

Welcome to Survey Practice

Saturday, August 23, 2008, 2:56:27 PM | Editor

Survey Practice began as an idea of Bob Groves, Sandy Berry, and a few others who thought that The American Association for Public Opinion Research (AAPOR) needed a publication that could provide good, sound information to new survey researchers, be relatively flexible, be able to address current issues, and be easy to read. The AAPOR executive council agreed and approved the concept.

Then, a larger group that included the editors (John Kennedy, David Moore, Diane O'Rourke, and Andy Peytchev) and many members of the editorial board developed the scope and mission of Survey Practice. The mission of Survey Practice:

Survey Practice provides current information on issues in survey research and public opinion. The articles in Survey Practice emphasize useful and practical information designed to enhance survey quality by providing a forum to share advances in practical survey methods, current information on conditions affecting survey research, and interesting features about surveys and people who work in survey research. Survey Practice is intended for practicing survey and public opinion researchers.

Survey Practice is not an academic journal. Its articles will not have extensive literature reviews, elaborated hypotheses, or difficult statistics. Survey Practice is not a competitor of Public Opinion Quarterly.

Survey Practice will have both short research articles and features. The research articles will focus on issues of interest and utility to public opinion and survey researchers. These articles are listed as "public opinion" and "methods" in the Categories field. The features will include interviews with interesting survey researchers, an "Ask the Experts" column, and a FieldNotes column that has short blurbs about surveys and those who conduct them. We plan to have two or three research articles and one or two features in each monthly issue

As an Internet publication, Survey Practice will be interactive. We expect that readers will comment on the articles and pose questions to the authors. We hope that through these exchanges, the collective wisdom of AAPOR members can be used to improve our survey methods and our understanding of important public opinion issues.

The editors have an applied orientation to survey research and public opinion. John Kennedy has directed a survey center for more than 20 years and worked at the U.S. Census Bureau for four years. David Moore has a distinguished career in public opinion research including working in The Gallup Organization. Diane O'Rourke worked at the University of Illinois Survey Research Laboratory for over 30 years, where she was Assistant Director of Survey Operations at the time of her retirement. Andy Peytchev is a recent University of Michigan grad and is currently a survey methodologist at RTI International. The members of the editorial board are practicing survey researchers who are involved in applied, practical survey research and public opinion every day.

To serve practitioners well, we need your help. You can do this by commenting on the articles which could include how you handled similar issues. Please tell us what kinds of articles you want in Survey Practice. We plan to have two or three research articles and one or two features in each monthly issue but we can publish more to meet your needs.

We encourage you to send research articles and research notes to Survey Practice. While we maintain quality standards, we are also very author-friendly. We also encourage you to send blurbs for FieldNotes, suggestions for interviews, questions for Ask the Experts, and any other ideas you have for articles or features that would make

Survey Practice more useful to you.

Survey Practice is intended to help our profession and everyone is welcome to contribute.

Comments: 1

 Comments

What's in Survey Practice This Month

Saturday, August 23, 2008, 2:52:25 PM | Editor

Survey and public opinion research is both messy and robust. We often find messy situations such as election polling when conditions change rapidly. We also find that survey methods are so robust that decreasing response rates have not necessarily produced less useful data.

In this first issue, we can see examples of both the messiness and the robustness of our practices. The [Bishop and Mockabee](#) article provides evidence that a standard public opinion question may not measure what we assumed it measured. The [Tom Smith article](#) shows that despite careful research, the transition from a four-point to a five-point scale was not as neat as hoped. In his [interview](#), Ken Prewitt describes some messy situations that survey research may find itself. However the [Keeter et al article](#) shows how robust survey methods can help us to understand better an increasingly important group in the U.S. - Muslims.

[FieldNotes](#) has short notes that we hope are interesting. For the first month, we show some topics that could be included. Please send similar notes to Survey Practice for next month. FieldNotes also has a "Question of the Month" that readers could discuss.

This month, we list some questions that might be asked in the [Ask the Experts](#) column. Your questions are welcomed.

Let us know what you think about the articles and the issues they raise.

John Kennedy
Diane O'Rourke
David Moore
Andy Peytchev

Comments: 0

 Comments

An Interview with Kenneth Prewitt

Saturday, August 23, 2008, 2:48:04 PM | Editor

Conducted by Andy Peytchev,
Survey Practice Associate Editor

Kenneth Prewitt is a Professor of Public Affairs at Columbia University. He has formerly held positions as director of

the U.S. Census Bureau, director of the National Opinion Research Center, president of the Social Science Research Council, and senior vice president of the Rockefeller Foundation. We asked him about his career and the major issues the survey field is facing. In particular, Dr. Prewitt has raised concerns about privacy and confidentiality, and declining cooperation in surveys, at a time with increasing amounts of information from multiple sources.

SP: What led you to a career in surveys? Did you have plans to do something different?

Prewitt: I was teaching political science and chairing the department at the University of Chicago. I was involved with NORC, though in a quite peripheral way. The Provost called and said that the University wanted to appoint me as Director of NORC. To which I replied, "I have no experience in running anything so big and complicated." In the best U. of Chicago tradition, the Provost replied: "Well that's the idea. This is a faculty-run university; we don't want professional managers making academic decisions." I became NORC director, though certainly not with the intention of initiating a "career in surveys," and in fact don't see my career that way. As it turned out, however, NORC was relevant to my appointment as Director of the Census Bureau. The Clinton administration was looking for someone who at least on paper had the right credential - and my earlier role at NORC provided that credential. It allowed the Democratic administration to present me to a Republican Congress (the position requires Senate confirmation) as a nonpartisan academic with relevant experience in scientific management. My career, however it might be characterized, is more an accident than a plan. I see myself as an academic who happened to do some other things in foundations, scientific organizations, and the government.

SP: What do you think are the most pressing problems facing surveys in the near future?

Prewitt: Public cooperation is a very serious problem and we'll talk in more detail about that. But there is a larger issue of which public cooperation is just one element. I believe that over the next quarter-century or so, the government will increasingly merge administrative data and survey data. What we today understand as the "national statistical system" will more properly be thought of the "national information system." Sample-based survey data will be part of that system, but less dominant than it has been in the previous half-century or more. For example, the new SIPP [Survey of Income and Program Participation] that is under consideration might be based on 50% administrative records and 50% survey data. If so, that is an indicator of where the whole system is going.

One reason for the turn to administrative data - and other data sources, such as commercially provided scanning - is the survey response rate problem: the unit cost of each respondent to a survey is high and getting higher. If, as some have suggested, we can control the response rate issue with incentive payments, there will be further cost increases - as well as data quality problems. Nonresponse, by the way, is not just how many respondents answer but also item nonresponse. There has been less attention to item nonresponse, but in the 2000 Census there was a sharp increase of item nonresponse, reaching into the 20% range on several questions.

SP: In your article in *Science* you point out that the public has serious privacy and confidentiality anxieties and that voluntary cooperation with surveys is declining. Why do you think that the two are linked, as opposed to other factors affecting responding to surveys?

Prewitt: There are two pieces of relevant empirical work. After the 1990 census and again the 2000 census, Eleanor [Singer] reported a correlation between levels of privacy concerns and responses to the census. ...Then, during the 2000 census, another study in which I was involved used Knowledge Networks (an Internet survey firm) in a design that took real-time surveys, six in all, as the census was taking place. When the privacy outcry erupted as the long-form reached households, we even added an experimental design on its impact on census cooperation. The results are presented in *The Hard Count*, Hillygus, et al. (Russell Sage Foundation). Here we estimated that the privacy uproar over the long form depressed the mailback response rate by as much as 5 percent. This work, along with that of Eleanor Singer and her colleagues, offers strong evidence of the association between census cooperation and

privacy/confidentiality issues.

You are correct to suggest that privacy concerns are not the only thing affecting fall-off in survey cooperation. Junk mail and push/pull marketing research turn the public off. Half the population, no doubt including you and me, have refused to cooperate with phone surveys. I'm sure you do what I do and try to find out how serious it is. If it's serious, you cooperate, but most people aren't going to find out if it is serious. So if half the population, and that's the half that responded to the survey asking whether they had refused to cooperate, are saying they've already turned down surveys, it is highly likely that more is going on than privacy concerns. It may have to do with disgust over the whole marketing agenda that disrupts our dinner hour. In terms of the larger picture there is a serious problem with response rate, some portion of which is attributable to privacy/confidentiality concerns, but how much of the variance we attribute to that factor is uncertain at present.

As I said, however, there is a larger, more complicated challenge to survey data. It will occupy a steadily decreasing role in the nation's information system. Already a number of European countries, especially the Nordic countries, will tell you that less than 25% of the information used by the government comes from surveys. The administrative data are already collected by the government for program management purposes; why not use it in lieu of survey data to understand the economy and society? Even if we were not facing a response rate problem, the sheer density of administrative and surveillance data presents a challenge to our traditional reliance on survey data as the platform for the national statistical system. By the way, by "surveillance," I do not have in mind the Patriot Act so much as the data we provide every time we use a credit card or book a flight. This is the digital footprint each of us leaves. The sheer amount of digitized data is enormous and we are at the early stages of its expansion and of the data mining methodologies used to extract information from it. We cannot be surprised if the government (now, for example, facing a full cycle 2010 decennial census, which includes the American Community Survey, that will exceed \$12b) asks "cannot it be much cheaper to see what we can learn from all of this administrative and digital data than to try to find people and convince them to answer what they see as our intrusive survey questions?"

As I wrote in the *Science* essay (or have written somewhere), it has taken nearly a century to get survey data to the level of quality we now expect - measuring sampling and non-sampling error, using cognitive science to improve question wording, etc. The amount of serious scholarship on the error structure of administrative data is miniscule in comparison, and even less on the error structure of scanned data or other surveillance sources.

Certainly one of the big challenges looming in front of us is the quality of administrative data. From the perspective of quality, administrative data have a troubling characteristic. Collected to administer a program, what matters is the accuracy of the variables that are germane to the particular program. The Social Security Administration really wants to get my age right, but does not need to be precise about my earnings above the threshold level that determines how much they collect and then will have to distribute. The IRS wants to get very exact data on my income, but can be more casual about my age. Neither of these programs has to be overly concerned about my current residential address. School records, in contrast, care less about my income and age, but if I am sending my children to a local school will want to know in which district I live. My address matters.

Administrative records are case-rich but variable-poor, that is, a large number of observations but only a small set of information about each observation. To be useful for population analysis, then, they have to be linked. This invites matching errors, and we know those to be serious. Nevertheless, a heavier use of administrative data is part of our future. All, or nearly all, of the questions on the American Community Survey could be addressed with federal and state administrative data or from private-sector data. There is an item on home mortgages - why not call banks for that information? Of course the blurring of the boundary between survey and administrative data, to say nothing of blurring the boundary between public and private-sector data, raises a host of issues other than data quality - issues of coverage and representativeness and, of course, of privacy and confidentiality. We are at the earliest stages of assembling scientific talent to take up these questions.

SP: You have been concerned with privacy and confidentiality as increasing threats to the US Census and the American Community Survey; why do you think these concerns are increasing?

Prewitt: For reasons unrelated to the Census and the ACS, there is a public reaction to the intrusiveness of the survey industry more generally. This intrusiveness irritates the public irrespective of data confidentiality concerns, but the Census can only rely on promising confidentiality. The person saying, "just leave me alone" is not going to be persuaded when the Census Bureau says, "your answers are confidential." The irritation and the response don't match.

This is not to ignore the problem of confidentiality, and here the Census is vulnerable to the more generalized anxiety over matters such as identity theft. Nearly half the public already discounts the pledges of confidentiality by the government. When asked, "Do you think your census data are being kept confidential?" about 40-45 percent of the population says no, and I fear that the percentage will increase because of news coverage of missing laptops from the VA, etc. The sheer volume of data collected in so many different places and via so many different methodologies guarantees an increase in incidents of leakages. There is a huge information market, largely driven by commercial interests. It is gathering up everything possible in order to sell products or, in this season, to win elections. There will be inadvertent as well as deliberate misuse of data. As the public experiences this, it will discount the privacy/confidentiality problem on grounds that there is not much that can be done anyway - short of throwing away credit cards, staying off the Internet, not visiting a doctor or catching a plane. But a public irritated by intrusiveness and knowing that there is risk that private information will not always be carefully handled can take it out on the census and other government surveys. It is easier to say "no" to a census-taker than to quit shopping on the Internet.

SP: What do you think we could be doing to protect the mission of a survey organization from the public withdrawal from cooperation?

Prewitt: Survey data, whether collected by the government or by reputable private organizations, are, we know, a public good. We can do a better job at packaging and presenting this public good data in a manner that is of value to the general public rather than just to government and commercial decision-making. For example, the new effort to create a national indicator system, under the leadership of a new non-profit, The State of the USA, is such an attempt. Those of us involved in that effort intend to design key national indicators that will be used by schools, churches, community organizations, local governments, and dozens of similar settings. The original data will largely come from the federal statistical information system, but will be returned to the public for its purposes and goals. This is providing a better tool than what is currently available, so that citizens can take advantage of the information that, after all, only exists because they answered government survey questions in the first place.

Put more bluntly, I don't think we can protect survey data by simply saying we're going to keep these data confidential. We've said that and said that, but even if census data are well-protected (and they are), it is not enough. We have a different educational project before us - to remind the public about the source of news coverage of the housing market or immigration issues or school reform. Hardly a day goes by in which the *New York Times*, for example, fails to have a story that does not cite the American Community Survey. We have to convince people that we can only tell you about your own community if you cooperate with surveys such as the American Community Survey.

I want to connect census cooperation less to a pledge of confidentiality and more to data accessibility and usability. This, I believe, is the basis on which to protect government surveys.

SP: Selzer and Anderson presented data recently that indicates the Census Bureau released identifying information during WWII. Do you think it's possible that any statistical agencies are currently releasing identifiable information as

part of the war on terror?

Prewitt: I think not, though from the outside we cannot offer a definitive answer. If a statistical agency were releasing identifiable information to, for example, Homeland Security, the agency would have to deny it. But a statistical agency has much less useful data on the specifics of flight school enrollment or who is learning to drive large trucks or purchasing chemicals than many alternative sources. The Anderson-Selzer paper was a great piece of detective work, and an important corrective to the historical record. After 50 years of denying that there had been any release of micro-data in connection with the treatment of Japanese-Americans, we stand corrected. This makes the statistical agencies even more determined to prevent this misuse.

I worry less about statistical agencies and more about data mining of administrative data, an issue that gets more troubling as the boundary blurs between administrative data and survey data - which returns us to where we started. There is a new "information system" in the making. My plea is that we subject it to the same array of quality standards, principles of confidentiality, and accessibility practices that we have worked so hard to ensure for the survey-based statistical system.

References

Hillygus, D.S., Nie, N.H., Prewitt, K., and Pals, H. (2006). *The Hard Count: The Political and Social Challenges of Census Mobilization*. New York: Russell-Sage Foundation.

Prewitt, K. (2004). What if We Give a Census and No One Comes? *Science*, 304, 5676, 1452-1453.

Comments: 0

 Comments

Questionnaire and Fieldwork Challenges in a Probability Sample Survey of Muslim Americans

Saturday, August 23, 2008, 2:10:35 PM | Editor

Scott Keeter, Gregory Smith, Pew Research Center
Courtney Kennedy, University of Michigan
Chintan Turakhia, Mark Schulman, Abt SRBI
J. Michael Brick, Westat

Short Overview of Research Methods

In 2007, the Pew Research Center conducted what is believed to be the first-ever national telephone survey of a probability sample of Muslim Americans, a rare, dispersed, and highly diverse population. The study examined the political and social values, religious beliefs and practices, and life experiences of Muslims living in the U.S. today. Conducting the study presented several difficult challenges.

First, since Muslims constitute a relatively small portion of the overall adult population in the U.S. - approximately 0.5% of typical English-only survey samples - simply contacting and identifying a large sample of Muslims required a considerable investment. We oversampled from counties with a high prevalence of immigrants from predominantly Muslim countries and supplemented our sampling frame by re-contacting and interviewing self-identified Muslims

from previous Pew surveys. In addition, we sampled from a commercial database of individuals with Muslim names. Further details about the sampling design and methods can be found in the full report on the survey, located on the [Pew Research Center website](#).

Second, since the Muslim population in the U.S. includes many immigrants, a sizeable portion of the target population is unable to complete a telephone survey in English. Extrapolating from U.S. Census data and our survey, we estimate that between 9% and 22% of adult Muslims in the U.S. speak one of three languages: Arabic, Urdu or Farsi. Accordingly, we employed interviewers who are fluent in these three languages, with the result that approximately 12% of all interviews (and 17% of all weighted cases) were ultimately conducted in these languages. One consequence of this is that Pew's estimate of the number of Muslims living in the U.S. was somewhat larger than that derived from other survey-based methods, which have typically depended on English-only surveys.

A third challenge was the sensitivity of the population to being interviewed. We designed the study and tailored the questionnaire to meet this concern by beginning the interview with neutral questions about satisfaction with the community, personal happiness, and other innocuous topics. After considerable experimentation, we decided to reveal the purpose of the study to respondents early in the interview, immediately following the opening sequence of neutral questions. The logic for revealing the focus of the study - a practice not common in survey research - was that Muslim respondents would quickly realize that the study was focused on Muslims and Islam, and that there would be a greater chance of establishing a bond of trust by revealing the intent of the survey in its early stages. Indeed, pretesting of the questionnaire indicated that to allay fears about the purpose of the study, it was essential to describe the purpose of the study and have available clear explanations of the sampling method for the interviewers to use in response to questions. Cumulatively, these efforts aided in gaining respondents' trust and maximizing cooperation rates.

In addition to the major challenges listed above, we also confronted many practical fieldwork challenges. These included the massive volume of calling necessary to locate eligible respondents in a portion of the RDD sample and potential cultural issues related to interviewer/respondent gender matching. The calling volume had serious implications for interviewer morale and readiness, since the vast majority of interviewer effort was devoted to unsuccessful screening interviews. We made an effort to give all interviewers a mix of RDD sample and targeted sample, which meant that they would more frequently have the opportunity to complete an interview with an eligible respondent. Similarly, we instituted a protocol of having interviewers first ask for respondents of their own gender, which we believe yielded a better cooperation rate.

Comments: 0

 [Comments](#)

Understanding the Meaning of the “Mood of the Country”

Saturday, August 23, 2008, 2:02:27 PM | Editor

George F. Bishop and Stephen T. Mockabee
Department of Political Science
University of Cincinnati

Assessing the “mood of the country” has become a staple in the diet of American public opinion polling (Hugick and DiAngelo, 2006; Ladd, 1992). Major polling organizations routinely ask questions about whether Americans are “satisfied or dissatisfied with the way things are going in the United States at this time” or whether “things in this

country are heading in the right direction or...off on the wrong track." However, as Hugick and DiAngelo (2006) rightly remind us, "Interpreting the results of these direction questions...is a challenge for professional pollsters, let alone more casual consumers of national media polls."

What do changes in these mood indicators actually mean? Do the readings of the "mood" mean the same thing at one time as they do at another? If not, in what sense are they leading or misleading indicators of the "the mood of the country"? These are among the questions we address in this investigation.

Comparability of Question Wording, Order, and Context

Hugick and DiAngelo's (2006) analysis addresses two key challenges to interpreting the results of the "mood of the country" indicator: comparability of question wording and comparability of question order. Hugick and DiAngelo (2006) have demonstrated that the variously worded versions of the mood item produce similar results, once the data are adjusted for differences among survey houses in the percentage of volunteered "don't know" responses. As to question order and context, Hugick and DiAngelo's (2006) analysis of six question order experiments with presidential approval and national mood items in Newsweek polls did find some evidence for order effects, particularly in two polls conducted just prior to the 2004 election. But when the data for all six polls were combined, none of the order effects achieved statistical significance. In sum, the jury on order effects is still out.

Temporal Comparability of Meaning

The problems of interpretability posed by differences in the wording and context of "mood" indicators pale by comparison with a more fundamental problem: variability over time in the meaning-and-interpretation of identically worded and sequenced questions. A cardinal assumption researchers make in asking any survey question is that it should mean the same thing to all respondents (Bishop, 2005; Brady 1985; Foddy 1993). Fowler (1995, p.84) has expressed essentially the *same psychological stimulus principle* in his recommendations for improving the wording of survey questions: "...A survey question should be worded so that every respondent is answering the same question." Furthermore, as Groves (1989, p. 450) has noted, "Although the language of the survey questions can be standardized, there is no guarantee that the meaning assigned to the questions is constant over respondents." This becomes critical because "A fundamental tenet of scientific measurement is that the measuring device is standardized over different objects being measured" (Groves 1989, p.449).

A corollary to the *same psychological stimulus principle* is that the question should also mean the same thing to respondents at time two that it did at time one: the *constant psychological stimulus principle* (Bishop, 2005). As Nie and his colleagues (1979, p.11) have captured the issue, "Even if the same question is asked at two different points in time, is it really the same question? The fact that times change may mean that the meaning of the question undergoes change." And if so, this violates the fundamental *invariance of meaning assumption* we make in trend analysis.

The objective of our investigation is to demonstrate this problem of comparability in meaning, using the "mood of the country" as a prime example. Consider the vagueness and ambiguity of the standard Gallup question: "*In general, are you satisfied or dissatisfied with the way things are going in the United States at this time?*" Just on its face, this question almost certainly means different things to different respondents. For some respondents in the present period, it may mean how things are going with regard to the war in Iraq. For many other people, however, it may now mean various domestic issues such as problems in the housing market or high gasoline prices.

Method and Data Sources

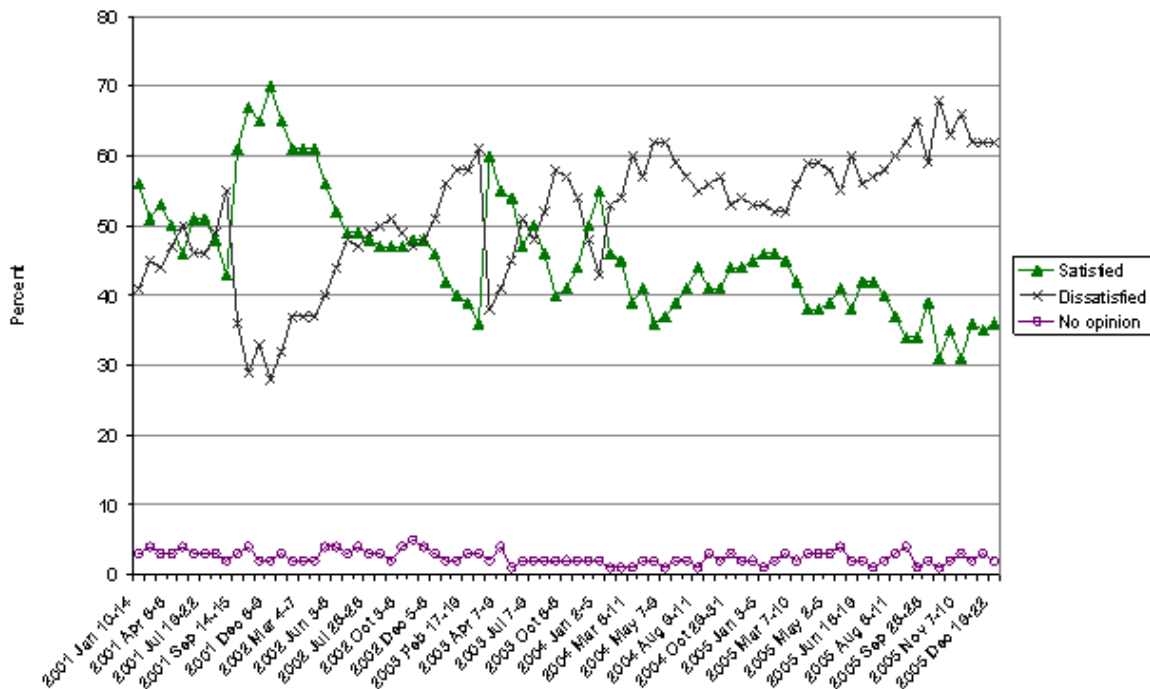
We use data from numerous Gallup surveys conducted from February 2001 to December 2005(1) to estimate

multivariate models of the direction-of-the-country item, with presidential approval, congressional approval, ratings of the current state of the economy, future economic expectations, and party identification as independent variables. Because the dependent variable is dichotomous, binary logistic regression is used. For each poll, then, we generate a measure of the relative potency of the independent variables as predictors of the “mood of the country.” If our *incomparability-of-meaning hypothesis* is correct, we should observe considerable variation over time in the relative potency of these predictor variables—even when the overall rating of the “mood of the country” appears to be relatively stable in the aggregate.

Findings

Figure 1 shows the trend in the “mood of the country” item. In 2001, satisfaction ranges from 56 percent to as low as 43 percent just prior to the events of September 11, 2001. Satisfaction then soared as part of a so-called “rally event” reaching 70 percent in December of 2001. After that five-year high point, the mood of satisfaction slowly declined to a low of 36 percent prior to the invasion of Iraq, then took a leap to 60 percent shortly after the war began in the middle of March 2003, closely tracking the increase in the President’s approval. But by early 2004, a majority of Americans reported that they were dissatisfied, and by the end of 2005 only 31-36 percent said they were satisfied, tracking rather closely the President’s dismal approval ratings.

