Extreme Response Style among U.S. Hispanics vs. Non-Hispanics

Martin Cerda[*], Ilgin Basar[†]

It is well documented that U.S. Hispanics tend to give extreme rating scale responses more so than white non-Hispanics (Marin, Gamba, and Marin 1992). However, there is no agreement on how to best measure extreme response style (ERS) among Latinos and control for its effects in the data analysis and interpretation. Furthermore, there are a small number of documented cases and academic papers written on the subject, and the few related to U.S. Hispanics are very limited or flawed from a study design perspective.

To shed new light, Encuesta, Inc. carried out a small scale experiment within a nationally representative study conducted for its *pro bono* non-partisan Americanos Poll series. The study was conducted in such a manner as to obtain proper representation of all U.S. Hispanics by language usage and acculturation level. The findings will provide evidence of extreme response style and highlight differences and possible solutions when conducting research among Hispanics.

OBJECTIVES

Given these common survey research challenges and a review of findings from numerous sources, a study was designed that explored topics related to:

- measuring ERS using various methods,

- determining how Hispanic and non-Hispanic subjects respond to a rating scale, specifically how they map their somewhat "elastic" individually held subjective categories onto the response categories provided, and

- stimulating thought on how to lessen ERS, acquiescence response style (ARS) and other "response set effects" by using proper question and rating scale design and the effective presentation and rotation of rating scales to reduce unwanted effects.

[*] **Institution:** Encuesta Inc.

[†] **Institution:** Encuesta Inc.

METHODOLOGY

A nationally representative telephone survey was conducted with n=358 U.S. Hispanic and n=302 non-Hispanic (all other races and ethnicities in a representative mix) adults. Field interviewing was between November 6 and November 27, 2009. A proprietary random probability landline (household frame random digit dialing (RDD) and listed Hispanic last name, n=443) and cell phone (RDD, n=217) frame sampling approach was used. The survey had two phases, the initial contact phase where the "test" questions were administered and a re-contact phase where "retest" questions were administered as part of a brief survey with the same respondent.

Interviews were conducted in English or Spanish, according to respondent preference. Of all interviews conducted among Hispanics, 56 percent were in Spanish and 44 percent were in English.

Of Hispanics interviewed, 46 percent were Spanish-Dominant, 39 percent were Bilingual, and 15 percent were English-Dominant according to the Encuesta, Inc. / Marin Acculturation Scale, a scale based on a series of questions that determine language usage in different situations.

The cumulative AAPOR RR3 for the household tier was: Hispanic 0.197, non-Hispanic 0.198, and cell phone tier 0.152. The cumulative AAPOR COOP3 for the household tier was: Hispanic 0.415, non-Hispanic 0.382, and cell phone tier 0.322. Overquotas were treated in the quota filled disposition as part of the calculations. No weighting was applied as a quota sample approach was employed.

DESCRIPTION OF THE EXPERIMENT

The interviewing process had two phases, testing and retesting of six experimental "bedrock" questions (i.e. opinions should not shift over a brief period of time). Each question represented a different construct (e.g., importance of family) with a unique single item rating scale used for each question (i.e. a mix construct-specific and commonly used response options).

For the test phase, questions related to several topics were asked with a mixture of 5-point Likert (with each point anchored plus a don't know/refused category which was not read) and 10-point end-point only anchored scales. This approach would help reduce "response set effects" as the use of same response format was avoided (Hui and Triandis 1985).

An example of a test (Q.E) and retest question (Q.1) is presented below in Table 1 for the single item construct we called "family".

Table 1 Example of a test and retest question.

Test Question
(READ: Now I am going to ask you some questions regarding different aspects of your life.)

(SCREENER)
(ROTATE Q.E–Q.H)
E. How important is having a close relationship with your family? Would you say it is...?
(ACCEPT ONE RESPONSE)

EXTREMELY IMPORTANT
VERY IMPORTANT
SOMEWHAT IMPORTANT
NOT VERY IMPORTANT
NOT AT ALL IMPORTANT
[DON'T KNOW/REFUSED]

RETEST QUESTION

(ROTATE SERIES A & B)
SERIES A
(ROTATE Q.1–Q.4)

1. How important is having a close relationship with your family? Using a scale from 1 to 10, where 1 means "not at all important" and 10 means "extremely important", please tell me how important having a close relationship with your family is.
1=NOT AT ALL IMPORTANT
10=EXTREMELY IMPORTANT
99=[DONT KNOW/REFUSED]

Within three days of the initial interview, the retest phase was conducted where the respondents were re-contacted and a brief interview was administered.

In the retest phase, a sub-sample of the total sample used for the study was randomly divided into two key cells (see Table 2):

Table 2 Sample cell.

	Total Sample	
	Hispanics (A)	Non-Hispanics (B)
Base: Total Respondents	n=358	n=302
Base: Control Cell Only	n=100	n=100
Base: Experimental Cell Only	n=146	n=102

- The control cell – This cell was given the exact same six "bedrock" questions with the exact scales from the test phase. This allowed assessing the reliability of the six "bedrock" questions (see Table 3).

