguish between technical and conceptual equivalence. Technical equivalence refers to equivalence in grammar and syntax, while conceptual equivalence refers to the absence of differences in meaning and content between two versions of an instrument. A technically equivalent instrument is a literal translation using the "equivalent denotative meaning" of the words in the original survey. However, different terms may have a different "connotative" or implied meaning in different cultures, requiring an assessment of conceptual equivalence in the translation of instruments (Marin and Marin, 1991).

Conceptual equivalence includes item and scalar equivalence of the source and translated surveys. Item equivalence signifies that each item has the same meaning for subjects in the target culture. Scalar equivalence is achieved when the construct is measured on the same metric in two cultures (Hui and Triandis, 1985). Health surveys
generally use categorical rating scales where response choices are ordered along a hypothesized response continuum (e.g., excellent to poor). It is important to determine if there is equivalence in the distances between the response choices in the two cultures (Keller et al., 1998).¹

Cultural competence refers to the requirement that the translated instrument adequately reflect the cultural assumptions, norms, values, and expectations of the target population (Marin and Marin, 1991). Cross-cultural researchers differentiate between universal or common meaning across cultures ("etic") and group-specific ("emic") constructs or ideas. The source survey reflects the assumptions and values of the researcher’s culture and in translating surveys, it is generally assumed that the constructs of the source survey are etic. Translated surveys should include both etic and emic items in order to reflect properly the reality being studied. This implies the development of new items that reflect the emic aspects of a concept in the target culture (Brislin, 1986).

Linguistic appropriateness refers to the language readability and comprehension of the translated instrument. The goal is to develop instruments using wording at a level easily understood by the majority of potential respondents. An instrument developed in the source language at an eighth grade reading level does not automatically preserve the same reading and comprehension level upon translation, and may actually increase considerably. The problem of equivalence in

¹For a discussion of the Thurstone scaling exercise applied to the SF-36 see Keller et al. (1998).
reading level is further compounded if the target population is at a lower average reading level than the source language population.

In order to cross-culturally adapt survey instruments, we propose a framework (Figure 1) that comprises the following activities:

- Translation (steps 1 to 4)
- Qualitative analysis (step 5)
- Field test and analyses (step 6)

Based on the results of the field test, additional qualitative analysis may be necessary. The International Quality of Life Assessment (IQOLA) project group has used a similar protocol in translating the SF-36 Health Survey into different languages (Bullinger et al., 1998; Gandek and Ware, 1998).

**Translation (Steps 1 to 4)**

Most researchers today agree that it is no longer acceptable to use a direct-translation technique (or one-way-translation) for translating survey instruments. A review of the literature indicates that the most accepted approach to translation is one in which a variety of techniques are used to ensure the reliability and validity of the translated survey instrument (Brislin, 1986; Bullinger et al., 1998; Marin and Marin, 1991). The rationale behind this approach is that no single technique adequately demonstrates and improves the equivalence of an instrument, and that only a multi strategy approach that provides and evaluates different types of equivalence can produce an adequate translation. We recommend a process for translating surveys
that includes translation, back-translation, independent review, and review by committee.

**Forward-translation**

Professional translators (two or more) experienced in translating similar survey instruments, preferably native speakers of the target language, are retained to translate the survey instrument. The translators used for this task should have familiarity with the target population and with data collection procedures. Before starting the translation, the translators should be briefed on the objectives of the study, the demographic characteristics of the sample, the interviewing mode to be used, and the targeted reading level of the translation.

**Back-translation**

Once the instruments are translated they go through a process of back translation. In this process the translated instrument is given to two translators, native English speakers, who are instructed to translate the questionnaire back into English. It is important that this translator not have access to the original English language versions of the instrument and that he/she does not consult with the first translators.

**Independent Review and Comparison**

The third step in the translation process it to give the translated versions of the survey instruments to one or more bilingual reviewers. The reviewers are provided with the original English
versions and the back-translated versions and are instructed to compare the two, highlighting any discrepancies in meaning or equivalence.

**Review by Committee**

Once the review process is completed, the forward-translators, the back-translators and the reviewer(s) hold a series of meetings to discuss problems found during the review process, to correct errors in grammar and syntax and to resolve problems of equivalence found among the versions. Decisions on wording and corrections are made by consensus. The rationale behind this is that a translator or back-translator can introduce his or her own bias or error into a translation. The review-by-committee approach is useful in neutralizing the cultural, social, and ethnic bias that can be introduced when using only one translator and one back-translator.

**CAHPS® Translation**

Rather than produce multiple, population-specific Spanish translations, CAHPS® sought to produce an instrument that would be understood by most respondents by using “broadcast Spanish,” and that maintained a reading and comprehension level that would be accessible to most respondents. “Broadcast Spanish” refers to a type of Spanish that is understood by most Spanish speakers regardless of their country of origin or ethnic background (Marin and Marin, 1991).

A professional translator experienced in translating survey instruments similar to the CAHPS® instrument was retained. The translated instrument was then given to a bilingual reviewer experienced in designing and translating survey instruments for cross-
cultural research. The reviewer focused on identifying syntax and typographic errors, identifying questions or terms that sounded awkward and identifying terms that were conceptually problematic. Once this process was complete, the reviewer was provided with the English version and was asked to compare the two instruments, highlighting any discrepancies in meaning or equivalence.

In an effort to adhere as closely as possible to the English version, the translator produced an initial Spanish version of the survey instruments that was technically equivalent to the English version, but in many instances was not conceptually equivalent, and in some cases, not linguistically appropriate for the target population (by using terms that are seldom used in Spanish, anglicisms, or words that are too sophisticated for the target population). The translator had been instructed to aim for a translation that would be appropriate for a Spanish-speaking Medicaid population likely to have less than 6 years of formal education. However, this proved to be difficult to accomplish while maintaining equivalence to the English version.

A member of the RAND CAHPS® team met with the translator and the reviewer to go over discrepancies related to equivalence. The reviewer and the translator back-translated problem areas in the Spanish version to further distinguish the source of the problems before decisions were made about addressing them. A final review of the original English version, the translation, and the back-translation, was conducted by the committee—the translator, the reviewer, and CAHPS® team member—and alternative wording for problematic terms was implemented. Table 1 shows terms that were problematic because they were not conceptually equivalent, were too sophisticated for the target population, or were
too-infrequently used by most Spanish speakers. The alternative wording in the final version comes closer to the conceptual meaning in the English version and is easier for the respondents to understand.

**Qualitative Analysis (Step 5)**

Qualitative research consists of “research methods employed to find out what people do, know, think, and feel by observing, interviewing, and analyzing documents” (Shi, 1997, p. 398). These methods should be viewed as complementary to quantitative methods. Qualitative methods are particularly useful in assessing the cultural competence or content validity of the translated survey instrument\(^2\). It is important to evaluate whether the survey measures the group-specific domains of the phenomenon under study for the target population. Qualitative methods assist in identifying the “etic” (universal) and “emic” (culture-specific) constructs or behaviors of a group. This constitutes an evaluation of the “subjective” culture whereby consistencies or patterns in responses by members of a group are used to identify the group’s cognitive structure (Marin and Marin, 1991). The assumption is that the group’s norms, values, and expectancies influence the observed consistencies or similarities in responses of a given cultural group. Qualitative methods can also be used to assess the conceptual equivalence and linguistic appropriateness of the translated survey.

\(^2\) Herdman, Fox-Rushby and Badia (1997) recommend that qualitative methods of instrument evaluation precede the translation of survey instrument.
We are using qualitative methods to investigate the appropriateness of the CAHPS® survey content for Spanish-speaking Latino patients enrolled in Medicaid. First, we want to determine whether the items and scales currently contained within CAHPS® address the key concerns and expectations of Latino patients with respect to their health care providers and health plans. Second, we want to verify that the translated survey items, initially developed in English, have similar meaning in Spanish. Finally, we want to determine the readability level of the Spanish language survey instruments and determine whether it is appropriate for the Spanish-speaking Medicaid population.

There are three types of qualitative research pertinent to cross-cultural research: focus groups, cognitive interviews, and readability assessments. In this section we discuss the use of focus groups and cognitive interviews. For a discussion on readability assessments and its application to the CAHPS® surveys see Morales et al. (1999) in this conference proceedings.

**Focus Groups**

Focus groups are a research tool that relies on group discussions to collect data on a given topic (Morgan, 1996). Participant interactions help to reveal experiences, values, beliefs, and feelings. In addition, group discussion helps uncover extent of consensus or diversity, and its sources. Focus groups have been used extensively in marketing research to obtain customer input on new products (Burns and Bush, 1995); however, their use in cross-cultural research has been more limited. The primary objective of the focus groups in cross-
cultural research is to assess whether the domains currently covered in the survey adequately address the needs and expectations of the target population, and to assess the need for developing new domains or expanding current domains. The focus group process usually starts with a literature review and analysis of health surveys that focus on the target population, to aid in the identification of issues and concepts particular to the cultural group.

Stewart and Shamdasani (1990) have identified eight steps in the design and conduct of focus groups:

- Formulation of the research question
- Identification of sampling frame
- Identification of moderator
- Generation and pre-testing of structured protocol
- Recruiting the sample
- Conducting the focus group
- Analysis and interpretation of data
- Writing the report

A group size of 8 to 12 respondents per focus group is recommended (Burns and Bush, 1995). Homogenous groups based on demographics or other relevant characteristics are also recommended. This is important to elicit conversation among participants. Focus groups in cross-cultural research generally involve culturally
homogenous groups. However, the researcher may consider additional relevant demographic characteristics in forming the groups. For example, elderly Latinos versus teenager Latinos.

The moderator is the most crucial factor to ensure the effectiveness of the focus group. The focus group moderator conducts the entire session and guides the flow of group discussion across specific topics. According to Burns and Bush, the moderator "must strive for a very delicate balance between stimulating, natural discussion among all of the group members while at the same time ensuring that the focus of the discussion does not stray too far from the topic" (1995, p. 200).

In analyzing the data, the qualitative statements of the participants are translated into categories or themes and an indication is given of the degree of consensus apparent in the focus groups. The results of the focus groups inform the development of new items for the survey and the modification of existing measures as needed.

**CAHPS® Focus Group**

A focus group was conducted on November 7, 1998 at one of the clinics of a local health plan. The participants were recruited from among the Latino patient population of the health plan’s clinics in two Los Angeles County communities with high concentrations of Latinos. In order to be considered for participation in the focus group, patients had to be adults (18 and over) and primarily Spanish speaking.

A member of the RAND CAHPS® team moderated the focus group using a scripted discussion guide. The focus group was conducted entirely in
Spanish and lasted for approximately two hours. Twelve women, ranging in age from 24 to 73 years, attended the focus group. Eleven of the participants were from Mexico and one was from Nicaragua. All of the women had been in the U.S. for many years, ranging from 10 to 23 years.

The specific objectives of the focus group included:

- Determining Latino patients’ perceptions about health providers;
- Collecting information on communication issues between Latino patients and their providers;
- Gathering information on the use of interpreters by Latino patients;
- Seeking information on the role of the family in health seeking behavior and in making decisions about healthcare;
- Collecting information on Latino patients’ satisfaction with their health care, and
- Determining the most important aspects related to health care for Latino respondents.

Briefly, the results of this focus group raised interesting points:

- Provider’s communication is highly valued by Latinos: that a doctor spend enough time with them, that he/she ask them questions, and that he/she provide sufficient information about the patient’s illness and medications.
Participants were less concerned with the doctor’s Spanish speaking ability (although they do value it) and with the doctor’s race or gender.

- Participants reported some dissatisfaction with the care that they received from their health plan. Their chief complaints related to issues of promptness of care. Specifically, patients complained of difficulty obtaining timely appointments and long delays in seeing the doctor once arrived at the clinic.

- Most of the participants reported problems in using interpreters. They complained about the quality of the translation. In addition, patients reported not discussing certain personal health problems because of being ashamed to speak in front of their interpreter.

- Some participants reported going to Mexico to receive health care and the rest reported that they too would seek health care in Mexico if they could afford it financially. Among the reasons for preferring the care received in Mexico were the promptness of care, continuity of care, and provider’s communication and approach to care.

The findings from the focus group suggested that the substantive issues covered in version 2.0 of the CAHPS® Survey Instrument are culturally and substantively appropriate. Two of the findings from the focus group are not addressed as part of the survey and require further exploration. The first of these findings centers on the use and quality
of interpreters and how this affects provider-patient communication. Although the CAHPS® supplemental item set contains items that ask about the need and availability of interpreters, it does not cover the issue of interpreter quality and the effect of interpreters on communication between a provider and his/her patient. The second of these findings relates to patients who travel to Mexico to seek health care in spite of the fact that they can receive health care from their health plan. This information is being used to field test additional CAHPS® survey items addressing care in Mexico.