Figure 1. Gallup “Mood of the Country” Satisfaction Ratings, 2001-2005

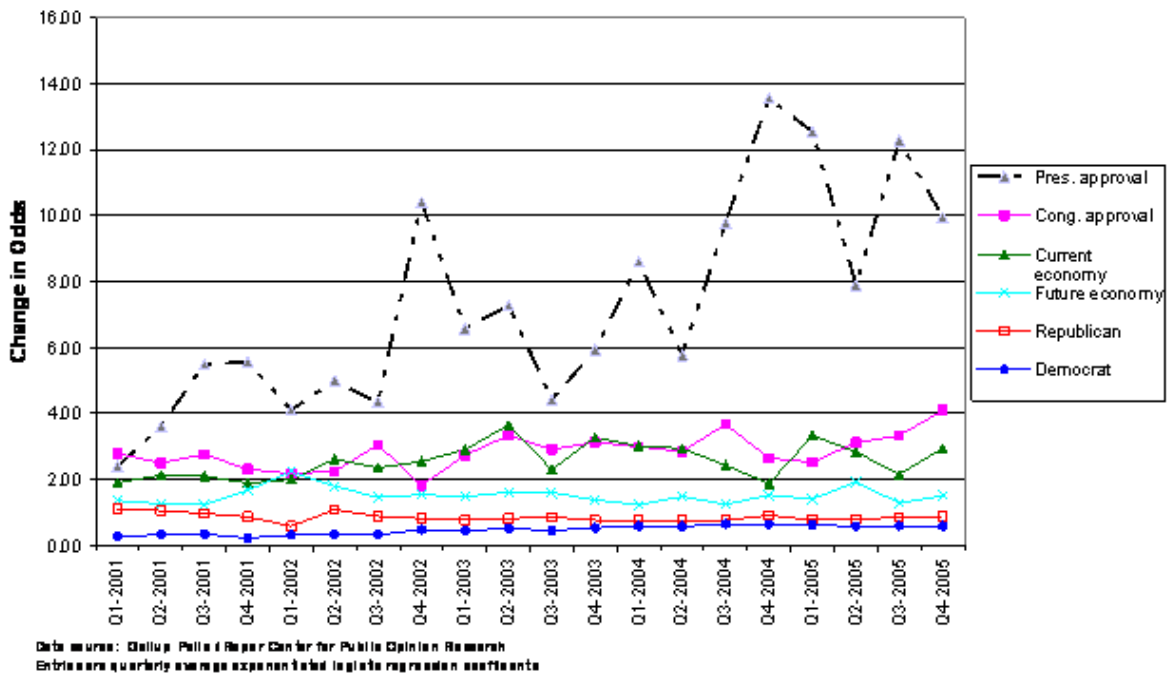


Mood of the Country

The data patterns in Figure 2 give us some further insight into the meaning of the trend in the “mood of the country.” The figure is based on the logistic regression analysis described above, showing changes in the relative potency of the predictor variables, controlling for party identification. The entries forming the trend line are exponentiated logistic regression coefficients, which indicate the predicted change in the odds of being satisfied with “the way things are going in the United States at this time”—given a one unit change in the independent variable. For

easier viewing, the figure presents quarterly averages of the exponentiated coefficients.

Figure 2. Predictors of Satisfaction Ratings: 2001-2005



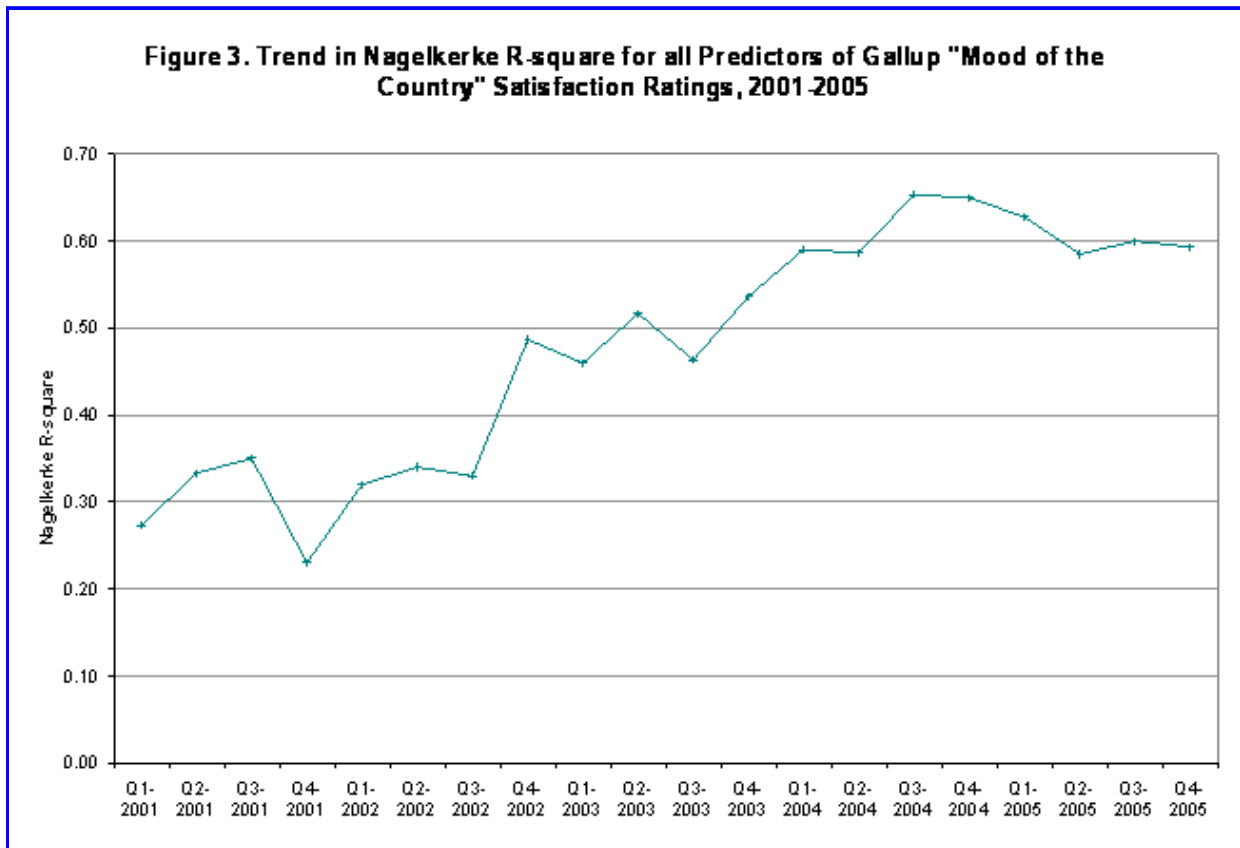
Predictions of Satisfaction Ratings

First, notice that the relative predictive power of presidential approval grows considerably over time. Notice too some of the key points at which presidential approval peaks as a predictor of the national mood in Figure 2: following September 11, 2001; just before the congressional election of 2002; shortly after the beginning of the Iraq war; and shortly before the election of 2004—all times at which the leadership of President Bush was highly salient to the public. And so it seems quite plausible that respondents would use the heightened visibility of Bush’s presidential persona in the mass media to disambiguate and interpret the very general question about “the way things are going.”

The trend data in Figure 2 also indicate that approval of “the way Congress is handling its job” becomes somewhat more important as a predictor of the mood of the nation beginning in 2003, and then reaches a plateau. So too do respondents’ assessments of the current condition of the economy and their expectations about the future of the economy loom larger over time in predicting overall satisfaction. While important, all of the changes in these other predictors are much less significant by comparison with the growing power of presidential approval as a key to understanding the meaning of changes in the national mood.(2)

But perhaps the most revealing pattern appears in Figure 3, which shows the pseudo R-Square for the model over time.(3) The same set of predictors explains more and more of the variance in the “mood of the country” with the passage of time. What starts out at the beginning of the Bush administration as a vague and ambiguous indicator of the “mood of the country” that is weakly or moderately linked to how the President, the Congress, and the economy are doing becomes steadily less fuzzy in the minds of respondents. By 2004-2005 the question’s meaning is much

more about how George W. Bush is dealing with *"the way things are going in the United States"* though also more too about how Congress is handling *things*, and how *things* appear to be going in the economy. Changing events had altered its meaning-and-interpretation over time.