- The experimental cell – This cell was given a variation of the original six "bedrock" questions with a different rating scale. Specifically, some of the retest questions were converted to 5-point and some to 10-point scales (see Table 3).

Table 3   Question design.

Constructs	Control		Experimental	
	Test (A)	Retest (B)	Test (C)	Retest (D)
Family	5-PT	5-PT	5-PT	10-PT
Breakfast	5-PT	5-PT	5-PT	10-PT
Universe	5-PT	5-PT	5-PT	10-PT
Pets	5-PT	5-PT	5-PT	10-PT
Health	5-PT	5-PT	5-PT	10-PT
Financial	10-PT	10-PT	10-PT	5-PT

Black measures are based on 5-point rating scale and light gray on 10-point rating scale data.

Each cell was further divided by Hispanic and non-Hispanic subjects. In most cases, a random assignment procedure that would yield equivalent groups of subjects per cell was used. However, in some cases the Hispanic and non-Hispanic group measures from both the control and experimental cells were not identical by all demographic variables, likely a product of the small sample sizes used per cell.

In addition, a complete cross-correlation analysis of the six constructs used indicated few notable inter-item (i.e., or construct in this case) correlations that might have yielded an unwanted effect. In other words, responses for the constructs were not correlated with one another.

RATING SCALE RELIABILITY AND IMPACT OF CHANGES

The control cell was used to assess the reliability of the test rating scale series. This was favorably indicated as there was a high degree of correlation between the test and retest means and distributions when using the exact same question and rating scale.

An analysis of the control cell indicated similar use of the rating scales (see Table 4). For example, when we look at the control cell in the retest phase, the raw means for the breakfast, universe, health, and financial constructs between Hispanics and non-Hispanics are similar (i.e., there are no significant differences). Also, there were no strong signs of ERS among Hispanics. For example, in the control cell, among Hispanics, while the raw mean for the family construct was 4.70 in the test phase and 4.54 in the retest phase, the raw mean for the universe construct was only 3.04 in the test phase and 3.17 in the retest phase (i.e., the raw mean for the universe construct was closer to a mid-point rather than any of the extreme points such as 1 or 5. The control cell was used to assess the reliability of the test rating scale series. This was favorably indicated as there was a high degree of correlation between the test and retest means and distributions when using the exact same question and rating scale.

Table 4 Rating scale raw means and correlation analysis.

	Control				Experimental		
	Hispanic		Non-Hispanic		Hispanic		Non-Hispanic
	Test (A) n=100	Retest (B) n=100	Test (C) n= 100	Retest (D) n=100	Test (A) n=146	Retest (B) n=146	Test (C) n=10
Mean (raw)							
Family	4.70	4.54	4.58	4.75B	4.77	9.88D	4.73
Breakfast	4.13C	4.12	3.86	3.96	4.24C	8.87D	3.79
Universe	3.04	3.17	3.28	3.46	3.03	5.97	3.57
Pets	4.29	4.04	4.44	4.55B	4.09	8.01	4.52
Health	3.50	3.42	3.68	3.61	3.08	7.69D	3.55
Financial	5.62	5.60	5.85	5.92	4.90	2.24	5.40
Correlation (T vs. RT)							
Family	→	0.191*	→	0.169*	→	0.082	→
Breakfast		0.388***		0.433***		0.466***	
Universe		0.545***		0.507***		0.557***	
Pets		0.542***		0.416***		0.615***	
Health		0.451***		0.740***		0.570***	
Financial		0.546***		0.501***		0.367***	

Black measures are based on 5-point rating scale and light gray on 10-point rating scale data.

Letters denote significant difference at a 0.100 level (2-tailed).

Paired sample correlation or t-test is significant at the *0.100, **0.050, ***0.001 level (2-tailed).

The use of a 10-point scale on the experimental cell as part of the retest phase indicated a higher degree of use of higher points of the rating scale among Hispanics compared to non-Hispanics. In other words, when asked with a 10-point scale, the raw mean for the family, breakfast, and health constructs were significantly higher among Hispanics compared to non-Hispanics. This was not evident in the case of the retest phase use of the 5-point scale. On the contrary, when asked with a 5-point scale, the raw mean for the financial construct was significantly lower among Hispanics compared to non-Hispanics.

It was apparent that Hispanics were more prone to ERS when using the 10-point scale (see Table 5). This was evident when all of the five constructs (i.e., family, breakfast, universe, pets, and health constructs) that were asked with a 5-point rating scale in the test phase were asked with a 10-point rating scale in the retest phase. For all these constructs, the mean point shift (the difference between the mean in the retest phase and the mean in the test phase) was higher among Hispanics compared to non-Hispanics. However, this was not seen in the financial construct, the only construct that was asked with a 10-point rating scale in the test phase and a 5-point rating scale in the retest phase. The mean point shift was similar among Hispanics and non-Hispanics for this construct.

