

The Effect of Inadequate Language Translation On Hispanics' Responses to Health Surveys

EMIL BERKANOVIC, PHD

Abstract: Data bearing on the effect of language of the interview on Hispanics' responses to a health survey in which no back-translation was undertaken reveal both lower reliabilities and lower bivariate correlations among Hispanics interviewed in Spanish than among Hispanics interviewed in English. An independent back-translation aimed at creating an English version of the questionnaire that was linguistically equivalent to the Spanish version indicated several in-

stances in which the Spanish version was unidiomatic. Differences between the Spanish and English version in the idiomatic quality of the interview items, while not affecting meaning, appear to have affected the seriousness with which the interview situation was perceived. These perceptions, in turn, appear to have led to the response discrepancies observed. (*Am J Public Health* 1980; 70:1273-1281.)

Recently, the importance of survey research in the planning and evaluation of health services has been reemphasized.^{1,2} In addition, attention has been called to the importance of collecting accurate health data on the growing Hispanic minority in the United States.³⁻⁵ The methodological difficulties in the collection of health data from Hispanics has been discussed by Aday, *et al.*³ Salber and Beza,⁶ and Rice, *et al.*⁷ The present paper offers data on one aspect of these difficulties—the effect of inadequate translation on Hispanics' responses to health surveys.

Although the effect of both language and culture on the stimulus value of health survey items has long been of concern to survey researchers working in multilingual communities, there is little evidence bearing on the practical consequences of questionnaire translation. For example, one might be interested in whether indices demonstrating adequate levels of reliability in one language achieve similar levels of reliability when translated. Similarly, one might ask whether predictive relationships observed among the variables in one language are replicated among those to whom the questionnaire was administered in translation.

Such questions are important for health survey research among Hispanics because health surveys in many parts of the country allow the respondent to choose between an English and a Spanish version of the interview instrument. Typically, the questionnaire is written in English and then translated into Spanish. The recommended technique is back-translation in which the Spanish language version is translated back into English by someone other than the original translator.* This process is repeated until an English translation is made that approximates the original. The Spanish version of the questionnaire that was used for the final Eng-

lish back-translation is then used in the field. As a practical matter, however, the number of independent iterations most survey research organizations are either willing or capable of doing is limited.

Method

The present paper compares data derived from Hispanics responding to an interview in English with data derived from Hispanics responding to the Spanish language version of the interview. The data were taken from a health survey conducted in a large West Coast city. In 1976, a sample of 1,210 individuals statistically representative of this city were interviewed. All field work for this study was conducted by a large, academically based survey organization.

There were 202 individuals who identified themselves as Hispanics. These respondents were asked if they wished to be interviewed in English or Spanish. Eighty-six chose to be interviewed in Spanish. The Spanish language version of the questionnaire was developed by the organization conducting the field work. Because of severe time constraints, there was no back-translation of the Spanish language questionnaire. It is the impact of this translation on Hispanics' responses to certain items in the questionnaire that is reported herein.

The interviews included several sets of Likert items intended to form indices of health beliefs. These indices were to be examined for their ability to predict several variables related to the decision to seek medical care. In this analysis, the reliability levels of these indices are examined separately for those Hispanics responding in Spanish and those responding in English. Item standard deviations are also exam-

*For an excellent example of the use of this technique as well as a review of the literature bearing on the translation problem, see Gilson, *et al.*⁸

Address reprint requests to Emil Berkanovic, PhD, Associate Professor, School of Public Health, University of California, Los Angeles, CA 90024. This paper, submitted to the *Journal* May 8, 1980, was revised and accepted for publication August 8, 1980.

ined wherever discrepancies in the reliability of the two language versions are found. This is done to rule out the possibility that any differences in reliability are a function of differences in item variance between the two translations. Additionally, the relationship between these indices and one dependent variable is examined separately for the two translations. Finally, the implications of an independent back-translation of the questionnaire conducted after the survey was completed are examined.

All Spanish language interviews were conducted by bilingual, bicultural interviewers. Eleven of the 51 interviewers participating in this study met these criteria. In order to maximize the efficiency of the field operation, an attempt was made to assign all interviews that were likely to be with Hispanics to bilingual interviewers. Thus, about 50 per cent of the Hispanics responding in English were interviewed by these bilingual interviewers.

TABLE 1—Alpha Reliability Coefficients for Health Belief Scales

	Total Sample (N = 1,210)	Hispanics	
		English Speaking (N = 121)	Spanish Speaking (N = 86)
Acceptability	.80	.83	.75
Accessibility	.73	.71	.73
Susceptibility	.67	.65	.70
Motivation	.65	.56	.69
Cost Concern	.79	.76	.78
Seriousness	.62	.66	.37
Efficacy of Care	.61	.53	-.05

Results

Seven indices of health beliefs were developed for this research. Table 1 presents the alpha reliabilities for these indices for three groups of respondents: the total sample of 1,210, the 121 Hispanics who were interviewed in English, and the 86 Hispanics who were interviewed in Spanish. Substantial discrepancies exist between the two Hispanic groups for two of the seven indices: the perceived seriousness of illness, and the perceived efficacy of medical care. In both instances, the reliabilities are considerably higher among Hispanics responding in English than among Hispanics responding in Spanish. In both instances, the reliabilities for the Hispanics responding in English are similar to those for the entire sample, but the reliabilities for those responding in Spanish are unacceptably low.