Summary and Conclusions

Because its meaning-and-interpretation is continuously shifting over time, comparisons of the "mood of the country" indicator over time become essentially invalid. One is not comparing an apple with an apple. To that extent, this standard question about the "mood of the country" is an ambiguous and potentially misleading indicator of American public opinion.

The notion that an identically worded question undergoes changes in meaning over time may seem foreign to some readers, but imagine we were talking here about the changes in "mood" resulting in part from differences in question wording or in question order and context; in that case, almost no one would disagree that we have a serious artifact problem on our hands. In practice, changes in the meaning of an identically worded and sequenced question are functionally equivalent to changes in wording or context. Temporal incomparability is temporal incomparability.

However, there is an upside to our findings as well, because we now have the beginnings of a new and simple, quantitative metric to systematically monitor how the meaning of the "mood of the country" question changes over time in tandem with real-world events and other indicators of the national mood.

References

Bishop, George F. 2005. *The Illusion of Public Opinion: Fact and Artifact in American Public Opinion Polls*. Lanham, Md., Rowman & Littlefield.

Bishop, George F., Stephen T. Mockabee, and Eric W. Rademacher. 2006. "Presidential Approval Ratings: Meaning, Measurement and Validity". Paper presented at the annual meeting of the American Political Science Association, Philadelphia, PA, August 31-September 3, 2006.

Brady, Henry E. 1985. "The Perils of Survey Research: Inter-Personally, Incomparable Responses." Political Methodology 11 (3-4): 269-90.

Foddy, William. 1993. Constructing Questions for Interviews and Questionnaires. Cambridge: Cambridge University Press.

Fowler, Floyd Jackson Jr. 1995. Improving Survey Questions. Sage Publications.

Groves, Robert M. 1989. Survey Errors and Survey Costs. New York: Wiley.

Hugick, Larry, and Stacy DiAngelo. 2006 (June). "Related Measures: Direction of the Country and Job Approval." www.publicopinionpros.com

Krosnick, Jon., and Laura Brannon. 1993. "The Impact of the Gulf War on the Ingredients of Presidential Evaluations: Multidimensional Effects of Political Involvement." American Political Science Review 87: 963-75.

Ladd, Everett C. 1992. "Big Swings in the National Mood Are a Staple of Contemporary Politics." Public Perspective, Vol.3, No.2 (January/February), pp.3-5.

Nie, Norman H., Sidney Verba, and John R. Petrocik. 1979. The Changing American Voter. Cambridge: Harvard University Press.

Pampel, Fred C. 2000. Logistic Regression. Thousand Oaks, CA. Sage Publications.

Condensed and revised version of a paper presented at the 62nd Annual Conference of the American Association for Public Opinion Research, Anaheim, California, May 17-20, 2007.

1. Data Source: Roper Center for Public Opinion Research, University of Connecticut.

2. We know from other trend analyses (Bishop et al., 2006) that the events of September 11th and its aftermath, followed by the war in Iraq, were the dominant factors that drove respondents' interpretations of the Gallup question about how the president was generally handling his job during the same time period.

3. Pseudo R-square indicates the "...reduction in the model 'error' due to the independent variables..." (Pampel, 2000, pp.48-49). For easier viewing, the figure presents the quarterly averages of the pseudo-R² values.

Comments: 0

 Comments

The Impact of Alternative Response Scales on Measuring Self-ratings of Health

Friday, August 22, 2008, 5:13:10 PM | Editor

Tom W. Smith
NORC
University of Chicago

Introduction

Following the First Law of Studying Societal Change, the General Social Survey (GSS) strives for consistent measurement over time by employing constant measures.⁽¹⁾ However, in certain cases measures have been changed for various reasons. When such alterations occur, the GSS has introduced the revised version in a controlled manner, typically using some combination of across-subjects experiments and within-subjects repetition. This procedure is important so that variation due to measurement effects is not confounded with studying true change. This report considers a possible change in the GSS measure of self-rated health.

Self-rated Health

Since 1972, the GSS has included a self-rated health measure (HEALTH - Would you say your own health, in general, is excellent, good, fair, or poor?). This simple item is widely used in health studies and is a notable predictor of mortality and other health outcomes, even controlling for other variables such as specific health history and medical evaluations.⁽²⁾ The GSS wording came from Gallup surveys in 1941 and 1950. In the 1970s about half of major US national studies measuring self-rated health employed a 4-category response scale and half used a 5-category version (Danchick and Drury, 1986). When the National Health Interview Survey (NHIS) was redesigned in 1982, it switched from a 4-category version to a 5-category format.⁽³⁾ Consistent with that decision, virtually all US governmental, health surveys now use 5-category versions (e.g. the National Health Examination Survey, the Health and Retirement Study, the Study of Assets and Health Dynamics, the Behavioral Risk Factors Surveillance Study), as do most other health scales (e.g. SF-36). Besides the GSS, relatively few studies continue to employ a 4-category version.⁽⁴⁾

As Kovar and Poe (1985) note, the NHIS study switched to five categories in order “to improve the ability to differentiate among people” and others have preferred it for similar reasons. The unarticulated expectation was that the finer measurement would more accurately measure health status and produce stronger associations with health variables and demographics.

On the GSS and other studies a variety of comparisons between the different response scales used for the subjective health measures exists. These include non-experimental comparisons and experiments using both inter-subject and intra-subject designs. [Table 1](#) examines the impact of the 4- and 5-category response scales on marginals. [Table 1A](#) Table 1A looks at non-experimental comparisons in which different surveys of similar populations were conducted at approximately the same time and [Table 1B](#) covers intra- and inter-subjects experiments. In the intra-subjects design people were asked both versions of the self-rated health question in different parts of the survey. In the inter-subjects design, different random samples were given 4 or 5 categories versions.

Adding the fifth “very good” category takes responses from the more positive “excellent” option and the less positive “good” option and reduces both. The declines in “excellent” range from 4.9 percentage points to 16.8 points and “good” decreases from between 15.4 points to 21.2 points. There is considerable difference as to whether most of the “very good” responses appear to come from “excellent” or “good”. The decline in “excellent” apparently contributes as little as 19.5% of the “very good” responses ([Table 1A-2](#)) to as much as 61.5% ([Table 1A-3](#)). The differences are even notable within the experimental studies. There is little impact on the distribution of “fair” and “poor” response across response scales. An [intra-subjects design](#) among employed adults on the 2002 GSS confirms the very limited impact on these two more negative responses. The impact of the changes in response scales on distributions is large, but variable, making any simple comparison across the response scales difficult.

Next, the associates of health are examined (Table 2). This examines whether the two items reveal the same structural relationships, and tests the hypothesis that the finer scale yields stronger correlations. Overall there is no meaningful difference in the strength or statistical significance of associations. The average absolute correlations were .130 for the former and .132 for the latter.

