

Speaking the Same Language: Effective Techniques for Reaching Spanish-speaking Households in a Mail Survey

Andrew Zukerberg, National Center for Education Statistics ¹
Saida Mamedova, American Institutes for Research

Abstract:

Reaching non-English speaking households is a challenge for many surveys, especially those conducted by mail. Unlike telephone surveys, where the interviewer can immediately identify a language issue and route the case to an interviewer that speaks the respondent's language, a mail survey must identify ways to target the household prior to contact. As part of the transition from a telephone administered to a mail self administered design, the National Household Education Survey (NHES) has conducted a number of experiments to look at optimal ways to identify and reach Spanish-speaking households. The issue of correct identification of the language spoken in the household is especially acute for the NHES, as it is a two-phase study, where sampled households are screened with a simple household roster to determine the presence of eligible children. If eligible children are present, within-household sampling is performed to select a reference child. The household is then sent a longer and more complex topical survey. The screener is used to determine the language of the topical survey. In the 2007 NHES administration, 4.8 percent of screener interviews were conducted in Spanish. This paper examines the results of experiments designed to: determine the best questionnaire approach to reach Spanish-speaking households; determine if there is a 'backfire' effect whereby response in English-speaking households is suppressed by the inclusion of Spanish language materials; and to identify the most effective mailing strategy to maximize identification of Spanish speakers.

Background:

While English is the only language spoken at home by the vast majority of Americans, the percentage of those that use another language as primary mode of communication inside the home has been rapidly increasing. Amongst those that speak a language other than English, Spanish is the most prevalent language spoken (Shin et al., 2010). Nearly one third (29%) of the population that speaks primarily Spanish at home reported that they cannot speak English well or at all (Shin et al., 2010). Households where the adults cannot speak English well or at all are referred to as Linguistically Isolated (LI). Many telephone surveys employ bilingual interviewers and have procedures to shift a case to these interviewers when a language issue is suspected. For mail-based surveys language can be a challenge as it is difficult to identify households that do not speak English ahead of time. Previous research has tested sending separate English and Spanish forms in the same mailing, bilingual forms with English on one side and Spanish on the other (De la Puente and Wobus, 1994) and a bilingual form where the same question is presented in

¹ This discussion is intended to promote the exchange of ideas among researchers and policy makers. The views expressed during discussion are part of ongoing research and analysis and do not necessarily reflect the position of the U.S. Department of Education

two languages next to each other (Bouffard and Tancreto, 2006). Both of these experiments found that the addition of the Spanish language option increased response rates. However, a study conducted as part of a pilot test of the National Household Education Survey in 2009 found no detectable difference in response rates between a treatment group that received a bilingual screener and one which received only English materials in a linguistically isolated sample (Zukerberg and Han, 2010). The study showed the possibility that including a bilingual form traded coverage improvements for response rate. That is, while the bilingual materials allowed households with limited or no English-speaking ability to participate, it suppressed response in some English-speaking households. However, the design of the study did not fully allow the evaluation of this possibility. Wobus and De la Puente (1995) found some evidence that non-Hispanic households could be turned off by the inclusion of Spanish materials. In a debriefing of non-Hispanic households, nearly 13% indicated that they thought a Spanish language Census form was a “bad idea.”

In this paper we further explore the impact of offering materials in English and Spanish for the NHES using three distinct samples and different contact strategies. The experiments were designed with the following goals:

- 1) Determine the best questionnaire (dual Spanish/English questionnaires or bilingual) to reach Spanish-speaking households;
- 2) Determine if there is a ‘backfire’ effect whereby response in English-speaking households is suppressed by the inclusion of Spanish language materials;
- 3) Determine the most effective mailing strategy to maximize identification of Spanish speakers.

From this, we present an optimal mailing strategy for reaching Spanish-speaking households in a two-phase mail survey.

Methods:

The NHES is a two-phase survey. In the first phase, households were screened by mail to determine if there are any eligible children living at the address. If there were children in the household, a ‘topical’ survey of education-related questions was sent at the second phase. Households received a reminder postcard and up to three questionnaire mailings at each phase. In addition, some households received telephone non response follow up after the third questionnaire mailing. The screener first determined if any children age 20 or younger lived in the household. If there were no eligible children, the respondent marked a box indicating this and was asked to return the questionnaire. If there were eligible children, the respondent was asked to provide the name, age, enrollment status and grade for each child and a contact phone number for the household. A reference child was selected from each household for reporting in the topical questionnaire. The language used to respond to the screener in the first phase determined the language that the topical questionnaire would be sent in. Three versions of the screener were tested: 1) an English only version 2) a Spanish only version and 3) a bilingual version that had the questions in English on one side and Spanish on the other side. Different timing approaches were tested. For example, some households received a Spanish and English screener together at each mailing (referred to as a dual mailing), while other households received an English form at the first mailing and dual (English and Spanish) forms at the second and third mailings. Table 1 shows the different mailing approaches that were tested with each sample.

Table 1: Treatment path and sample

Treatment path name	1st mailing	2nd mailing	3rd mailing	Sample tested in
Dual all	English and Spanish forms	English and Spanish forms	English and Spanish forms	Linguistically Isolated (LI), Hispanic Surname (HS), National Sample (NS)
Bilingual	Bilingual form	Bilingual form	Bilingual form	LI,HS
Dual 2nd	English form	English and Spanish forms	English and Spanish forms	NS
Dual 3rd	English form	English form	English and Spanish forms	LI,HS
English only	English form	English form	English form	LI,HS,NS

A nationally representative sample and a targeted Spanish sample were used in this test. The samples were drawn from a frame of all addresses in the United States maintained by Marketing Systems Group (MSG). The frame is based on the USPS delivery sequence file and is enhanced with data from other sources to allow for stratification and sampling. The sample included Post Office boxes.

The National Sample (NS) was an independently drawn nationally representative sample of addresses in the US based on the enhanced USPS delivery sequence file created by MSG.

The targeted Spanish sample was selected from MSG's enhanced USPS deliver sequence file. The sample was divided into two mutually exclusive subsamples. A high linguistically isolated Spanish-speaking (LI) sample that included addresses in Census blocks where 13 percent or more of the households had no one over the age of 14 who spoke only English or who spoke English 'well or very well.' The allocation of tracts to this group was made using Census 2000 data and approximately 48 percent of all linguistically isolated Spanish-speaking households in the U.S. were represented by these tracts. The second subsample was a Hispanic surname (HS) group. This consisted of addresses where the frame indicated that the head of the household had a Hispanic surname. The indicator is based on matching the surname to a Census Bureau file of surnames that are commonly shared by people of Hispanic origin. These cases were selected from Census tracts that were not eligible for the LI sample.

Table 2 shows the sample size for each sample and treatment group. While initial screener sample sizes are large enough to detect modest differences between groups, approximately 30 percent of screener respondents are eligible for the follow up topical interview on which much of the analysis in this paper is based. The smaller sample size

at the second phase contributes to instability in some of the estimates. These are noted in the tables below and caution should be used in interpreting the results.

Table 2: Sample size and treatment path

Sample and form type	Sample size
Linguistically Isolated sample	8,600
English only	1,400
Dual all	2,900
Bilingual	2,900
Dual 3 rd	1,400
Hispanic Surname sample	10,200
English only	1,700
Dual all	3,400
Bilingual	3,400
Dual 3 rd	1,700
National sample	3,400
English only	2,200
Dual 2nd	600
Dual all	600

Each questionnaire package included a cover letter that explained the purpose of the survey and a postage paid return envelope. The cover letter was in English for the English only questionnaire and had English on one side and Spanish on the other for the dual form (English/Spanish) and bilingual questionnaire mailings. All respondents were randomly assigned to receive either a \$2 or \$5 cash incentive with the first mailing. A reminder postcard was sent one week after the first mailing. The English treatment group received a postcard in English that included a toll free number and information that the survey was available in other languages by request. The Dual and Bilingual groups received a bilingual postcard which had Spanish on one side and English on the other. The content of the two postcards was identical. The first two questionnaire mailings were sent by first class mail. The third mailing was sent by FedEx. Some respondents received a telephone follow up after the third mailing. However, only the results obtained by mail are discussed in this paper. A total of five different treatment paths were tested. Data collection began in early January 2011 and ended in June 2011.

Findings:

Table 3 shows the response rate by mailing strategy and treatment group for the different samples in the NHES field test.

Table 3: Response rate by treatment group

Treatment group	Overall response rate	Comparison	Significance
Linguistically Isolated			
A. English only	51.6	A:B, A:C, A:D	~,~,~
B. Dual all	54.1	B:C	~
C. Bilingual	54.2	C:D	~
D. Dual 3 rd	53.4		
Hispanic Surname			
E. English only	59.6	E:F, E:G, E:H	~,~,~
F. Dual all	61.3	F:G	~
G. Bilingual	62.9	G:H	*
H. Dual 3 rd	60.5		
National Sample			
I. English only	67.8	I:J, I:K	** , ~
J. Dual 2 nd	72.0	J:K	*
K. Dual all	67.4		

Note:

* p < 0.10 significance using two-tail test

** p < 0.05 significance using two-tail test

~ no difference detected using two-tail test

There was no statistically significant difference in response rates between the treatment paths within the LI subsample. In the Hispanic surname subsample the Bilingual path performed better than the Dual 3rd path. However, there was not a detectable difference between the English only and Spanish treatment paths in this subsample either. This may be in part a sample size issue. Combining the Dual All and Bilingual treatment paths generates a detectably higher response rate at the 0.10 level of significance than the English only group in both targeted Spanish subsamples. Within the national sample the Dual 2nd group performed significantly better than the English only and Dual all groups.

The language the respondent used to complete the screener questionnaire determined which language version of the more complex topical follow-up questionnaire would be sent. Thus, more completed interviews in Spanish were considered an indicator of improved identification of Spanish speaking households. Table 4 shows the percent of screener forms completed in Spanish by each treatment path. A case was considered Spanish language if they completed primarily the Spanish column of the bilingual screener form or returned a Spanish screener questionnaire. Cases in the English only treatment path could call and request materials in Spanish. A very small number made this request.

Table 4: Language of screener form completed by treatment path

Treatment path	Percent of screeners completed in English	Percent of screeners completed in Spanish	Comparison	Significance
Linguistically Isolated				
A. English only	100	#	A:B, A:C, A:D	**, **, **
B. Dual all	72.5	27.5	B:C	~
C. Bilingual	72.1	27.1	C:D	**
D. Dual 3 rd	93.8	6.2		
Hispanic Surname				
E. English only	99.9	.1‡	E:F, E:G, E:H	**, **, **
F. Dual all	79.3	20.7	F:G	*
G. Bilingual	80.6	19.0	G:H	**
H. Dual 3 rd	95.5	4.5!		
National Sample				
I. English only	100	#	I:J, I:K	~, **
J. Dual 2 nd	98.3	1.7‡	J:K	**
K. Dual all	95.7	4.3!		

Note:

* p < 0.10 significance using two-tail test

** p < 0.05 significance using two-tail test

~ no difference detected using two-tail test

‡ Reporting standards not met (too few cases).

! Interpret data with caution (estimates are unstable).

Rounds to zero

Table 4 shows that the survey identified a higher percent of Spanish-speaking households in both the LI and HS subsamples when Spanish materials were sent. There was a slightly better identification of households using the Dual form approach compared to the other treatments in the HS sample (20.7% compared with .1%, 19% and 4.5%). Within the national sample, the Dual 2nd approach was less effective at identifying Spanish-speaking households than the Dual All approach (1.7% vs. 4.3%).

Sending Spanish forms with all mailings (Dual All) in the national sample reduced overall screener response rates in the first phase relative to sending them with the second mailing (72% for Dual 2nd compared to 67.4% for Dual All). We wanted to better understand the characteristics of the households that responded to the screener and their participation in the second phase to determine if the treatment paths were bringing in different types of respondents. Since the screener only captured basic information on the children in a household, it is necessary to look at data from the topical questionnaire to learn more about household characteristics. Table 5 shows selected characteristics from topical respondents in the NS by treatment path. Although the Dual 2nd path had the highest overall screener response rate, it had a lower topical response rate than the Dual All path (68% vs. 82%). There was some indication that holding the Spanish form back

to the second mailing changed the profile of respondents. A higher percent of Dual 2nd respondents were white (77% vs. 56%) when compared with the Dual All path respondents. Not surprisingly, dual 2nd path respondents were less likely to report Spanish as parent's primary language than did Dual All path respondents (4% vs. 10%).

