

# **The General Social Survey (GSS) The Next Decade and Beyond**

**National Science Foundation  
Workshop on Planning for the Future of the GSS**

**May 2-3, 2007**

**National Science Foundation  
Arlington, