

## Survey Practice This Month

---

Thursday, October 30, 2008, 3:55:18 PM | Editor

This month, we try something new - a website review. Tim Johnson reviews a [website](#) that contains a large amount of top-quality information on cross-national, cross-cultural survey methods. Four experts (Gene Lutz, Tom Guterbock, Molly Longstreth, Bob Craddock) provide information about their experiences with [cell phone samples](#) in the "Ask the Experts" column. Our third feature is a thought piece from a group that is attempting to develop [alternative methods](#) to measure survey representativeness. Rui Wang and her team summarize a large project that analyzes [Asian and Asian-American response patterns](#). Sid Groneman uses a variety of datasets to look at [American attitudes towards Jews](#).

### A word about WordPress

We are using blog software - WordPress - to publish Survey Practice. We (the editors) chose WordPress because it is free (except for some purchased upgrades) and relatively easy to use. Two editors already had WordPress blogs. However, we find some limitations, especially in displaying tables and charts. Each month, we learn a little more about how to use the software. If you find some display problems or have suggestions for improving the display, please send them to us.

### Your evaluation of Survey Practice so far

We designed Survey Practice so that it could be flexible and not lock us into a format. We continue to experiment with types of articles and a variety of content. If you have a minute, please send us some feedback on the first three issues. You can post a comment at the bottom or send the comments to [survprac@indiana.edu](mailto:survprac@indiana.edu).

John Kennedy     Diane O'Rourke  
David Moore     Andy Peytchev

Comments: 0

 Comments

## A Review of the Cross-Cultural Survey Guidelines Website

---

Thursday, October 30, 2008, 3:50:03 PM | Editor

By Beth-Ellen Pennell, Kirsten Alcser and Janet Harkness  
Reviewed by Tim Johnson, Director, Survey Research Laboratory, University of Illinois at Chicago

Earlier this year, a set of best practices for the conduct of cross-national and cross-cultural survey research were announced during the Berlin International Conference on Survey Methods in Multinational, Multi-regional, and Multicultural Contexts (<http://www.3mc2008.de/>). Available at: <http://ccsg.isr.umich.edu/>, these recommendations are intended to provide guidance to researchers who might be developing plans for cross-national, and other forms of cross-cultural, survey research.

The conduct of cross-cultural survey research has rapidly grown in importance over the past several decades, and

there are an increasing number of ongoing multi-national collaborative survey studies now being fielded on a regular basis by governmental, business and academic consortiums. With this increased activity, practical knowledge and experience are also being accumulated, although there have been few attempts (although see Harkness, van de Vijver & Mohler, 2003) to systematically organize this information and make it available to the larger survey research community, until now.

This project was no small undertaking. The efforts of working groups who contributed to these guidelines had been ongoing for several years. They grew out of a CSDI (Comparative Survey Design and Implementation) initiative spearheaded by Beth-Ellen Pennell (University of Michigan) and Janet Harkness (University of Nebraska & ZUMA, Mannheim, Germany). Thirteen additional authors from multiple institutions are also credited with having contributed to these guidelines, which have been further vetted by an impressive list of reviewers.

The guidelines are organized into a series of modules that cover the various steps necessary when developing, conducting and processing data from a cross-cultural survey. Some of these modules deal with topics that will be very familiar to survey practitioners (e.g., sample design, questionnaire design, pretesting), albeit tailored to address cross-cultural considerations. Other modules, in contrast, focus directly on the additional methodological and operational challenges that must be confronted when conducting survey research across cultures (e.g., translation, adaptation, harmonization of survey and statistical data).

Although different in format, these various modules can be thought of as independent book chapters (although structured very differently) that enable readers to quickly locate the specific information of interest. Thus, while anyone contemplating a cross-cultural research project would greatly benefit from a thorough familiarity with each of the recommendations contained in this document, those needing advice on very narrow matters should be able to find what they are looking for very quickly.

The integrative nature and complexity of cross-cultural survey work is clearly recognized here. There is something in these guidelines for all collaborators in cross-cultural surveys. The needs of survey methodologists, survey technicians, and survey administrators are all addressed. While most of the available survey literature tends to focus on methodological topics, and there are some good technical manuals available, there are relatively few easily accessible documents that directly consider the administrative aspects of survey research practice. And, as with all other aspects of cross-cultural survey research, their administration is also likely to be far more complicated and challenging than a typical mono-cultural project. It would have been very easy to overlook this all-too-often-ignored topic when constructing these guidelines. The organizers are to be commended for acknowledging its obvious importance to successful cross-cultural collaborations.

Each module is organized in a practical format that directly confronts relevant issues. Following a brief introduction, guidelines are immediately presented. A clear rationale is formally presented for each, followed by the specific procedural steps necessary for its implementation. Each guideline concludes with a brief section on "lessons learned." Although the specific format varies some across the modules, these lessons typically present brief case study examples and/or tips from the experts regarding potential obstacles that researchers may encounter when following each recommendation. By themselves, these lessons learned make for excellent reading. Within the context of each module, they highlight the importance of each guideline and help the document avoid the common pitfall of reading like a dry checklist of required activities.

The guidelines are also carefully documented, as each module includes appropriate references and a glossary of relevant terminology. Some modules, such as the one on Survey Instrument Design, also contain one or more appendices. Others contain additional lists of recommended reading on specific topics. The authors have provided a valuable service, all-the-more-impressive in that it is freely accessible to all on the web, rather than being available only as an expensive text.

Some might challenge the need for, or appropriateness of, cross-cultural guidelines, citing the principle that none have the right to dictate standards of conduct to researchers working in other cultural contexts (other critics might go even further and argue that survey research itself is an inherently Western methodology that is in some instances systematically imposed on less-quantitative cultures - this equally important issue can be addressed some other day). I would like to briefly defend them, however, by first noting that the authors of these guidelines do continually acknowledge that various elements of the survey process may well need to be adapted, albeit in a manner that preserves the scientific rigor of the cross-cultural elements of the survey, to local cultural conditions. As Thomas Kuhn (1962) has observed "men whose research is based on shared paradigms are committed to the same rules and standards for scientific practice." These guidelines adhere to the spirit of Kuhn's basic principle while also respecting and valuing the cultural diversity that drives this form of research.

In summary, the guidelines provide a valuable service to the profession, well above-and-beyond what is currently available in the published literature. Even experts in the field of survey methodology and/or cross-cultural survey research practitioners will find new and useful information here. For persons looking for a basic introduction to the practice of cross-cultural survey research, this is also a great place to start.

This document is an ongoing project, and a few sections remain incomplete as of October 2008. It currently exceeds 300 pages in length and will hopefully continue to be updated as new knowledge accumulates. Some may view the length as a barrier to use, but as suggested earlier, the various modules are stand-alone pieces and each is actually quite reasonable in length (none currently exceeds 20 pages in length). I would strongly urge readers to take advantage of this valuable resource, which concentrates in one place, arguably for the first time, a set of recommended practices for collecting cross-cultural survey data.

## References

Harkness, Janet, van de Vijver, Fons J.R. & Mohler, Peter Ph. (2003). *Cross-Cultural Survey Methods*. New York: Wiley.

Kuhn, Thomas. (1962). *The Structure of Scientific Revolution*. Chicago: University of Chicago Press.

Comments: 0

 Comments

## Cultural Differences: Why Do Asians Avoid Extreme Responses?

---

Thursday, October 30, 2008, 3:44:38 PM | Editor

Rui Wang, Survey Sciences Group  
Brian Hempton, Survey Sciences Group  
John P. Dugan, University of Maryland  
Susan R. Komives, University of Maryland

The United States is a multi-racial and multi-cultural society. Social scientists conducting surveys face one problem: they are dealing with people from different cultural backgrounds. Similar to the challenge of international studies, we have to make sure that the survey measures are comparable across subpopulations with different cultures. To ensure comparability, we should consider two important factors: (1) equivalence of presenting the measures (whether the presentation of the stimuli is equivalent and comparable across different cultures); (2) and equivalence of interpreting/responding to the measures (whether respondents would interpret and respond to the stimuli in the

same way).

Research has found differences in responding to survey questions between different racial/ethnic groups. Bachman and O'Malley (1984) found Blacks were more likely than Whites to select extreme response categories, particularly the positive end of "agree-disagree" scales. Hui and Triandis (1989) found Hispanics had a stronger tendency of selecting extreme responses on 5-point scales. Marin et al. (1992) also found Hispanics tended to choose the extreme responses and agree with a given item more often than Whites did, and the level of acculturation among Hispanics affected the level of extreme and acquiescent responses, with the more acculturated Hispanics choosing these types of responses less often.

Studies that include Asians and Asian Americans suggest that they are more likely to select the midpoints and avoid extreme responses on Likert scales. Lee et al. (2002) studied a group of Chinese, Japanese, and Americans recruited at ethnic or general supermarkets, and found Chinese and Japanese selected midpoints more often on items that involved admitting to a positive emotion. Chen et al. (1995) compared response styles between East Asian and North American students, and found students from the two collectivist cultures (Chinese and Japanese) demonstrated a greater preference for midpoints and less preference for extreme values than those from the individualist cultures, especially the U.S. students. In another study of college students, Grandy (1996) found Asian American students tended to endorse middle options and avoid extreme responses on a 5-point Likert scale more than Whites did.

Social scientists do not agree on why Asians or Asian Americans would prefer middle options and avoid extreme responses, but most believe it is culturally-related. East Asian cultures emphasize the need to fit in with others and avoid conflict in society. Hoy (1993) had referred to this "aversion to the spotlight" as "cultural shyness." Johnson et al. (2005) discussed two cultural orientations: individualism and collectivism, suggesting that a middling response style should better fit the cultural norms and imperatives of persons living in collectivist cultures, since collectivism is associated with a greater emphasis on interpersonal harmony and a lesser emphasis on individual opinions (Triandis et al., 2001). Therefore, we would assume that people living or growing up in East Asian countries and those who share similar cultures tend to behave modestly and politely and maintain harmony in social relationships by avoiding expressing extreme opinions.

## Methods

The analyses were based upon data collected from the 2006 Multi-Institutional Study of Leadership. Over 170,000 students in 53 institutions across the U.S. were selected, and 55,623 students responded to the survey, including 49,283 completes and 6,340 partials.

In this on-line survey, students were asked to indicate their agreement or disagreement with 68 statements on the Social Responsible Leadership Scales (SRLS), which were designed to measure leadership outcomes across eight critical values (5-11 items per each value). The 68 items, displayed in grids of 11-12 questions per screen, were randomly re-ordered so that items of same topics would not group all-together on the same screen. For each of the 68 SRLS items, students were asked to report on a fully labeled 5-point Likert scale, ranging from "Strongly Disagree" (1) to "Strongly Agree" (5) with "Neutral" (3) as the midpoint.

A series of new variables were generated to capture respondents' reporting preferences to the 68 SRLS items, indicating whether they selected the extreme responses, the middle points, or other response options. The percentages of times a student selected middle, extreme or other response options were then calculated among the SRLS items they actually answered, as a summary of their overall response preferences.

The key independent variables include respondents' self-reported race/ethnicity, which was re-coded into five categories - Asian/Asian American, White, Black, Hispanic and others, and respondents' citizenship status - whether

they were born in the U.S. or in foreign countries. Two other demographics at the individual level, gender and class standing, were used as basic control variables.

We included two cultural-related institutional factors - the region of the institution where students study and the diversity of the institution's cultural environment, measured by the percentage of Asian/Asian American students and the percentage of foreign students in the total student population.

#### *Hypothesis and Analytic Plan:*

The first step of the analysis is to determine whether Asian/Asian American students would select the middle points and avoid extreme responses more often on the Likert scale. The second goal is to offer some explanations to the differences (if we found any) between U.S. born Asian American students and those born in foreign countries.

In addition, we will summarize the logistic regression result for each of the 68 SRLS items. Institutional characteristics will also be included in the analytic model to explore whether these factors would influence response behaviors.

### **Results**

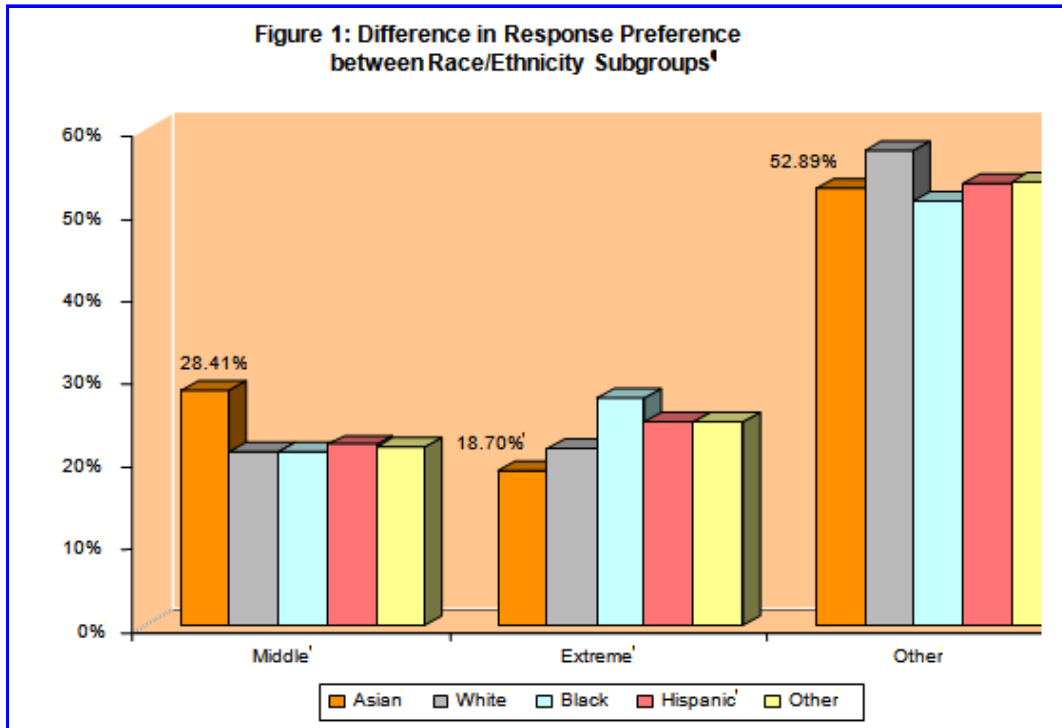
Table 1 summarizes the weighted proportions and unweighted counts of the independent variables. The percentage of Asian/Asian American students in the total student population is about 5.9% on average for 53 institutions, ranging from 0.1% to 33.8%. The percentage of foreign students is about 4.9% on average, ranging from 0.2% to 14.3%.

**Table 1: Descriptive Statistics of Key Predictors and Demographics.**

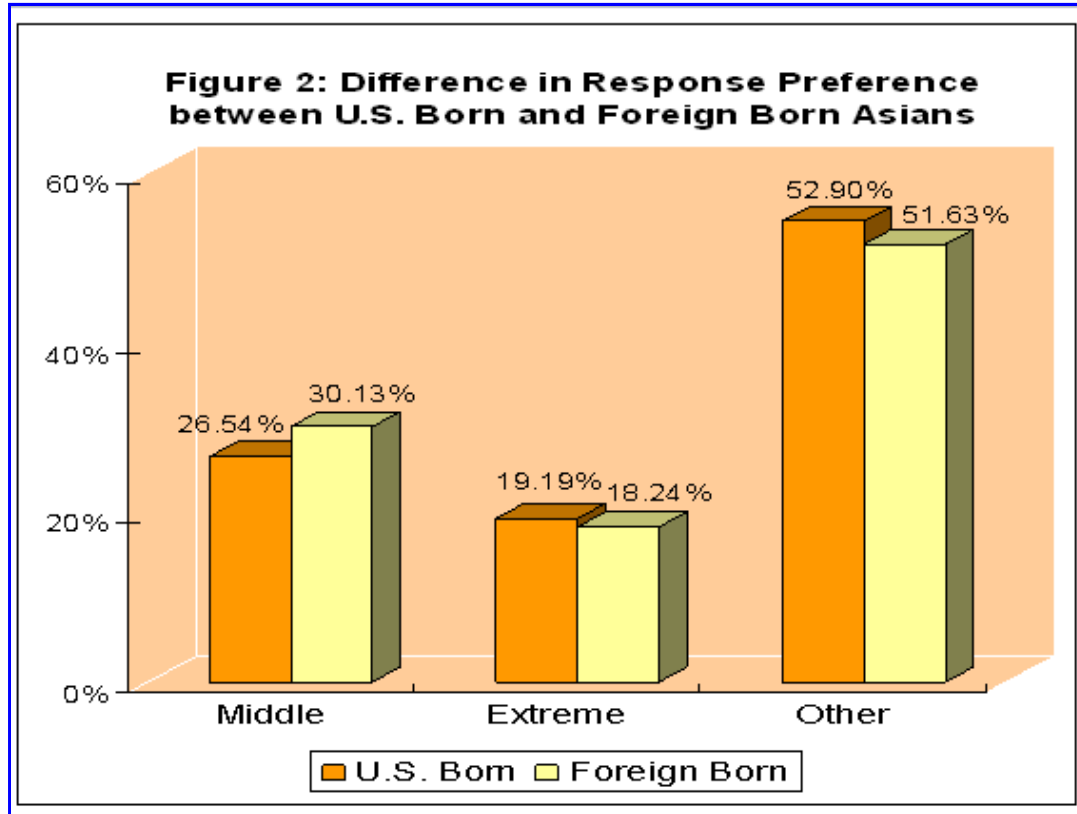
Variables	Weighted % (S.E. %)	Unweighted Count	Total Count
<b>Race/Ethnicity</b>			
Asian	8.0% (1.2%)	4191	53200
Black	6.6% (1.2%)	3145	
Hispanic	4.9% (0.9%)	2490	
Other	10.3% (0.5%)	5385	
White	70.1% (2.5%)	37989	
<b>Citizenship</b>			
U.S. born	89.1% (1.1%)	44071	49277
Foreign born	10.9% (1.1%)	5206	
<b>Gender</b>			
Male	56.6% (1.4%)	32430	54071
Female	43.4% (1.4%)	21641	
<b>Class</b>			
Freshman	23.8% (1.3%)	13190	55623
Sophomore	22.1% (0.7%)	12246	
Junior	26.0% (0.7%)	14541	
Senior	28.1% (1.3%)	15646	
<b>Region</b>			
Northeast	20.6% (6.5%)	11606	

Midwest	23.8% (6.4%)	13566	55623
South	36.1% (7.1%)	19665	
West	19.5% (5.8%)	10786	

We found Asian/Asian American students consistently reported lower values to the 68 SRLS items compared to other racial/ethnic groups (Table 2). This is attributed to the fact that Asian/Asian American students chose the middle option substantially more often and extreme options less often than the other groups (Figure 1).



Our second analytic goal is to explore whether the transition from Eastern to Western cultures would affect their response styles. We found that U.S. born Asian American students selected middle options significantly less often and the extreme responses more often than foreign born Asian students (Figure 2).



The overall response preferences for Asian/Asian American students were correlated with the density of the Asian/Asian American student population in the institution. The multivariate model indicated that Asian/Asian American students selected middle options more often and extreme options less often as the proportion of Asian/Asian American students among the student population increased.

We also conducted logistic regressions on each of the 68 SRLS items to examine students' response preference on each single measure (Table 2). Among the 68 items, the differences in their response preferences between Asian and White students were significant for 53 items (78%); the differences between Asian and Black students were significant for 65 items (96%); the differences between Asian and Hispanic students were significant for 64 items (94%); and the differences between Asian and other racial/ethnic groups were significant for 66 items (97%). The consistent pattern suggests that Asian/Asian American students favor middle options and avoid extreme responses more often than any other racial/ethnic groups, no matter what the question was asked about.

**Table 2: Summary of the Bivariate Regression Analyses on All the SRLS Measures.**

Summary	Comparison or Predictor	Same Direction as Hypothesized	Same Direction		Reverse Direction	Total Number of Items
			Significant at 0.05 Level	Significant at 0.1 Level		
Linear Regression	White vs. Asian	64	54	57	4	68
	Black vs. Asian	66	62	63	2	68
	Hispanic vs. Asian	66	61	61	2	68

	Other vs. Asian	66	61	61	2	68
Logistic Regression	White vs. Asian	64	53	56	4	68
	Black vs. Asian	68	65	66	0	68
	Hispanic vs. Asian	68	64	66	0	68
	Other vs. Asian	67	66	66	1	68
Logistic Regression among Asian Subgroup	U.S. vs. Foreign born	59	32	38	9	68
	Percent of Asian	61	53	55	7	68

Similarly, logistic regressions on 68 SRLS measures showed some evidence of the different response preferences between U.S. born Asian American students and foreign born Asian students; however, the evidence was not consistent across all the items. Although 87% of the items suggested the same direction as our hypothesis that U.S. born Asian American students were less likely to select the mid-points on the scale, only 47% were significant (Table 2).

There was also evidence that the response style of Asian/Asian American students was correlated with the diversity of the institution's cultural environment. For 78% of the items, the odds of selecting middle options for Asian/Asian American students increased as the percentage of Asian students in the total student population increased.

## Conclusion

We found respondents with different race/ethnicity had different response preferences when answering questions on Likert scales. Although our study focused on undergraduate students, the finding was consistent with previous research on the general population, which suggests that Asians or Asian Americans prefer middle options while Blacks and Hispanics prefer extreme responses. There was also evidence that U.S. born Asian American students differ in their response preferences from foreign born Asian students due to their cultural similarity to the westerners; however, the evidence was not consistent across all the Likert scale items.

We also found that response preferences of Asian/Asian American students were correlated with the cultural environment of the institution. The larger percentage of Asian/Asian American students in the institution, the more likely they would select the middle options and the less likely they would select the extreme responses. This finding supported our hypothesis that the distinctive response style of Asians/Asian Americans is culturally related. One possible explanation could be that Asian students would have more opportunities to socialize with other Asian students or be exposed to eastern cultures within institutions that have a larger Asian student population, thus their behaviors would be more like the easterners than the westerners.

Several limitations of the study should be acknowledged. Since the sample selection was implemented by each institution, we do not know the sampling rate applied by the institution to draw the simple random sample. All the analyses were conducted by ignoring the different sampling rate in each institution. However, nonresponse weights were applied to account for potential response errors, and possible clustering effects (homogeneity within each institution) were also taken into consideration in the analysis. Although we had statistically significant findings that supported our hypotheses, we were cautious to make any generalization to a larger population due to the lack of information about the sampling procedures.

The implications of the findings are more important than identifying the differences. It is crucial for survey

researchers to be aware of this distinct response style when their studies are involved with asking Asians/Asian Americans and other racial/ethnic groups to report on Likert scales. False conclusion can easily be drawn in surveys like the Leadership study, where the consistent lower scores Asian/Asian American students obtained on the SRLS items was not a good evidence of their weak leadership capacities. Therefore, we suggest survey practitioners use Likert scales with more cautions.

More efforts should be made to minimize the differences in the response preferences between different racial/ethnic groups and make the items measured by the Likert scales more equivalent across people with different cultural backgrounds. We also suggest that survey researchers, especially those who are responsible for designing and testing questionnaires, further examine (1) whether other types of questions would better capture Asians/Asian Americans' opinions than the Likert scale (2) and whether scales with more (or fewer) points or even points (no midpoint) would better capture their opinions than the 5-point scale.

### References:

Bachman, J. G., and O'Malley, P. M. "Yea-saying, Nay-saying, and Going to Extremes: Black-White Differences in Response Style." *Public Opinion Quarterly* 48.2 (1984): 491-509.

Chen, C., Lee, S., and Stevenson, H. W. "Response Style and Cross-cultural Comparisons of Rating Scales among East Asian and North American Students." *Psychology Science* 6 (1995): 170-175.

Grandy, J. "Differences in the Survey Responses of Asian American and White Sciences and Engineering Students." GRE Board Professional Report No. 93-25P, ETS Research Report 96-25: July 1996  
<<http://www.ets.org/Media/Research/pdf/RR-96-25-Grandy.pdf>>

Hoy, R. R. "A 'Model Minority' Speaks Out on Cultural Shyness." *Science* 262 (1993): 1117-1118.

Hui, C. H., and Triandis, H. C. "Effects of Cultural and Response Format on Extreme Response Style." *Journal of Cross Cultural Psychology*, 20.3 (1989): 296-309.

Johnson, T., Kulesa, P., Cho, Y. I., and Shavitt, S. "The Relation Between Culture and Response Styles: Evidence from 19 Countries." *Journal of Cross Cultural Psychology*, 36.2 (2005): 264-277.

Lee, J. W., Jones, P. S., Mineyama, Y., and Zhang, X. E. "Cultural Differences in Responses to a Likert Scale." *Research in Nursing & Health* 25.4 (2002): 295-296.

Marin, G., Gamba, R. J., and Marin, B. V. "Extreme Response Style and Acquiescence among Chinese-American Children." *Journal of Counseling Psychology*, 37 (1992): 208-212.

Triandis, H. C., Carnevale, P., Gelfand, M., Robert, C., Wasti, S. A., Probst, T., Kashima, E. S., Dragonas, T., Chan, D., Chen, X. P., Kim, U., De Dreu, C., Van De Vliert, E., Iwao, S., Ohbuchi, K-I., and Schmitz P. "Culture and Deception in Business Negotiations: A Multilevel Analysis." *International Journal of Cross Cultural Management*, 4.1 (2001):73-90.

Comments: 2

 Comments

## How Much Do Americans Like Jews?: Approaches to Social Attitude Measurement

---

Thursday, October 30, 2008, 3:34:47 PM | Editor

Sid Groeneman, Groeneman Research & Consulting, Inc.

How others regard Jews has been the focus of considerable social science research, typically conceptualizing Jews as a cultural out-group, and often motivated by a concern with anti-Semitism. The emphasis has been on measuring ethno-centrism and negative attitudes. This research maintains that reliance on direct affect measurement tools results in more comprehensive, more affirmative, and less misleading assessments than evaluations based on stereotypes or social distance measures.

### Approaches to Measuring Social Attitudes

Attitudinal research about Jews (and other minority groups) has relied primarily on three types of measures. "Social distance" measures include questions about comfort level in interactions: living in a neighborhood with Jewish residents, working with Jewish co-workers or superiors, having a close relative marry a Jew, and so forth. Negative responses to these kinds of questions have been used as indicators of dislike.

For example, the National Opinion Research Center's 2000 General Social Survey found that 9% of Americans objected to living in a neighborhood "where half of your neighbors were Jews"; and 13% were opposed to having "a close relative marry a Jewish person". A 2006 Gallup/USA Today Poll found that 5% would not like to have Jews as neighbors. Such results are taken as uniformly negative feelings. However, even these modest proportions might not be expressing negativity. The problem is that desire to avoid closeness might represent a preference to be with members of one's own ethnic, religious, or racial group (or fear of assimilation), and not antipathy.

Similar considerations about misleading or exaggerated interpretation also apply to the second method, measurement of stereotypes. Agreement with stereotypic statements about Jewish "control" of Wall Street or the media, or of negative Jewish personality traits, have been used as indicators of hostility. Like the social distance approach, ascertaining the prevalence of stereotyping has a storied tradition in social research. Some of the measures used — for example, Jews being avaricious or scheming - denote indisputably negative perceptions. Other stereotypes, such as disproportionate Jewish influence in the entertainment industry, or Jewish wealth, convey more ambiguous meanings.

In some cases, erroneous inferences can result from an absence of context: A 1996 Pew Research Center survey found 49% preferring that Jews have less political power — which, in isolation, would be a strong negative indictment. However, the same survey also found the public preferring that Catholics (53%), Evangelicals (51%), and even Protestants (44%) have less power. Jews were hardly the only target of this perception.

Even the oft discussed notion that Jews were responsible for the crucifixion of Christ does not always indicate negativity. During the height of interest in Mel Gibson's controversial 2004 film *The Passion of the Christ*, 26% responded in a Pew Research Center survey that Jews were responsible for Christ's death. Some took this as disturbing evidence of contemporary anti-Semitism. But, when ABC News asked respondents if they agree or disagree that "All Jews today bear responsibility for the death of Jesus," only 8% agreed — a sharply lower figure.

Despite disagreement over meaning and exaggeration of its significance in some instances, stereotypes of Jews are still quite prevalent, some of them held by upwards of 25-30% or more of U.S. adults. (1) That said, it is also necessary to recognize that a person can hold a negative stereotype of a group while simultaneously having overall positive feelings toward that group.

The foregoing is not to argue against all measurement of stereotypes or social distance — only that these

approaches have sometimes produced exaggerated estimates and overblown conclusions about the degree of implied negativity.

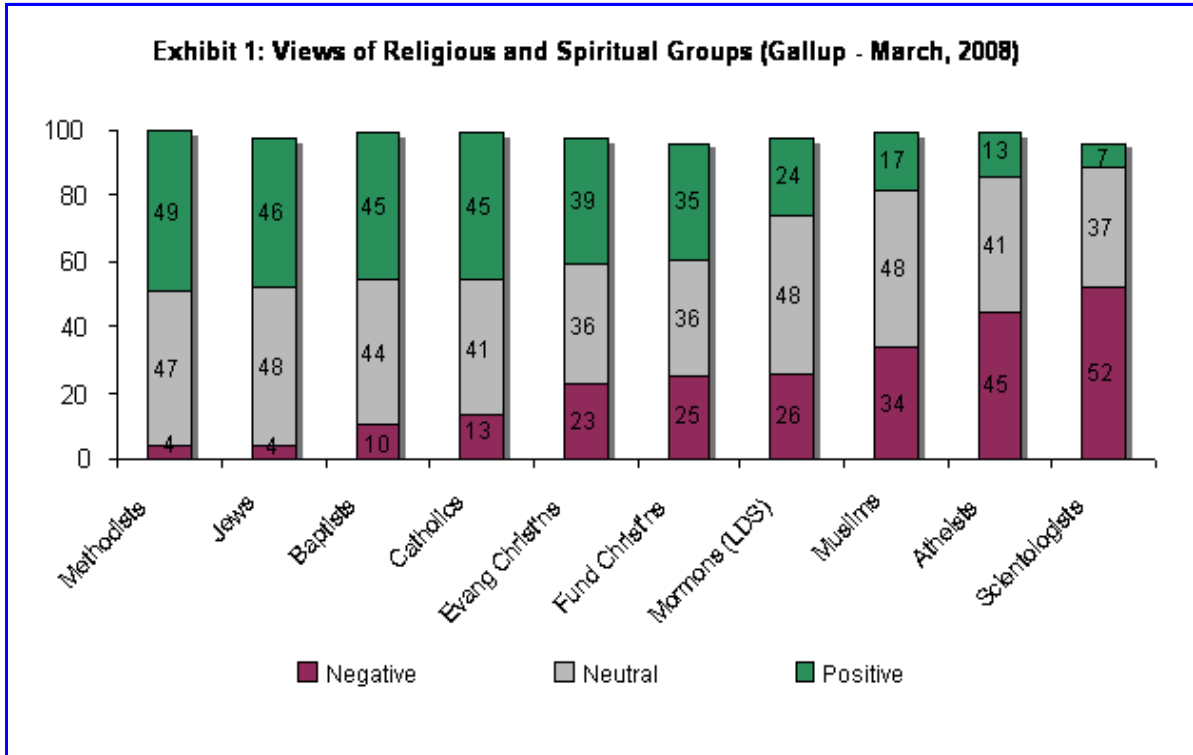
The third type of attitude research relies on purer affect measures - those which most directly gauge both positive and negative feelings. This type of measure comes in a variety of flavors, but the most common asks about favorability of opinion, as in the Pew Research Center series of surveys:

**Table 1: Favorability Toward Jews  
(Pew Research Center)**

	% Very/Mostly Favorable	% Very/Mostly Unfavorable	% Don't Know/Can't rate/Other
Spring 2008	77	7	17
March 2007	69	10	21
Spring 2006	77	5	17
May 2005	76	7	16
March 2004	77	8	15
June 2003	72	9	19
February 2002	74	9	17
March 2001	72	10	18
August 2000	77	8	15
June, 1997	82	9	9

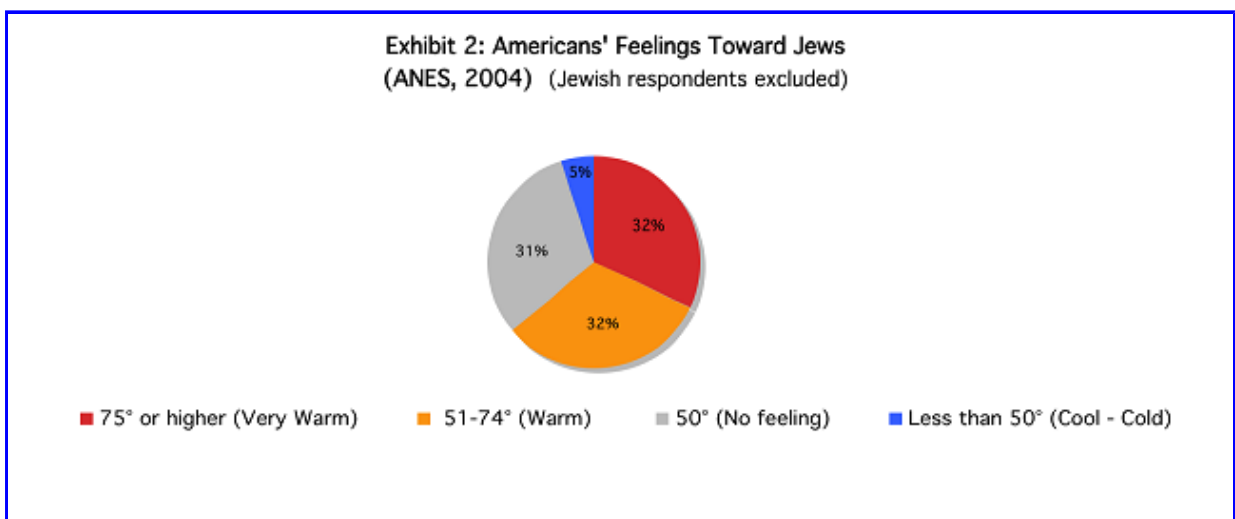
The level of unfavorable opinion toward Jews over the past decade ranges quite narrowly between 5-10%, and with no apparent trend. The fairly high proportion of "Don't know / other" responses is ambiguous. It probably signifies a combination of things: unfamiliarity with Jews, (2) unwillingness to generalize, a desire to answer neither/neutral/in-between, and, in some cases, an unwillingness to admit unfavorable views. A 2000 ABC News / Washington Post poll tried to address the latter possibility directly: "If you honestly assessed yourself, would you say that you have some negative feeling about Jews?" Those results (4% yes, 95% no, 1% no opinion) found few admitting a reluctance to disclose negativity.

Exhibit 1 displays results from a Gallup Poll taken this past Spring, which asked how positive or negative people feel about "...different religious or spiritual groups in the United States". This question offered an explicit "Neutral" response option. Compared to others, Jews are viewed quite favorably, with net positive scores only narrowly behind Methodists, slightly higher than Catholics and Baptists, and substantially ahead of six other groups.



Another type of direct affect measure, the “feeling thermometer,” has been widely applied in measuring attitudes of public figures and political/ethnic/religious groups. It ranges in one-degree increments from 0° (Very Cold or Unfavorable Feeling) to 100° (Very Warm or Favorable Feeling). With a greater number of discrete scale points than other measures, the feeling thermometer has the ability to capture finer gradations of sentiment. Like Gallup’s measure but unlike Pew’s, it has an explicit middle point at the 50° mark (labeled “No Feeling At All”), facilitating a valid reply from respondents perceiving the group neither favorably nor unfavorably.

Exhibit 2 shows the latest feeling thermometer ratings of Jews:

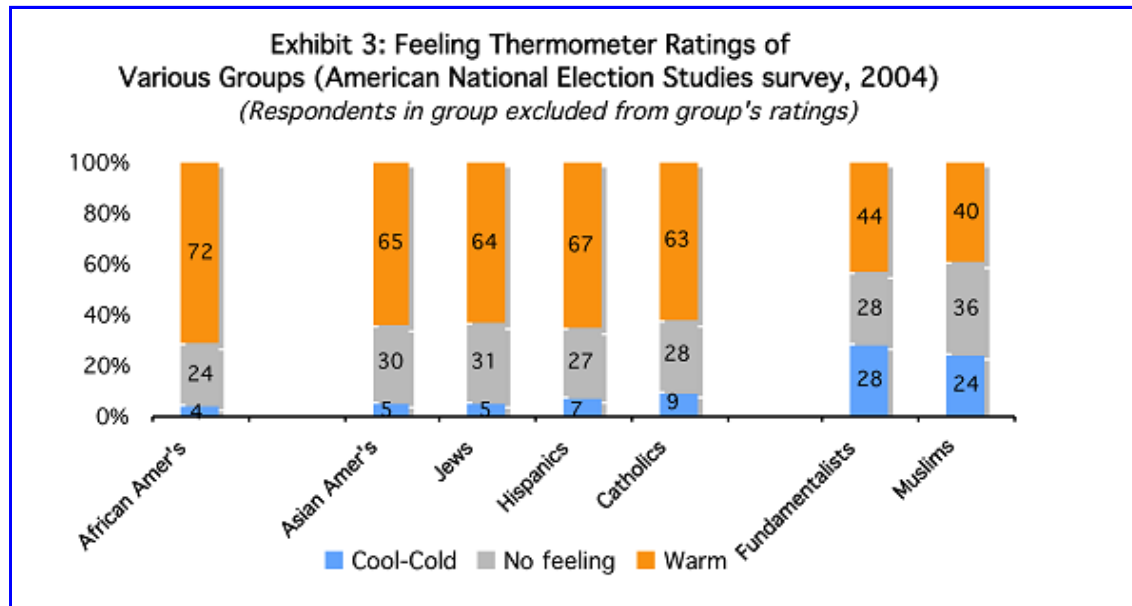


These ratings too suggest that Americans feel quite favorable toward Jews: Over half (64%) hold warm feelings (>50°) toward Jews, with half of those attitudes above 50° - or nearly one-third of the total — being quite favorable, at 75°

or warmer. Another 31% gave ratings of 50° (no feeling). Only 5% chose cool-to-cold scores (less than 50°). (3)

### Applying the Feeling Thermometer to Other Groups

Exhibit 1 suggested that Jews are viewed positively compared to other *religious* groupings. Does this conclusion hold when temperature thermometer comparisons are made with other groups in American society? Since Jews are also an ethnic or cultural group — not only adherents of a distinctive religion — it is appropriate to examine how they are viewed among a broader set of social groupings (Exhibit 3).



Among the seven groups rated in the ANES survey, Americans feel as warmly toward Jews as toward Catholics, Hispanics, and Asian Americans — and considerably more so than toward Christian Fundamentalists and Muslims. African Americans receive warmer ratings. (4) These data present a moderating counterpoint to the more strongly favorable image of Jews suggested by the previous data: While Jews are warmly regarded by most Americans, so are other ethnic minorities.

African Americans' higher thermometer score is unexpected. Except perhaps for Muslims — who constitute a small fraction of the U.S. population (5) but suffer a special burden in our highly charged post-911 atmosphere — African Americans likely encounter greater prejudice and discrimination than other groups. So, it is puzzling that they are the recipients of the most positive ratings. Might the expressed attitudes reflect compensatory sentiment?, a genuine response to the sustained teaching of the immorality of racial prejudice?, or possibly a reflection of greater reluctance to admit one's true feelings toward African Americans?

### Concluding Thoughts

Americans like Jews at least as much as other religious and ethnic/racial groups in American life, with the possible exception of African Americans. Although other research documents that certain stereotypes about Jews remain prevalent and that small proportions of adults prefer to maintain some social distance, overall attitudes are quite favorable.

This research expresses a preference for direct affect measurement over exclusive reliance on stereotypes and social

distance because it results in more accurate attitude assessments. A few questions beg further investigation. First, how many of the “neutrals” and “don’t know” responses are hiding negative feelings, and what do the answers of the rest signify — unfamiliarity?, unwillingness to compare social groups? Some form of research less intrusive than interviews might begin to answer the first question. In-depth follow-up probing might help us understand the second. As a corollary question: To the extent that some are disguising their antithetical views, is this manifested equally across different social groups, or are Americans less inclined to divulge their true feelings about some (African Americans?) than others? - an issue much on the minds of pollsters and poll watchers this election season.

### Endnotes

1. The most extensive data on stereotypes of Jews has been compiled by the Anti-Defamation League - <http://tinyurl.com/2bdgvd> - and by the American Jewish Committee - <http://tinyurl.com/8mcdq>. See also Gary A. Tobin and Sid Groeneman, *Anti-Semitic Beliefs in the United States*, Institute for Jewish & Community Research, 2003: [http://www.jewishresearch.org/PDFs/A\\_S\\_Report\\_web.pdf](http://www.jewishresearch.org/PDFs/A_S_Report_web.pdf).
2. The National Opinion Research Center’s 2000 General Social Survey found that 43% of U.S. adults reported not personally knowing any Jews.
3. There are moderate subgroup differences by age and education. Those with more formal education rate Jews more warmly (based on mean thermometer response): < HS grad: 60.1°, HS grad: 65.9°, Some college: 68.6°, College graduates: 69.6°. So do older compared to younger respondents: 18-34 (63.2°), 35-54 (65.8°), 55+ (71.2°).
4. The mean thermometer scores are: African-Americans-69.5°, Asian-Americans-67.2°, Jews-66.7°, Hispanics-66.5°, Catholics-66.0°, Fundamentalists-54.2°, Muslims-53.6°.
5. In 2002, Tom W. Smith, Director of the GSS, calculated a best estimate of the U.S. Muslim population to be 0.67% (1.9 million people) - and possibly as many as 1.0% of the population (2.8 million) if “high-side” estimates based on local surveys, figures from mosques, and ancestry and immigration statistics are given more weight than the survey-based number. Citation: “The Muslim Population of the United States: The Methodology of Estimates,” *Public Opinion Quarterly*, Vol. 66, No. 3 (Fall, 2002), p. 414.

Comments: 0

 Comments

## Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools

---

Thursday, October 30, 2008, 3:34:24 PM | Editor

R. M. Groves, J. M. Brick, M. Couper, W. Kalsbeek, B. Harris-Kojetin, F. Kreuter, B. Pennell, T. Raghunathan, B. Schouten, T. Smith, R. Tourangeau, A. Bowers, M. Jans, C. Kennedy, R. Levenstein, K. Olson, E. Peytcheva, S. Ziniel, and J. Wagner

It is increasingly clear (e.g., Keeter et al., 2000; Curtin, Presser and Singer, 2000; Groves, 2006) that nonresponse rates are poor indicators of nonresponse error of survey estimates. However, the field (and the public at large) has been taught to use response rates as quality indicators. Indeed, with the exception of the sample size, the response rate is probably commonly held to be the most important criterion of survey quality. As of yet, however, there is no alternative indicator of relevance to nonresponse error that has been proposed

We thus seem to be in a moment of uncertainty - where old procedures have been questioned, but no new options have been forwarded. The value of the response rate historically was

- a) its simplicity (a single number to characterize quality), and
- b) its transparency (it was not a complicated statistical function of sample observations).

The common perspective that higher response rates produce lower nonresponse error led to the popular use of response rates.

The plague of nonresponse research is the fact that very little information is typically available on nonrespondents. Based on recent meta-analyses it's clear that there is commonly more variation in nonresponse error across estimates of the same survey than across surveys with different response rates. For example, since

- a) the key to nonresponse error is the correlation of the survey variables with response propensities,
- b) these correlations vary as a function of covariances with the causes of propensities and the variables, therefore,

no single measure of nonresponse (at the survey level) can inform the researcher to the diverse levels of nonresponse error that various estimates might hold. This is not dissimilar to the problem the field faces with characterizing the effect of clustering on standard errors. Design effects vary across different estimates within a single survey. While no one design effect is appropriate for all estimates, sometimes average design effects are useful.

**Alternative Approaches.** To get the conversation moving, it seems that there are two "ideal-type" approaches of alternative indicators:

1. a single indicator at the survey level
2. individual indicators at the estimate level

Indicators at the survey level differ from the simple response rate computations by incorporating into the computation some key auxiliary variables that are correlates of the survey variables (or at least desirable measures on which the respondent pool should be balanced). Indicators at the estimate level are specific to a single estimate, providing some information based on variation in success of measuring groups that vary on the expected value of the estimate.