Table 5: Characteristics of topical respondents in National sample by treatment path (self-report)

	National sample	
	Dual 2nd	Dual All
 Screener Final Mail Response Rate 	72.0	* 67.4
 Percent of Households with Eligible Children 	32.9	35.9
 Topical Response Rate 	67.7	** 82.0
 Reported characteristics of topical responders 		
Race of Parent 1 or Mom		
White	77.4	** 56.2
Non-White	20.2	! ** 39.0
Two or more races	1.2 ‡	1.9 ‡
Missing	1.2 ‡	2.9 ‡
Language spoken of Parent 1 or Mom		
English	82.1	73.3
Spanish	3.6 ‡	* 9.5 ‡
A language other than English or Spanish	6.0 ‡	2.9 ‡
English and Spanish equally	1.2 ‡	2.9 ‡
English and another language equally	2.4 ‡	3.8 ‡
Missing	4.8 ‡	7.6 ‡
Highest level of education of Parent 1 or Mom		
High school or Less than High school	20.2	! 19.0 !
Some College	36.9	! 35.2
Bachelor or higher	41.7	42.9
Missing	1.2 ‡	2.9 ‡
House arrangement		
Owned	75.0	72.4
Rented	23.8	! 23.8 !
Other arrangement	1.2 ‡	3.8 ‡
Total Household Income		
Less than \$40,000	29.8	! 36.2
\$40,000-60,000	16.7	! 17.1 !
\$60,000-75,000	15.5	! 7.6 ‡
\$75,000+	38.1	39.0

Note:

* p < 0.10 significance using two-tail test

** p < 0.05 significance using two-tail test

‡ Reporting standards not met (too few cases).

! Interpret data with caution (estimates are unstable).

Rounds to zero

Table 6 shows the results for the targeted Spanish samples. In the LI and HS subsamples, the Bilingual and Dual treatment groups were combined into one ‘Spanish’ group for comparison. This combined group had higher topical response rates than the English only (LI: 71% vs. 64%, HS: 71% vs. 66%) and higher rates of Spanish as the parent’s primary language (LI: 41% vs. 32%, HS: 32% vs. 24%). In the LI sample, the combined Spanish offering had more Hispanic respondents (75% vs. 64%) and fewer Asian/Pacific Islander respondents (5% vs. 9%) than the English only path. Within the LI sample, topical responders in the English only path were more likely to have a Bachelor’s degree or above (23% vs. 15%) and less likely to be renting their home (46% vs. 53%), as well as a higher household income level (75,000 or more 15% vs. 9%). Within the HS sample, responders in the English only path had higher income levels than the responders in the combined bilingual/dual all path (Less than 40,000 43% vs. 49%) and were more likely to report that a parent was 2 or more races (.3% vs. 1%).

Table 6: Characteristics of topical respondents in Spanish sample by treatment path (self-report)

	Linguistically Isolated		Hispanic Surname	
	English only	Combined Bilingual and Dual	English only	Combined Bilingual and Dual
 Screener Final Mail Response Rate 	51.6 *	54.1	59.6 *	62.1
 Percent of Households with Eligible Children 	43.3	44.9	50.4	51.0
 Topical Response Rate 	64.2 **	71.3	66.2 *	70.5
 Reported characteristics of topical responders 				
Race of Parent 1 or Mom				
White	12.6!	9.7	21.9	19.7
Black	6.3!	6.7	1.3‡	1.8!
Asian/Hawaiian/Pacific Islander	8.6! *	4.5!	5.6!	3.4!
American Indian/Alaskan	0.6‡	0.2‡	1.3‡	0.5‡
Hispanic	64.4 **	74.6	65.2	70.1
Two or more races	1.7‡	0.8‡	0.3‡*	1.1!
Missing	5.7!	3.6!	4.3!	3.4!
Language spoken of Parent 1 or Mom				
English	39.1	32.5	53.3 **	46.8
Spanish	31.6 **	40.5	24.2 **	32.2
A language other than English or Spanish				
English and Spanish equally	3.4‡	3.3!	1.7‡	1.4!
English and another language equally	9.8! *	14.4	10.6!	11.2
Missing	4.6‡	2.0!	1.7‡	1.3!
Missing	11.5! *	7.4	8.6!	7.1
Highest level of education of Parent 1 or Mom				
High school or Less than High school	48.3	54.0	42.1	41.5

Some College	23.0	27.2	31.5	30.2
Bachelor or higher	23.0 *	15.2	22.2	24.9
Missing	5.7!	3.6!	4.3!	3.4!
House arrangement				
Owned	50.6	45.3	66.2	67.6
Rented	46.0 *	53.3	31.5	31.2
Other arrangement	3.4‡	1.3‡	2.3‡	1.2!
Total Household Income				
Less than \$40,000	58.6 **	73.3	43.0 *	49.1
\$40,000-60,000	19.0 **	12.4	20.2	17.6
\$60,000-75,000	7.5!	5.4	11.3! *	7.9
\$75,000+	14.9! **	8.9	25.5	25.4

Note:

* p < 0.10 significance using two-tail test

** p<0.05 significance using two-tail test

‡ Reporting standards not met (too few cases).