The lack of any meaningful and consistent difference in correlations is not surprising since several previous GSS studies showed little or no impact on associations of using response scales with more categories.⁽⁵⁾ It is also expected because on the 2002 GSS the correlation between 4- and 5-category health items is 0.85 and if Excellent on the 4-category scale is considered consistent with Excellent or Very Good on the 5-category scale and likewise Good with Very Good or Good, that means that 93.6% of the cases are on the diagonal when crosstabulating the items. Also, as indicated above, there is little impact on the bottom two categories and Singer (1994) argues that the “predictive value of self-rated health is driven by ratings of fair or poor health”.

Summary

This evaluation of self-rated, health items indicates that 1) no discernable difference in the explanatory power of the two scales occurs, 2) major shifts in the distributions happen at the positive end, but little at the negative end, 3) the variation in the contributions from Excellent and Good to the added Very Good option would not allow trends in these categories to be reliably estimated across scales and, as a result, would restrict trend analysis combining both 4- and 5-category data points to comparing the bottom two responses with the combined top two or three categories, and 4) correlations across studies using the 4- and 5-category scale might be compared since they do not produce different estimates.

The large impact of the shift in response scales over part of the distribution and the unexpected nil impact on correlations underscores that survey researchers must be careful whenever changing methods. Changing methods should always be presumed to muddy, if not eviscerate, valid comparisons. Additionally, changes will often not yield the improvements expected. When modifications are introduced, experiments and other rigorous designs should be utilized and any expected improvements need to be verified.

Table 1

Comparisons of the Distributions of Self-Rated Health

Using 4 or 5 Response Options

A. Non-Experimental

1. 1981 and 1982 National Health Interview Survey (NHIS)

	1981	1982
Excellent	42.0%	32.2%
Very Good	—	25.4
Good	41.2	25.8
Fair	12.7	11.5
Poor	4.1	5.1

Source: Danchik and Drury, 1986

2. 1979 NHIS and 1979 Fourth Quarter Evaluation Study (FQES)

	NHIS	FQES
Excellent	42.8%	30.6%
Very Good	---	28.8
Good	40.3	24.7
Fair	12.8	11.4
Poor	4.1	4.4

Source: Danchik and Drury, 1986

3. 1976 NHIS and National Health and Nutrition Examination Survey II (ages 20-74)

	NHIS	NHANES
Excellent	43.9%	27.1%
Very Good	---	27.3
Good	40.1	27.9
Fair	11.9	12.5
Poor	3.7	5.0
Missing	0.4	0.2

Source: Forthofer, 1983

B. Experimental (Intra- and Inter-Subject Designs)

1. NHIS Inter-Subjects Experiments, 1979

	Standard	Variant
Excellent	48.0%	36.0% ^a
Very Good	---	28.0
Good	39.0	21.0
Fair	10.0	8.0
Poor	3.0	3.0

B.1.Variant total adds up to only 96% in original source.

Source: Kovar and Poe, 1985

2. General Social Survey, 2002 (Intra-Subjects Experiment; Employed People)

	Standard	Variant
Excellent	35.9%	31.0%
Very Good	---	25.1
Good	48.7	30.0
Fair	13.3	12.1
Poor	2.1	1.8

	1193	1186
--	------	------

Source: GSS

3. General Social Survey, 2004 (Inter-subjects experiment)

	Standard	Variant
Excellent	35.7%	26.3%
Very Good	---	30.6
Good	47.8	26.5
Fair	12.2	11.4
Poor	4.3	5.3
	466	517

Source: GSS

Table 2

Correlates of 4-Category and 5-Category Health Self-Ratings (Pearsons r/probability)

A. 2002 GSS (Employed People)

	4-Category	5-Category
Age (AGE)	.027/.355	.044/.129
Gender (SEX)	-.023/.425	.013/.644
Race (RACE)	.064/.026	.040/.160
Education (EDUC)	-.196/.000	-.177/.000
Occ. Prestige (PRESTGE80)	-.149/.000	-.165/.000
Attend Church (ATTEND)	-.091/.002	-.075/.009
Frequency of Praying (PRAY)	.028/.497	.005/.902
Happiness (HAPPY)	.258/.000	.234/.000
Life Exciting (LIFE)	.223/.000	.201/.000
Physical Health (PHYSHLTH)	.316/.000	.313/.000
Mental Health (MNTLHLTH)	.224/.000	.213/.000
Health Days, Month (HLTHDAYS)	.178/.000	.188/.000
Feel Used Up by Job (USEDUP)	-.140/.000	-.165/.000
Suffer Back Pain (BACKPAIN)	-.154/.000	-.175/.000
Pain in Arms (PAINARMS)	-.126/.000	-.154/.000
Hurt at Work (HURTATWK)	.050/.034	.043/.137
Gov Health Spending (NATHEAL)	-.057/.047	-.069/.017
Medical Confidence (CONMEDIC)	.125/.028	.117/.039

B. 2004 GSS (All Adults)

	4-Category	5-Category

Age (AGE)	.198/.000	.181/.000
Gender (SEX)	.008/.868	-.078/.076
Race (RACE)	.028/.530	.043/.718
Education (EDUC)	-.274/.000	-.328/.000
Occ. Prestige (PRESTGE80)	-.184/.000	-.218/.000
Attend Church (ATTEND)	-.035/.447	.015/.732
Mental Health (MNTLHLTH)	.285/.000	.061/.256
Job Stress (WRKSTRESS)	-.050/.049	-.122/.000
Gov Health Spending (NATHEAL)	.021/.338	-.041/.773
Respondent's Weight Judged by Interviewer (INTRWGHT)	.131/.051	.230/.000

Note: Variables names are in parentheses and these items can be found in Davis, Smith, and Marsden, 2005.

References

Alwin, Duane F., "Information Transmission in the Survey Interview: Number of Response Categories and the Reliability of Attitudes Measurement", in Sociological Methodology, 1992. edited by Peter V. Marsden. Washington, DC: Blackwell, 1992.

Danchik, Kathleen M. and Drury, Thomas F., "Evaluating the Effects of Survey Design and Administration on the Measurement of Subjective Phenomena: The Case of Self-Assessed Health Status", Proceedings of the Survey Methods Research Section, 1985. Washington, DC: American Statistical Association, 1986.

Davis, James A.; Smith, Tom W.; and Marsden, Peter V., General Social Surveys, 1972-2006: Cumulative Codebook. Chicago: NORC, 2007.

Davis, Wendy; Wellens, Tracy R.; and DeMaio, Theresa J., "Designing Response Scales in an Applied Setting", Working Paper in Survey Methodology, 96/7, US Bureau of the Census, 1996.

Ferraro, Kenneth F. and Farmer, Melissa M., "Utility of Health Data from Social Surveys: Is There a Gold Standard for Measuring Morbidity?" American Sociological Review, 64 (1999), 303-315.

Forthofer, Ronald N., "Investigation of Nonresponse Bias in NHANES II", American Journal of Epidemiology, 117 (1983), 507-515.