Regardless of its merits, the notion of having a separate indicator of nonresponse error for each estimate may be too large a leap in complexity for current practitioners. It seems useful, therefore, to evaluate compromises between a single indicator and estimate-specific indicators. These would be indicators of nonresponse errors for sets of estimates (ones that might have similar correlations with different auxiliary variables).

## 2. A Single Indicator at the Survey Level

A single indicator to characterize nonresponse error must perforce assert that a set of auxiliary variables measured on nonrespondents and respondents usefully capture the major correlates of *all* the survey estimates. We know this to be strictly false, but it is the smallest useful step away from response rates.

There are several that could be forwarded:

- a) Variance functions of nonresponse weights (e.g., coefficients of variation of nonresponse weights)

*The value of this would be to indicate how variable the level of nonresponse is across sample cases. It could incorporate observations on respondents and nonrespondents as part of the data collection effort. Comment: these measure variation in response propensities without linkage to the survey variables and are dependent on how rich the poststratification weights are.*

- b) Variance functions of post-stratification weights (e.g., coefficients of variation of poststratification weights)

*This would probably be most appealing when there were no other nonresponse weights based on sample quantities and population totals served as the basis of both nonresponse and coverage adjustments. These would assert that the weights were correlated with the variables of interest and that the measures of these variables in the survey were equivalent to the measures used from the population data (e.g., that education measured in a survey was equivalent to the measure of education in the decennial census). Comment: these measure variation in response propensities without linkage to the survey variables and are dependent on how rich the poststratification weights are.*

- c) Variance functions of response rates on subgroups defined for all sample cases (both respondents and nonrespondents)

*These could be coefficients of variation of response rates across key subgroups in the sample. These might result from efforts of the survey to induce interviewer observations or other design features to capture such variables, ideally correlated with the key survey variables. Comment: These indicators are generally sensitive to the size of the samples in the subgroups and the level of response rates.*

- d) Goodness of fit statistics on propensity models

*It is more and more common to build response propensity models (often logistic regression models predicting the likelihood of a complete). When these models fit the data very well, then de facto predictor variables have identified groups with very different response rates. This indicator would assert that if such a case existed, there is evidence of nonresponse bias. Comment: the value of this is probably limited to estimates using survey variables highly correlated with the predictor variables of the response propensity model.*

- e) R-indexes, which are model-based equivalents of the above

*Shouten defines a whole set of indexes, which he calls R-indexes. Among these is a marginal R-index using a standardized regression coefficient in a logistic propensity model, to indicate the degree of imbalance among respondents and nonrespondents on an auxiliary variable. There is a large class of such indexes, some overlapping with the above. The distinctive subset would posit multivariate propensity models with auxiliary variable predictors. Comment: a current challenge in these indexes is their tendency to be affected by the level of response rates as well as the variation in rates.*

<http://www.cbs.nl/en-GB/menu/methoden/research/discussionpapers/archief/2007>.

### 3. Indicators at the Level of Individual Estimates

Clearly the field would be on the soundest theoretical footing if each survey estimate produced had its own nonresponse error indicator. However, such a stand would be a giant leap in complexity for users of survey estimates. Moving to this level permits us, however, to examine various functions of the relationship between response propensities and individual survey variables. We use "y" as the designator of an individual survey item. We use the term, "auxiliary variables," to describe any attributes that are known for respondents and nonrespondents

- a) Comparisons of respondents and nonrespondents on auxiliary variables

Any variables available on the sampling frame or used for postsurvey adjustment are measured on both respondents and nonrespondents. These means on these variables for respondents can be compared to those of nonrespondents (as well as the total sample). In essence, these comparisons permit nonresponse bias estimates on such variables.

*Comment: to the extent that these variables are correlated to survey variables these indicators may be predictive of nonresponse error on the survey variables. A useful exercise might be to group survey variables by the level of correlation with the auxiliary variables, as measured among respondents, to comment on which auxiliary variables are most like various survey variables.*

- b) Correlation between post-survey nonresponse adjustment weights and  $y$ , measured on the respondent cases

All surveys that use some adjustment procedure in hopes of reducing nonresponse error of estimates possess a variable on every respondent that acts to increase the influence of underrepresented cases and decrease the influence of overrepresented cases within the respondent pool. Correlations between the resulting "postsurvey" adjustment weights and the survey variables (estimated on the respondent cases) may be informative about the relationship between response propensity and those survey variables. *Comment: the correlations may not be the same within the nonrespondent pool and thus may mislead the researcher about the extent of nonresponse bias in the adjusted and unadjusted estimates.*

- c) Examine the means of survey variable  $y$  within deciles of the survey weights

With a graphical display of mean across deciles of nonresponse weights, this is a visual equivalent of the correlation in b) above. *Comment: the variation in means across propensity groups may not be the same within the nonrespondent pool as within the respondent pool and thus may mislead the researcher about the extent of nonresponse bias in the adjusted and unadjusted estimates.*

- d) Fraction of missing information on  $y$

*This is based on the ratio of between-imputation variance of an estimate and the total variance of an estimate, based on imputing values for all the nonrespondent cases in a sample (Little and Rubin, 2002). The imputation models can be diverse, but one of some appeal is that of a sequential regression imputation, utilizing all of the variables in the data set. The percentage of missing information would be high when the percentage of variation in the total estimate due to imputation was large. It would be small when the percentage of variation in the total estimate due to imputation was small. Comment: The value of this indicator is a function of the quality of imputation model, both its variance properties and its bias properties.*

#### **4. Indicators at the Level of Sets of Variables/Estimates**

Instead of each estimate in a survey having its own indicator of nonresponse error, for convenience, sets of similar estimates could be defined, such that the set would share a value of a nonresponse indicator. These would be based on a prior analysis, most likely among the respondent cases only, by which the researcher would establish the magnitude of relationships (covariances or other measure of the relationship) between the auxiliary variables and the likelihood of participation. The researcher would identify sets of survey variables or estimates that share high correlations with different auxiliary variables *on the respondent cases*. Based on that indirect information, separate variance functions above would be presented for each class of estimates. In some sense, this would resemble the technique of identifying sets of estimates subject to similar design effects (e.g., see the CPS Technical Paper 63RV, <http://www.census.gov/prod/2002pubs/tp63rv.pdf>).

## 5. Summary

The above is a listing of alternative indicators that might be useful to explore across several ongoing surveys simultaneously, in order to aid judgments about whether different indicators do indeed supply more information about possible nonresponse errors in survey estimates.

## 6. References

Curtin, R., S. Presser and E. Singer (2000). "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* **64**(4): 413-428.

Groves, R. M. (2006). "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* **70** (5): 646-675.

Keeter, S., C. Miller, A. Kohut, R. M. Groves and S. Presser (2000). "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* **64**: 125-148.

### List of Authors With Affiliations

Robert M. Groves, University of Michigan and Joint Program in Survey Methodology

J. Michael Brick, Westat

Mick Couper, University of Michigan and Joint Program in Survey Methodology

William Kalsbeek, University of North Carolina

Brian Harris-Kojetin, US Office of Management and Budget

Frauke Kreuter, Joint Program in Survey Methodology

Beth-Ellen Pennell, University of Michigan

Trivellore Raghunathan, University of Michigan and Joint Program in Survey Methodology

Barry Schouten, Statistics Netherlands

Tom Smith, National Opinion Research Center

Roger Tourangeau, University of Michigan and Joint Program in Survey Methodology

Ashley Bowers, Matthew Jans, Courtney Kennedy, Rachel Levenstein, Kristen Olson, Emilia Peytcheva, Sonja Ziniel, and James Wagner, University of Michigan

Comments: 0

 Comments

## Ask the Experts - Experiences With Cell Phone Samples

---

Thursday, October 30, 2008, 3:34:03 PM | Editor

Survey Practice sent a query to the Association of Academic Survey Research Organization's listserv to ask the members to describe their experiences with cell phone samples. Four organizations provided responses. Feel free to ask for more information in the comments section at the end.