! Interpret data with caution (estimates are unstable).

Conclusions:

We attempted to answer three questions in this study:

- 1) What is the best questionnaire (dual Spanish/English questionnaires or bilingual) to reach Spanish-speaking households;
- 2) Is there a ‘backfire’ effect whereby response in English-speaking households is suppressed by the inclusion of Spanish language materials;
- 3) What is the most effective mailing strategy to maximize identification of Spanish speakers?

The analysis indicated that there was not a strong difference between the bilingual or dual questionnaire approaches in terms of response rate. This is consistent with the findings of De La Puente and Wobus (1994). The dual form approach yielded a slightly higher identification of Spanish-speaking households in the Hispanic surname subsample. Sending the Spanish form with the first mailing rather than later mailings also led to increased identification of Spanish-speaking households.

The results were less clear on the potential for a ‘backfire’ effect in which sending Spanish materials suppresses response among some households. Based on the findings of other studies (e.g. Bouffard and Tancreto, 2006 and Govern and Reiser, 2008), we would expect that offering Spanish language forms would have an additive effect on response rates. That is, all of the households that would have responded in English continue to respond plus households that only speak Spanish can now respond. This would lead to a higher response rate in the Dual All treatment group. The Dual 2nd treatment group was designed to allow those respondents who would be turned off by the inclusion of a Spanish form to respond in the first mailing and those who needed a Spanish form to participate to respond in the second mailing. In the NHES, sending Spanish forms to all respondents at first screener mailing in the national sample (Dual All) reduced screener response rates compared to the group that received them at the

second mailing (Dual 2nd). There were some differences in respondent characteristics between these two treatment paths as well. The difference in response rate and respondent characteristics could be indicative of a potential backfire. However, this reduction in screener response rate was overcome by a higher second phase response rate among Dual All respondents compared to Dual 2nd. The difference in respondent characteristics between the two treatment paths was minimal. If there is any backfire in the NHES, it is likely small and overcome by the higher eligibility rate and second phase response rate obtained by the Dual All treatment.

Often researchers target high linguistically isolated areas for distribution of bilingual forms. This research demonstrated that respondents outside of high linguistically isolated areas would complete forms in Spanish. The results showed that addresses where sample frame data indicate that the household has a Hispanic surname were likely to complete forms in Spanish.

References:

- Bouffard, J., and Tancreto, J. G. 2006. "2005 National Census Test: Bilingual Form Analysis." 2010 Census Test Memoranda Series, No. 22, September 15. U.S Census Bureau Washington, DC.
- De la Puente, Manuel and Peter Wobus. 1994. "An Item Nonresponse and Log-Linear Analysis of the Spanish Language Forms Availability Test." 1994 Proceedings of the Section on Survey Research Methods, Volume II pp.583-588. American Statistical Association, Alexandria, VA.
- Govern, Kelly A. and Courtney N. Reiser. 2008. "Bilingual Census Questionnaire Design Test 2007 National Census Test." Proceedings of the Joint Statistical Meetings, American Statistical Association, pp. 1604-1611.
- Shin, Hyon B. and Robert A. Kominski. 2010. "Language Use in the United States: 2007, American Community Survey Reports." ACS-12. U.S. Census Bureau, Washington, DC.
- Wobus, Peter and Manuel de la Puente. 1995. "Results From Telephone Debriefing Interviews: The Census Bureau's Spanish Forms Availability Test." 1995 Proceedings of the Section on Survey Research Methods, Volume II pp.1040-1045. American Statistical Association, Alexandria, VA.
- Zukerberg, Andrew, and Daifeng Han. 2010. "Impact of offering a bilingual option in a mail survey of linguistically isolated areas: Results from the 2009 National Household Education Survey pilot." Paper presented at the 65th Annual Meeting of the American Association for Public Opinion Research, Chicago, IL.