

# EVALUATING TRANSLATIONS OF SURVEY INSTRUMENTS IN SPANISH AND ENGLISH: THE 1990 SPANISH LANGUAGE CENSUS LONG FORM

Susan M. Gabbard, Jorge Nakamoto, Aguirre International

Susan Gabbard, Aguirre International, 411 Borel Avenue, Suite 402, San Mateo, CA 94402

## Key Words: Translation, Census, Spanish

During 1992-1993, The Bureau of the Census sponsored and Aguirre International conducted an investigation of how Hispanics of different national origins (a) respond to questions in the 1990 Spanish language census form and (b) view the census process. A full description of the research and its findings is contained in Kissam, Nakamoto and Herrera 1993.

The results presented here derive from one component of the research design which consisted of in-depth interviews in which study participants were asked to perform an abbreviated census form completion task from the 1990 Spanish language long form followed by an in-depth debriefing. Study participants represented a cross-section of Hispanics in terms of national origin, literacy level, language usage, age, gender, and length and location of U.S. residence. Monolingual Spanish and Bilingual Spanish-dominant persons comprised an overwhelming majority of the study participants.

While some findings focused on more generic Spanish translations for English words and phrases, many of the findings illustrated the importance of other factors, primarily cultural background and educational level to accurate survey forms completion. Confusing cultural concepts included confidentiality, household, race, ethnicity and year of entry into the United States. For Hispanics with low levels of education, problems with text literacy were compounded by a lack of familiarity with surveys and a subsequent lack of document literacy. Overall, the results support the critical importance of actual field evaluation of survey instrument translations.

## Research Design

In-depth interviews consisting of an abbreviated census form completion task followed by an in-depth debriefing were used to: (1) explore specific semantic, syntactic, and conceptual issues affecting Hispanics' response to the census long form, (2) estimate respondent burden among distinct sub-groups among Hispanics, (3) assess the reliability of information elicited from respondents to the long form, and (4)

formulate recommendations regarding long form improvements to facilitate response.

To respond to the Census Bureau's concerns regarding diversity within a limited research budget and time frame, Aguirre International identified a purposive sample which would reflect the diversity of the Hispanic population in the U.S. Primary criteria for stratification included literacy level, national origin, and language usage (monolingual Spanish vs. bilingual Spanish-English).

Additional considerations in selecting study participants included achieving a reasonable gender balance, a reasonable representation of U.S.-born Hispanics, long-term immigrants, and recent immigrants, and a representative cross-section of Hispanics in terms of age. Monolingual Spanish and Spanish-dominant persons comprised an overwhelming majority of the study participants. Bilingual English-dominant U.S.-born Hispanics are somewhat underrepresented and monolingual English-speaking Hispanics are not included in the study population. The results reported here must be understood to relate primarily to the response patterns and perspectives of an immigrant Spanish-speaking Hispanic population.

Respondents were chosen from three separate geographical regions: the Washington, D.C. area, the San Francisco Bay area (San Francisco, Santa Rosa, San Jose), and the Los Angeles (Temecula, Riverside) area.

Using purposive sampling restricts the technical generalizability of the findings. However, the purpose of the research was to identify translation issues and not to quantify the prevalence of these issues. For this purpose, we believe that the findings are reliable and relevant. Our technical staff and other observers noted that the respondents selected seemed fairly representative of the larger Hispanic population. The reactions of the respondents was enlightening to observers, and many of their comments and recommendations are very reasonable.

In exploring Hispanics' responses to the census long form, Aguirre International addressed two distinct issues: (1) patterns of response and non-response, and (2) "quality" of response, (i.e., the extent to which information elicited from respondents might be comparable to information elicited from a hypothetical

"standard respondent"). The first issue relates to strategies for improving the Hispanic undercount. The second issue relates to the reliability of census data collected.

Study participants were asked to first perform an abbreviated census form completion task. They were then debriefed regarding their understanding of instructions, the meaning of questions, their assessment of the difficulty of various forms completion tasks, and possible alternatives for selected questionnaire terms and questions. Data collected includes ethnographic observations of participants' forms completion behavior, completed questionnaires, and observation protocols. The study data set also includes demographic information from a short participant screening interview. These data allow some cross-checking of information as reported on the census form and as reported orally to the interviewer. Additionally, there was discussion with all study participants about several general issues relating to the census form in general, the English vs. Spanish versions of the form, and the census process. These topics included the difficulty of the form, mailout-mailback vs. face-to-face enumeration, assistance in completing the census form, and strategies to increase motivation and response.

For each sample task item, we provide information on the type of assistance, support, or encouragement provided by the interviewer to the participant in order to enable them to respond to the question. Data on level of assistance rendered provide a measure of forms completion difficulty for each given item, even where it was necessary to assist the respondent to complete a question. The interviewer sought always to provide the minimal level of intervention required to keep the respondent from giving up on the form completion task. The abbreviated census form completion task consisted of responding to Questions 1a (Household Listing) through 7 (Hispanic Origin) for the respondent's entire household and responding to questions 10,12,13, 15, 21, and 26 for "Person 1" in the household. This 14-item abbreviated census form completion task thus consisted of reading instructions for, and responding to, the census form questions listed below :

Q. 1a (Household Listing)

Q. 1b (Temporary Residence)

Q. 2 (Household Composition and Relationships)

Q. 3 (Gender)

Q. 4 (Race)

Q. 5 (Age and Year of Birth)

Q. 6 (Marital Status)

Q. 7 (Hispanic Origin)

*Items only on the Long Form* [Completed only for "Person 1" in Household]

Q. 10 (Year of Immigration)

Q. 12 (Years of Schooling)

Q. 13 (Ethnic Origin)

Q. 15 (Home Language)

Q. 21 (Prior Week Work)

Q. 26 (Job Search)

The interview and debriefing procedure used in the experimental forms-completion session followed four steps:

1. orient study participants to the study procedure,
2. proceed through page two of the census form -- Q. 2 (Household Composition) through Q. 7 (Hispanic Origin),
3. query participants on housing and other long form questions, and
4. discuss with all study participants several general issues relating to the census form and the census process.

In the first step the study procedure is defined as consisting of (a) completing the census form and (b) reviewing and discussing the questions with the interviewer. Study participants read and completed the first page of the census form -- Q. 1a (Household Listing) and Q. 1b (Temporary Residence) before being queried by the interviewer about this initial task. The second step was for study participants to proceed through page two -- Question 2 (Household Composition) through Question 7 (Hispanic Origin). They were then interrupted to discuss these items with the interviewer. The third step was different for low-literate vs. high-literate respondents. Study participants with less than nine years of schooling were directed straight to the selected questionnaire long form items (Q. 10, 12, 13, 15, 21, 26), and skipped the series of housing questions. Study participants with nine years of schooling or more were queried on a list of terms from the series of housing questions (Q. H2 through Q. H26) corresponding to pages 4 and 5 of the census long form. They then continued on to the other long form questions. The fourth and final step was to discuss with all study participants several general issues relating to the census form and the census process. These topics included the difficulty of the form, mailout-mailback vs. face-to-face enumeration, assistance in completing the census form, and strategies to increase motivation and response.

The forms completion task and interviews were conducted, wherever possible, at the informant's home to assess the forms completion task in an environment as close to the "natural" social and physical environment of the respondent as possible. Thus, in some cases, the

forms completion task drew in other household members, making the forms completion task a collective one, with some similarities to what would occur in responding to the actual census. In-home interviews were conducted in a number of extremely crowded settings (e.g., a 200 square foot partitioned section of a garage occupied by a family of four); however, in some cases it was not feasible to conduct the interview at home and the interview took place at a neutral public place -- the office of a community organization or a local restaurant. Interview sessions were tape recorded.

## Findings

1. Hispanics with less than 6 years of schooling find it very difficult to complete the census long form without active and constant assistance. These difficulties are more related to a lack of literacy skills required for successful forms completion (document literacy) than to other factors (i.e., individual, specific problematic census concepts, lexical difficulties stemming from English to Spanish translation, or dialectical differences between Hispanic sub-groups of differing national origin.) Hispanics with 12 or more years of schooling experience no substantial difficulty in responding to the long form but still encounter some difficulties in responding to some questions.

2. A variety of non-linguistic and linguistic conventions commonly used in questionnaire and survey research (e.g., parenthetical instructions, instructions for skip patterns, schemata for pre-coded responses, ostensive definitions) contribute heavily to low-literate respondents' difficulties with the census form. The disparity between "forms literacy" skill requirements and "text literacy" skill requirements gives rise to respondents' perceptions that the census long form is extremely burdensome. Actual respondent burden for low-literate respondents substantially exceeds the OMB estimate and is difficult to estimate since it is doubtful that respondents would persevere in responding to the form without assistance.

The census form formatting style requires at several different points in the course of completion recourse to a meta-cognitive skill -- the ability to rapidly establish clearly-delineated frames of semantic reference within which terms are ordered hierarchically. This competency which is well beyond the information processing experience of low-literate Hispanic readers, since "basic" reading literacy for persons with an elementary school education consists primarily of the ability to read page-width linear text.

Smooth and easy progression through the form requires a substantial amount of "frame shifting" to minimize the burden of parenthetical clarifications (e.g., of the inclusion/non-inclusion rules) and fine-grid

elaborations of categories (e.g., of racial sub-groups). Complete and correct forms completion requires yet other meta-cognitive skills, primarily the ability to rapidly scan responses to detect inadvertent omissions (e.g., a box with no recorded response), coding errors (e.g., double-marking in age columns). While the partial form completion task in our study is not well-suited to examining respondents' ability to follow skip pattern instructions, it is likely this also creates difficulties. Overall, successful response to the census long form requires a general awareness of effective strategies in responding in the face of uncertainty -- skipping a particularly difficult question, prioritizing and re-reading the most critical instructions while skipping others.

Despite our interviewer assurances and Bureau of the Census assurances that census forms completion is not "a test," many respondents perceived it as such. Many study participants faced "the test" with some apprehension and left feeling a fair measure of anger, disgust, or humiliation. One study participant repeatedly requested that the tape recorder be turned off so her questions to the interviewer would not be included as part of her "test results." Other study participants made verbal comments on their performance deficits -- e.g., *¡Qué boba!* [How foolish] Other study participants made facial expressions of disgust in correcting household members' names to the "last name/first name" format.

3. The Census Bureau's implicit conceptual framework relating concepts of race, Hispanic origin, and ethnicity were considered inadequate by most study participants. The separate questions relating to race, Hispanic origin, and ethnicity were considered to be, on the one hand, redundant, since these questions were not perceived to successfully articulate and query distinct concepts. On the other hand, the questions were perceived as inconsistent -- "asking the same question different ways." Interpretations of the meaning of all three terms vary extensively from respondent to respondent. In contrast, the concept of "national origin" is, for most, a well-anchored concept.

4. Concepts of household structure implicit in Census long form completion instructions compromise the data quality of several key long form variables -- including household size and composition, age, length of time in the U.S. Presumably, these difficulties affect several other key data items such as income, earnings, and use of government programs which were not examined in the current study. A ubiquitous feature of the life of Hispanic and other immigrants is the existence of "complex households" consisting of one or more unrelated nuclear families and unrelated individuals sharing housing accommodations. These are often crowded and may include extensive sub-divisions within

the original housing unit. Internal inconsistencies in the household listing instructions, high respondent burden stemming from extremely detailed instructions, and lack of space for listing household members jeopardizes the validity of household listing and, consequently, the enumeration process.

5. Syntactic, conceptual and formatting difficulties in several key questions on the Census long form -- including those relating to year of arrival, home language and schooling -- compromise the data quality of these variables with respect to low- and mid-literate respondents.

6. There are a number of lexical and syntactic problems in the Census long form to be found in the series of questions relating to housing. Due to the overall respondent burden experienced by low-literate study participants, the interviewer was instructed to explore comprehension of presumed difficult housing terms only for persons with 9 years or more of schooling. Even among intermediate and fully literate respondents some terms were unfamiliar and presented difficulties. Study participants of various national origins were observed to have difficulties with various housing terms. This area is the one where we observed the most substantial response difficulties apparently linked to dialectical variations within Spanish.

## Conclusions

The current study addresses in detail the ways in which Spanish-speaking Hispanics respond to the census long form. One of the most important conclusions to be drawn from the study is that forms completion demands are so burdensome to low-literate individuals that response rates are likely to be adversely affected. This, in turn, indicates that there is indeed a differential undercount between various sub-groups within the Hispanic population (e.g., recently-arrived, unattached, undocumented male immigrants). The study also suggests there are a number of clearly-identifiable problems in terms of reliability of the responses provided by low-literate Spanish-speaking persons who did respond to the census form. These stem from a variety of causes. The aggregate result of sub-group differential undercount and impaired data quality for key questionnaire items serve to make some census-based analyses of issues affecting Hispanics unreliable.

Among the study population, omission of "marginal" persons appears to be the leading factor in partial household omission. Form simplification, even at the expense of precision (e.g., lack of guidance regarding children in boarding school) would more effectively elicit from low-literate respondents "correct" listing.

## References

Kissam, Edward, Jorge Nakamoto and Enrique Herrera. *Hispanic Responses to the Census Enumeration Forms and Procedures*. Final Report to the Center for Survey Methods Research, Bureau of the Census, Aguirre International, San Mateo, California March 1993.