Table 2 presents the means and standard deviations for the individual items comprising the two indices for which discrepancies were observed in Table 1. Although differences in item means have little impact on reliability, differences in the magnitude of the item variances could affect reliability by affecting the size of the interitem correlations. It is clear from Table 2, however, that the differences between the two language versions are quite small.

What is the impact of these discrepancies on the prediction of outcomes? Unfortunately, although the survey from which these data are drawn contains several dependent variables relevant to the use of health services, the low number of Spanish language respondents coupled with the infrequency with which they used health services limits the analysis to a single comparison permitting the inclusion of all Spanish language respondents.

Table 3 presents the partial correlations between self-reported tendency to delay seeking health care, and the Seri-

TABLE 2—Standard Deviations (SD) and Means for Items in Seriousness and Efficacy of Care Scales

	Hispanics			
	English Speaking		Spanish Speaking	
	SD	\bar{x}	SD	\bar{x}
1. Seriousness Scale				
a. When I'm sick I try to keep going on as usual.	.57	2.28	.51	2.26
b. When I'm sick I cut back on whatever I'm doing in order to get well.	.49	2.86	.39	2.87
c. I almost <i>never</i> take the illnesses I get seriously.	.56	2.67	.58	.258
d. When I'm sick I usually try to ignore it.	.56	2.68	.44	2.72
e. Whenever I'm ill, no matter how mild the symptom, I take it seriously.	.60	2.57	.58	2.78
2. Efficacy of Care Scale				
a. Most of the time when I'm sick I don't think a doctor can do much good.	.60	2.87	.58	2.79
b. I can take care of most of the illnesses I get as well as a doctor can.	.60	2.63	.52	2.79
c. I think a doctor can do a lot of good for most of the illnesses I get.	.53	2.79	.48	2.83

ousness and Efficacy of Health Care scales. Age, income, education, and occupational ranking were controlled in these partial correlations. As shown, both partial correlations are .28 ($p < .01$) for respondents interviewed in English, but only .06 and .12 for respondents interviewed in Spanish. This pattern coincides with what one would expect on the basis of classic test theory, i.e., lower reliability is expected to attenuate statistical relationships. Consequently, these data provide a textbook illustration of how differences in the reliability of two translations can affect the conclusions one draws. Also note that we have examined only half the problem. There may also be differences between the Spanish and the English language versions in the stimulus value of the dependent variables. Substantively, of course, if language of the interview impacts on the predictive conclusions one might draw from survey data, the value of such data for planning, monitoring, and evaluating services is compromised.

These data indicated that Hispanics interviewed in Spanish respond differently to some items than do Hispanics interviewed in English and that these differences can affect the conclusions one might draw regarding the level of association between two or more variables. Given these results and the fact that the Spanish language version had not been back-translated, an experienced, bilingual survey researcher who had no previous connection with this project was asked to back-translate the Spanish language questionnaire. Without being told the purpose of the present study, this individual was asked to provide an English language version that achieved a level of idiomatic usage equivalent to the Spanish language version. Thus, instead of seeking the best English language version, we asked for a version that most accurately reflected the linguistic properties of the Spanish version that had been used in the field.

The back-translation revealed that two of the four items comprising the perceived seriousness of illness index and two of the three items comprising the efficacy of care index were unidiomatic. Further, although the items comprising the other fixed indices were unaffected, there were language irregularities of less magnitude in most of the dependent variables used in this study. It was the opinion of the individual doing the back-translation that the intent of the questions would be understood by most of the respondents interviewed in Spanish, but that they would be somewhat amused by the lack of facility with which the language was used.

This insight suggests that one source of the response discrepancies observed in these data results from differences in the circumstances under which the interviews were conducted in each language. It is well established that responses to surveys are shaped by the total set of demands present in the interview situation as they are perceived by the respondent.^{9, 10} The particular questions being asked are only part of the overall "demand characteristics" operative during the interview. The fact that the Spanish language version is unidiomatic in several respects, while, perhaps not threatening to the intelligibility of the questions, suggests that the overall interview situation may have been perceived differently by Hispanics interviewed in Spanish than it was by Hispanics interviewed in English. Specifically, it is plausible that the English language interview was taken more seriously than

TABLE 3—Partial Correlations between Seriousness and Efficacy of Care and General Tendency to Delay

(1 = low delay)	(2 = medium delay) (3 = long delay)† General Tendency to Delay	
	English Speaking Hispanics (N = 116)	Spanish Speaking Hispanics (N = 81)
Health Belief Scale		
Seriousness	-.28**	-.06
Efficacy of Care	-.28**	.12

* $p < .05$; ** $p < .01$; *** $p < .001$

†Partial correlations control for respondent's age, sex, marital status, income, and socioeconomic rank (education and occupation).

the Spanish language interview. As a consequence, the English language respondents may have been more willing to do the work required to respond to the questions accurately than were those interviewed in Spanish.