**Gene M. Lutz**

**Director**

**Center for Social and Behavioral Research**

**University of Northern Iowa**

Our center conducts data collection for the Iowa Behavioral Risk Factor Surveillance System survey. For the period of February through September of 2008, both a traditional landline version and a cell phone pilot version of the survey were in the field, using samples of telephone numbers provided by CDC/Genesys. During this period there were 3766 landline completions and 245 cell completions. The landline questionnaire contained 156 items and the cell questionnaire contained 101 items. Cell respondents were offered a \$10 compensation; no compensation was offered for landline respondents. Some of the first findings are the following:

\* More dials were required to achieve completions by cell. The mean number of dials per complete were 87 for cell and 24 for landline; a ratio of 3.6 cell dials to every 1 landline dial to achieve a completion.

\* All AAPOR cooperation rates slightly favored cell over landline. E.g., Estimated AAPOR Cooperation Rate 2 was 0.825 for cell and 0.747 for landline.

\* Response rates were similar for cell and landline. Estimated AAPOR Response Rate 3 was 0.445 for cell and 0.443 for landline.

\* Comparably fewer refusals than completes occurred by cell for eligible, selected respondents. The ratio of completes to refusals was 4.6:1 for cell and 3.8:1 for landline. Estimated AAPOR Refusal Rate 2 for cell was 0.091 and 0.140 for landline.

\* Nine of ten cell respondents accepted the compensation offer. 75% accepted an electronic gift card, 14% accepted a physical store gift card, and 11% declined any compensation.

\* However, eight of ten cell respondents claimed they would have completed the interview without the compensation offer. 79% said "yes" and 21% said "no" to this post interview question.

\* Most cell respondents (16% strongly agree and 69% agree) said they had a positive experience using a cell phone for the interview.

\* Seven of ten cell respondents were multi-tasking while doing the interview. Only one-fourth (26%) were doing nothing else, but 22% were watching TV, 14% were watching children, 11% were using the computer, 11% were doing household chores, and lesser percentages were doing other things.

\* Most cell respondents (76%) said being able to do other things while completing the interview by cell made them more likely to do the survey.

\* Overall, the cell survey required more resources (sample pieces and dial attempts) for locating eligible respondents than the landline, but the actual cell respondents were generally positive about using a cell phone to be interviewed.

**Thomas M. Guterbock**  
**Director**  
**Center for Survey Research**  
**University of Virginia**

\* If you want to start doing cell phone interviewing, plan to involve your interviewers, supervisors, and technical staff in the process from the start. Expect to be revising your training materials, adding and defining new disposition (call outcome) codes, checking your partially collected data in new ways, and helping the entire staff adjust to something that is quite novel in some ways. Treat this transition as a learning moment for your entire organization.

\* We ran a small experiment here using cash incentives for those who agree to be interviewed via cell phone. One group was offered \$5, another group was offered \$10 for the interview. The rate of completions per hour was better with the \$10 incentive, so much better, in fact, that it more than pays for the \$5 difference in the incentive cost.

\* The people we reach by cell phone, even those who have a landline, are very different demographically from those we reach by landline. This suggests that it is probably not wise to 'screen out' dual-phone households from the cell phone sample. Better to interview all cell phone users you can reach, whether or not they have a landline phone in the household.

\* Costs for interviewing cell phone numbers are certainly higher. The time it takes per completion is a little more than double that needed for a conventional landline interview, even when the cell phone respondents are offered a cash incentive. Add in the incentive cost and the cost of a cell phone completion is about 2-1/2 times more.

**Molly Longstreth**  
**Director**  
**Survey Research Center**  
**University of Arkansas**

The following were found on 1,587 completed surveys of the general population of a single state conducted during October 2008.

\* Cell-phone interviews are slightly shorter, on average, than are landline interviews.

\* The number of calls to complete cell versus landline interviews are the same.

\* Hours per completion for cell are about one-third longer than for landline interviews.

**Bob Craddock**  
**University of Wisconsin Survey Center**

In Wisconsin, the state Department of Health Services conducts the Behavioral Risk Factor Surveillance System (BRFSS) survey. In 2008, a pilot project attempted to extend the BRFSS survey to cell-phone-only adults who would never be reached under the usual landline protocol. The BRFSS cell pilot (BCP) differed from the BRFSS landline (BLL) protocol in several ways: sampling, eligibility, compensation, content, and calling strategy (both overall and for refusals).

The BCP sample was pre-screened with the intention to include only cell numbers; the BLL sample was pre-screened to exclude them. BCP respondents were asked whether they also had a residential landline, and those that did were screened out. BCP included respondents living outside Wisconsin, as long as they had an in-state phone number and lived in one of several other states also involved in the pilot study. BCP treated any adult answering the cell phone as the selected respondent; BLL enumerated the household and selected one adult at random. BCP respondents were offered ten dollars as compensation for their airtime cost; BLL respondents were not compensated.

BCP used a shorter questionnaire averaging 12 minutes long; BLL interviews averaged 19 minutes and included more potentially sensitive questions about sexual behavior and orientation, binge drinking, and children. Calling protocols differed. BCP required 5 or 8 calls (depending on the month) before closing out unresolved cases, and no case received more than 15 calls; BLL used a 15-call minimum and a 25-call ceiling. BCP employed more experienced interviewers who had previously worked on BLL for several months.

Refusal conversion protocol differed significantly. BLL called back to convert all first refusals (except abusive ones). BCP did not call back refusals if the respondent referred to their cell phone in the process of refusing, although interviewers did try to convert those respondents during the initial refusal. Both studies ceased calling at the second refusal.

BLL and BCP were conducted by the University of Wisconsin Survey Center during the months of February through August 2008. Although the protocol variations make it difficult to draw rigorous conclusions about cell phones vs. landline effects, a few intriguing differences emerged.

\* The yield of completed interviews per cases called was much lower for BCP (4%) than for BLL (30%). (Table 1)

\* Protocol differences for refusals and number of calls meant BCP refusal conversion attempts (96%) were less likely than BLL conversion attempts (99%). (Table 2)

\* In those cases where refusal conversion was attempted, BLL was more successful at obtaining a completed interview (14%) than was BCP (1%). (Table 3)

\* However, the percentage of respondents cooperating with the conversion attempt long enough to either complete the interview or establish ineligibility was very similar (BLL 14% complete, 1% ineligible; BCP 1% complete, 12% ineligible).

\* Conversion attempts on BCP helped improve the response rate mainly by identifying those ineligible cases, rather than by obtaining completed interviews.

Table 1. Completed Interviews per Cases Called

Outcome	Landline (BLL)		Cell (BCP)	
	N	%	N	%
Complete	4111	(30%)	191	(4%)
Other	9649	(70%)	4879	(96%)
Total Sample	13760		5070	

Table 2. Conversion Attempts after First Refusals

Attempted	Landline (BLL)		Cell (BCP)	
	N	%	N	%
Yes	4309	(99%)	483	(96%)
No	38	(1%)	22	(4%)
Total Refusals	4347		505	

Table 3. Outcome of Refusal Conversion Attempts

Outcome	Landline (BLL)		Cell (BCP)	
	N	%	N	%
Complete	585	(14%)	5	(1%)
Ineligible	60	(1%)	56	(12%)
Unresolved	975	(23%)	184	(38%)

Second Refusal	2689	(62%)	238	(49%)
Total Attempts	4309		483	

Comments: 0

 [Comments](#)