Cognitive Interviews

Cognitive-testing techniques are often used in the process of questionnaire development to investigate, assess, and refine a survey instrument (Berkanovic, 1980). Cognitive testing can detect and minimize some sources of measurement error by identifying question items or terms that are difficult to comprehend, questions that are misinterpreted by the respondents, and response options that are inappropriate for the question or that fail to capture a respondent’s experience (Jobe and Mingay, 1991).

One of the most common forms of cognitive testing is the cognitive interview to examine the thought processes of the interviewee. There are two forms of cognitive interviews: the concurrent and retrospective approaches. With the concurrent technique, the respondent goes through a process of “thinking-aloud” or articulating the thought processes as he or she answers a survey item. In the retrospective or “debriefing” technique, the interviewer asks questions about the survey process after the respondent completes the
Verbal probes or follow up questions may be used in either type of cognitive interview. One common probe is to ask the respondent to paraphrase the survey question. This helps to understand whether the respondent understands the question and gives it the intended interpretation. This may also suggest more appropriate wording for the survey item.

Prior to conducting the cognitive interviews, a structured protocol is developed to ensure that all participants receive similar prompts from the facilitators. The structured protocol is translated. Interviewers are bilingual in the target language and are trained in using cognitive interview techniques. Using notes taken during the cognitive interviews and audiotapes of each of the interviews, each interviewer writes up a summary for each interview in English. These summaries are then combined into one report outlining the results of the cognitive testing.

CAHPS® Cognitive Testing

The CAHPS team completed 150 cognitive interviews in different geographic locations (Harris-Kojetin et al., 1999). Seven cognitive interviews were completed in Spanish in California during June-July, 1996. A concurrent think-aloud technique with scripted probes was used in this case. The Spanish-language interviews were completed with adult women on Medicaid who were receiving AFDC benefits and were enrolled in either an HMO or a fee for service plan through Medicaid.
The primary objectives of the cognitive interviews were:

- To assess whether respondents understood the CAHPS® survey instruments.
- To determine the optimal response categories for ratings and reports of care.
- To identify the source of problems in comprehension: translation, reading level, survey content and cognitive task involved.

The results of each cognitive interview were summarized in reports and analyzed for points of convergence. In addition, the interviewers were debriefed and asked to provide general feedback on how well the instruments were working and to discuss content areas or issues that were problematic.

For the overall ratings, an adjectival scale (excellent, very good, good, fair, poor) was compared with a numeric scale (0-10). There was less translation difficulty with the numerical than the adjectival categories. It was particularly difficult to translate “fair” and “poor” into Spanish (Harris-Kojetin et al., 1999).

The cognitive tests were also used to explore whether key words and concepts worked equally well in Spanish and English. Specific wording and terms that were particularly problematic for Spanish-speaking respondents were modified based on the results of the cognitive testing and used to produce instruments that were ready for pretesting.
The interviewers reported that the survey instruments worked better with the respondents who seemed to be more educated or acculturated. Another issue identified by interviewers as problematic was that the instrument presumed that all prospective respondents were reasonably familiar with the terminology and landscape of the health care system in the United States. Familiarity with the system may be common for most Medicare and Medicaid recipients, but it also is related to length of time in the United States and to levels of acculturation, usually lower for non-English-speaking respondents.

**Field Test and Analyses (Step 6)**

A field test of the translated survey instrument is also recommended. Psychometric analysis can then be used to assess the reliability and validity of the translated survey instruments. Psychometric testing can also be used to test for measurement equivalence across cultural groups. Three types of analysis commonly used are:

Reliability estimates, such as Cronbach’s (1951) alpha coefficients, to measure the internal consistency of the instrument. Cronbach’s alpha is based on the number of items in the scale and the homogeneity of the items. The homogeneity of the items represents an average of the inter-item correlations in a scale and measures to what extent items share common variance.

Factor analysis to examine the internal structure of the instrument or construct validity of the scales. In addition, factor analysis can be used to test measurement invariance across groups (Reise, Widaman, and Pugh, 1993).
Item Response Theory (IRT) methods provide an ideal framework for assessing differential item functioning (DIF), defined as different probabilities of endorsing an item by respondents from two groups who are equal on a latent trait. When DIF is present, trait estimates may be too high or too low for those in one group relative to another (Thissen, Steinberg, and Wainer, 1993).