Table 5   Rating scale mean point shifts.

	Control				Experimental		
	Hispanic		Non-Hispanic		Hispanic		Non-Hispanic
	Test (A) n=100	Retest (B) n=100	Test (C) n=100	Retest (D) n=100	Test (A) n=146	Retest (B) n=146	Test (C) n=102
Mean point shift							
Family	→	-0.16*	→	+0.17**	→	+5.11^	→
Breakfast		-0.01		+0.10		+4.63	
Universe		+0.13		+0.18		+2.94	
Pets		-0.25**		+0.11		+3.92	
Health		-0.08		-0.07		+4.61	
Financial		-0.02		+0.07		-2.66	

Black measures are based on 5-point rating scale and light gray on 10-point rating scale data.

Paired sample correlation or t-test is significant at the *0.100, **0.050, ***0.001 level (2-tailed).

^ indicates not stat tested as rating scale was changed from 5-point to 10-point.

EVIDENCE OF ERS

A commonly used method of measuring ERS was employed across both cells and corresponding phases of the experiment, simply counting the number of times a subject uses the extreme end of a scale within any given response item.

Several examples of ERS were noted in the Hispanic experimental cell compared to the non-Hispanic experimental cell (see Table 6). In other words, a higher proportion of Hispanics used the extreme end-points of the 10-point scale (i.e. top box rating of "10" or bottom box rating of "1") as part of the retest phase. Specifically, 9 percent of the Hispanics surveyed used an extreme end of the scale in all five possible cases, while the comparable number among non-Hispanics was 3 percent. The evidence of ERS was more acute among Hispanics when looking at the proportion of subjects that used an extreme end of the scale in either four or five out of the total five possible cases (30 percent among Hispanics compared to 18 percent among non-Hispanics). More important, no clear proof of ERS was noted in the Hispanic control cell compared to the non-Hispanic control cell when the 5-point scale was used (i.e., based on counts in all five possible cases).

Table 6 Measuring ERS based on extreme end-point use.

	Control				Experiment		
	Hispanic		Non-Hispanic		Hispanic		Non-Hispanic
	Test (A) n=100	Retest (B) n=100	Test (C) n=100	Retest (D) n=100	Test (A) n=146	Retest (B) n=146	Test (C) n=l46
Scale type (all five cases)	5-PT	5- FT	5-PT	10-PT	5- FT	5-PT	5-PT
5/5 cases	1%	2%	1%	9%H	2%	1%	2%°
4/5 or 5/5 cases	11%	7%	12%	30%H	12%	18%B	16%

Black measures are based on 5-point rating scale and light gray on 10-point rating scale data.

Letters denote significant difference at a 0.100 level (2-tailed).

Note: This table is based to five cases (not six) as this was the number of times the exact scale type was repeated by cell.

In summary, the experiment indicated that moving from a 5-point scale to a 10-point scale among the same subjects over a period of time (i.e., test vs. retest) led to a significant increase in ERS in spite of moving to a scale where subjects were asked to map their "elastic" individually held subjective categories from five test phase response categories onto the ten response categories provided as part of the experimental cell retest phase.

CONCLUSIONS

Based on this experiment, there are indications that 5-point Likert rating scales are more appropriate than 10-point anchored end points rating scales for use among Hispanics in order to improve variability and ease comparisons with non-Hispanics. More sophisticated methods of post-hoc adjustment, such as Greenleaf's (1992) ERS correction factor, are needed that when applied will preserve valuable construct "signals" while considerably eliminating extreme response style "noise". It is likely that we will not be able to "completely" remove ERS effects among Hispanics. However, below are some ideas worth exploring to possibly adjust for ERS effects and thus improve comparisons between Hispanics and non-Hispanics.

One possible solution is to develop "bedrock" or "neutral" questions to measure the extent of ERS among Hispanics, and subsequently, adjust Hispanic levels accordingly so that they are impartially comparable with non-Hispanic levels. Another potential solution is to try to develop rating scales for specific uses or constructs, rather than developing one universal rating scale to be used for all survey topics. It is likely that both aforementioned potential solutions will be most valuable on large scale or longitudinal studies.

Based on these findings, a more robust experiment is planned with larger sample sizes and mix of modes including interviewer-led surveys (telephone) and self-administered surveys (mail and online). Continued industry efforts are needed to develop an appropriate series of items and corresponding approach

for use in Hispanic marketing and opinion research within the United States.

REFERENCES

Greenleaf, Eric A. 1992. "Measuring Extreme Response Style." *Public Opinion Quarterly* 56 (3): 328–51.

Hui, Harry C., and Harry C. Triandis. 1985. "The Instability of Response Sets." *Public Opinion Quarterly* 49 (2): 253–60.

Marin, Gerardo, Raymond J. Gamba, and Barbara V. Marin. 1992. "Extreme Response Style and Acquiescence among Hispanics." *Journal of Cross-Cultural Psychology* 23 (4): 498–509.