Groves, Robert M.; Fultz, Nancy H.; and Martin, Elizabeth, "Direct Questioning about Comprehension in a Survey Setting", in Questions about Questions: Inquiries into the Cognitive Bases of Surveys, edited by Judith M. Tanur. New York: Russell Sage, 1992.

Hardy, Melissa and Pavalko, Eliza K., "The Internal Structure of Self-Reported Health Measures Among Older Male Workers and Retirees", Journal of Health and Social Behavior, 27 (1986), 346-357.

Idler, Ellen L. and Angel, R.J., "Self-Reported Health and Mortality in the NHANES-I Epidemiological Follow-up Study", American Journal of Public Health, 80 (1990), 446-452.

Idler, Ellen, L. and Benyamini, Yael, "Self-rated Health and Mortality: A Review of Twenty-Seven Community Studies",

Journal of Health and Social Behavior, 38 (1997), 21-37.

Kovar, Mary Grace and Poe, G.S., "The National Health Interview Design, 1973-1984 and Procedures, 1975-1983", Vital and Health Statistics, Series 1, No. 18. Washington, DC: GPO, 1985.

Mallinson, S., "Listening to Respondents: A Qualitative Assessment of the Short-Form 36 Health Status Questionnaire", Social Science Medicine, 54 (1982), 11-21.

Perry, M. et al., "Factors Associated with Self-Perceived Excellent and Very Good Health Among Blacks", MMWR, 45 (Oct. 25, 1996), 906-911.

Petersen, Bruce L., "Confidence: Categories and Confusion", GSS Methodological Report No. 31. Chicago: NORC, 1985.

Ratner, P.A.; Johnson, J.L.; and Jeffery, B., "Examining Emotional, Physical, Social, and Spiritual Health as Determinants of Self-rated Health Status", American Journal of Health Promotion, 12 (1998), 275-282.

Remle, R. Corey, "Self-rated Health Trajectories: Alternative Measures of Perceived Change in Health as a Predictor of Mortality", Paper presented to the American Sociological Association, San Francisco, August, 2004.

Schechter, Susan, "Investigation into the Cognitive Process of Answering Self-Assessed Health Status Questions", CDC Working Paper No. 2. Washington, DC: NCHS, 1993.

Sehulster, Jerome R., "Health and Self: Paths for Exploring Cognitive Aspects Underlying the Self-Report of Health Status", Proceedings of the 1993 NCHS Conference on the Cognitive Aspects of Self-Reported Health Status, edited by Susan Schechter. NCHS Working Paper No. 10, 1994.

Siegel, Paul Z., "Self-Reported Health Status: Public Health Surveillance and Small-Area Analysis", Proceedings of the 1993 NCHS Conference on the Cognitive Aspects of Self-Reported Health Status, edited by Susan Schechter. NCHS Working Paper Series, No. 10, 1994.

Singer, Eleanor, "Self-Rated Health Status: How Are Judgements Made?", Proceedings of the 1993 NCHS Conference on the Cognitive Aspects of Self-Reported Health Status, edited by Susan Schechter. NCHS Working Paper No. 10, 1994.

Smith, Tom W., "A Comparison of Two Confidence Scales", GSS Methodological Report No. 80. Chicago: NORC, 1994a.

Smith, Tom W., "A Comparison of Two Government Spending Scales", GSS Methodological Report

1. The GSS is the largest and longest-term project of the Sociology Program of the National Science Foundation. It has conducted 26 national, in-person, full-probability surveys of adults living in US households between 1972 and 2006 (Davis, Smith, and Marsden, 2007).

2. See Ferraro and Farmer, 1999; Hardy and Pavalko, 1986; Idler and Angel, 1990; Idler and Benyamini, 1997; Perry et al., 1996; Remle, 2004; Siegel, 1994.

3. The NHIS is the main, continuous health monitoring study of the household population conducted by the US government. For more information see www.cdc.gov/nchs/about/major/nhis/hisdesc.htm On the switch see Kovar and Poe (1985).

4. On the meaning of the self-rated health measure and how evaluations are done by respondents see Groves, Fultz, and Martin, 1992; Mallinson, 2002; Schechter, 1993; Schulster, 1994; Singer, 1994.

5. On the GSS, see Peterson, 1985; Smith, 1994a; 1994b. Alwin (1992) found a slight increase in reliability moving from 4 to more than 4 categories, but Davis, Wellens, and DeMaio (1996) found no gains between 4 categories and 5-6 categories.

Comments: 0

 Comments

FieldNotes

Friday, August 22, 2008, 4:03:14 PM | Editor

Fieldnotes - a not so random sample of ideas and activities in survey research

In April, the American Association for Public Opinion Research (AAPOR) Executive Council approved a [report from the Cell Phone Task Force](#). The report, along with the special POQ issue on cell phones, provides a comprehensive look at the impact of cell phones on telephone surveys. They also identify the research that is needed to bring cell phone interviewing into the everyday part of telephone surveys. The report is long but has important information in it. Well worth the time to read it.

Book Reaction (not a book review) - Earlier this year, I read Conrad and Schober's new book - *Envisioning the Survey Interview of the Future*. It is an edited volume and the chapters are written by experts in survey research and in other professions that may be involved in future developments in survey methodology. The book is important and contains much valuable information for survey researchers. You can learn more about IVR and ECAs (avatar interviewing).

You should read it, especially if you want to find out about ECAs and survey research. However, at the end, I didn't think that it had much for me, in my day-to-day activities as a survey center director, to act on now. It's still too early to predict how these new methods will evolve.

Takeaway: Read it; it's an important book that contains good information on where survey interview may go in the next few years. But, it doesn't have much information that would be used to change your organization or operations.

Speaking of survey centers, about 30 directors of **academic survey research organizations** met in Berkeley in early March to discuss issues related to our niche in the survey world. The group has been meeting informally since 1995. At the New Orleans AAPOR conference, the group passed the bylaws and the Association of Academic Survey Research Organizations (AASRO) is now organized.

Question of the month - do you think IRBs are easing up on survey research? From my observations and what I hear, I think they are starting to realize the nature of risk in survey research and are making it easier for

protocols to be approved. What do you think?

If you have thoughts or issues that you would like included in Fieldnotes, send them to survprac@indiana.edu.

John Kennedy

Comments: 0

 Comments

Ask the Experts

Friday, August 22, 2008, 3:55:34 PM | Editor

Starting next month, Survey Practice will have an "Ask the Experts" column. Please submit questions related to survey research and public opinion polling methods, processes, or procedures to the editor. I will select an interesting question and ask 2-3 experts to give relatively short answers to the questions.

Some example questions:

- What are good quality control procedures in postal mailing assembly?
- What's better: a four- or five-point scale? (the question I get most often)
- How are participants tracked in panel studies?
- What's a good question to measure political ideology?

To submit a comment, click on "comments" below or send an email message to survprac@indiana.edu.

Comments: 1

 Comments