**CAHPS® Field Test**

A pretest of preliminary drafts of the CAHPS® 1.0 survey instruments was conducted as part of the Medicaid field-test data collection conducted by RAND in 1996 (Brown, Nederand, Hays, Short, and Farley, 1999). Only 23 respondents completed the interview in Spanish. All 23 Spanish-speaking respondents completed the interview by telephone. The total number of completes in Spanish was insufficient to conduct sensitivity analyses to determine whether the Spanish-language instruments were performing like the English-language instruments.

**Conclusion**

Adept translation of a survey instrument is an integral part of the instrument-development process, but it alone does not ensure that a culturally appropriate survey instrument will result. Cross-cultural adaptation of survey instruments requires that the translated instruments be conceptually and technically equivalent to the source language, culturally competent, and linguistically appropriate for the target population. Producing a survey instrument that is culturally appropriate for Latinos in the United States requires subjecting the
Spanish-language instruments to rigorous testing. That testing must include conducting focus groups and cognitive interviews that evaluate the cultural appropriateness of the survey content as well as the cognitive task required in the survey instrument, determining the reading level of survey instruments in Spanish, and field testing the survey instrument to ensure that the survey measures perform equally well in Spanish and English.

The results of the cognitive interviews and the focus groups may require modifying the English version of the survey instruments by adding domains to capture the experiences of Latino consumers, modifying the construction of items in English to make them more “translatable” into Spanish, modifying the Spanish version to accommodate ethnic and regional variations in Spanish language use, and simplifying the translation to make the reading level of the document appropriate for the target population.

In order to assess the cultural appropriateness of the CAHPS® 2.0 survey instruments among different Latino ethnic groups and to account for regional variations in care, focus groups and cognitive interviews will be conducted in San Diego, New York, and Miami. By conducting focus groups across these sites, we will incorporate Latinos of Mexican, Puerto Rican and Cuban origins in our focus groups. The qualitative component of CAHPS® is being done later than we would like. Ideally this phase would have taken place before finalizing the English-language instrument. Currently, we are also conducting a field study of the CAHPS® surveys among a Medicaid managed care population in the San Diego area. Our goal is to obtain 50% of completed surveys in Spanish.
References


Weidmer, B., Brown, J., & Garcia, L. (1999). Translating the CAHPS 1.0 survey instruments into Spanish. Medical Care, 37, MS89-MS96.
<table>
<thead>
<tr>
<th>Original English</th>
<th>Alternative wording used</th>
<th>Back-translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health insurance plan</td>
<td>plan de seguro médico</td>
<td>medical insurance plan</td>
</tr>
<tr>
<td>Health provider</td>
<td>profesional de salud;</td>
<td>health professional</td>
</tr>
<tr>
<td>Rating/rate</td>
<td>calificación/califica</td>
<td>grade/grade</td>
</tr>
<tr>
<td>Usually</td>
<td>normalmente</td>
<td>normally</td>
</tr>
<tr>
<td>Preventive health steps</td>
<td>medidas de salud</td>
<td>preventive health measures</td>
</tr>
<tr>
<td>Listen carefully</td>
<td>escucharon atentamente</td>
<td>listen attentively</td>
</tr>
<tr>
<td>Health care</td>
<td>atención médica</td>
<td>medical attention</td>
</tr>
<tr>
<td>Prescription medicine</td>
<td>medicamentos recetados</td>
<td>prescribed medications</td>
</tr>
<tr>
<td>Male or female</td>
<td>niño o niña/hombre o</td>
<td>boy or girl/man or woman</td>
</tr>
<tr>
<td></td>
<td>mujer</td>
<td></td>
</tr>
<tr>
<td>Background</td>
<td>ascendencia</td>
<td>ascendancy</td>
</tr>
<tr>
<td>Grade</td>
<td>año</td>
<td>year</td>
</tr>
<tr>
<td>School</td>
<td>estudios</td>
<td>studies</td>
</tr>
<tr>
<td>Highest</td>
<td>avanzado</td>
<td>advanced</td>
</tr>
</tbody>
</table>
Figure 1: Cultural Adaptation of Survey Instruments

Step 1: Forward

Step 2: Backward Translation

Step 3: Independent Review

Step 4: Committee Review

Step 5: Qualitative Analysis

Step 6: Field Test and Psychometric Analyses