One test of this proposition that could be undertaken with the data at hand consists of comparing the interviewer ratings of the honesty of both groups of respondents. Whereas 85 per cent of the Hispanics interviewed in English were rated by the interviewer as very honest, 67 per cent of the Hispanics interviewed in Spanish were so rated. This difference was significant at the .005 level. That these interviewers perceived those responding in Spanish as less honest is consistent with the idea that Hispanics interviewed in Spanish may have taken the interview less seriously than Hispanics interviewed in English.

It is possible, however, that the differences in reliabilities observed between Hispanics interviewed in English and Hispanics interviewed in Spanish were due to differences in level of education. Thus, it may be that those interviewed in Spanish were seen by the interviewers as less honest not because they were less serious about the interview, but because they did not fully understand the questions. This possibility may be discounted, however, because the differences among the reliabilities persisted after education was controlled.

The possibility that differences in how the bilingual interviewers conducted the interviews may have affected the reliabilities of these indices was tested by comparing the reliabilities for the Spanish language interviews with the reliabilities for the English language interviews among the bilingual interviewers only. The results of this analysis were mixed. Thus, whereas both the reliabilities for the unaffected indices and the reliability for the efficacy of care index were similar to those observed in Table 1, the differences in the reliabilities for the seriousness index were substantially less than those observed in Table 1. It appears, therefore, that, in addition to difficulties with the translation, there may have been some interviewer effects for the seriousness index. Such effects would not be surprising among bilingual interviewers required to use an unidiomatic interview. Why they do not appear in the efficacy index is unclear. Unfortunately, there is no information available regarding how these interviewers felt personally about the Spanish language version.

Conclusion

At present, there is discussion regarding the appropriateness of any single Spanish language translation of materials intended for such culturally diverse groups as, for example, Mexicans, Cubans, and Puerto Ricans. Clearly, this is an issue that is best resolved experimentally. Indeed, one would expect that it is within the means of both the Bureau of the Census and the National Center for Health Statistics to undertake randomized studies of alternative translation strategies. What the present data indicate, however, is that considerable care needs to be taken in order to assure that any translation, whether culturally specific or not, achieves a level of language usage that is equivalent to the original source. Without such assurances, cross-language studies will be subject to the threat that between-group differences are an artifact of differences in the adequacy with which different languages are used. At a minimum, no cross-language survey should be fielded without the assurance of language equivalence offered by the blind back-translation technique.

REFERENCES

1. Phillips HT and Beza AG: The use of health surveys in health systems agency planning. *Am J Public Health* 1979; 69:221-222.
2. Siemiatycki J: A comparison of mail, telephone, and home interview strategies for household surveys. *Am J Public Health* 1979; 69:238-246.

3. Aday LU, Chiu Gy and Andersen R: Methodological issues in health care surveys of the Spanish heritage population. *Am J Public Health* 1980; 70:367-374.
4. Roberts RE and Lee ES: The health of Mexican Americans: Evidence from the Human Population Laboratory studies. *Am J Public Health* 1980; 70:375-384.
5. Roberts RE and Lee ES: Medical care use by Mexican-Americans: Evidence from the Human Population Laboratory studies. *Med Care* 1980; 18:266-281.
6. Salber EJ and Besa AG: The Health Interview Survey and minority health. *Med Care* 1980; 18:319-326.
7. Rice DP, Drury TF and Mugge RH: Household health interviews and minority health: The NCHS perspective. *Med Care* 1980; 18:327-335.
8. Gilson BS, Erickson D, Chavez CT *et al*: A Chicano version of the Sickness Impact Profile (SIP). Paper presented to the Hispanic Health Service Research Conference, Albuquerque, NM, September 5, 1979.
9. Cannell CF and Kahn RL: Interviewing, IN: Lindzey G and Aronson E (eds): *The Handbook of Social Psychology*, 2nd edition, Volume 2, Addison-Wesley, 1968.
10. Rosenthal R and Rosnow, RL (eds): *Artifact in Behavior Research*. Academic Press, 1969.

ACKNOWLEDGMENTS

Data reported in this paper were collected pursuant to Grant Number 5-R18-CA-18451, "Processes in Health Behavior and Cancer Control", awarded by the National Cancer Institute, DHEW. An earlier version of this paper was presented to the Hispanic Health Services Research Conference, Albuquerque, NM, September 4-6, 1979.

5th Southeastern Conference on Alcohol and Drug Abuse Scheduled

The Fifth Southeastern Conference on Alcohol and Drug Abuse, a symposium co-sponsored by Peachford Hospital and the American Medical Society on Alcoholism, will be held December 3-7, 1981 in Atlanta, Georgia. Registration for the program is limited to the first 1,000 applying.

Utilization of time-proven principles for the treatment and recovery of those addicted to alcohol and other drugs will be presented by the numerous well-known program participants. Continuing education unit credits for physicians and nurses can also be earned.

Early registration is urged, due to the attendance limitation. For further information, contact Mary Galos, Peachford Hospital, 2151 Peachford Road, Atlanta, GA 30338—404/455-3200 extension 550.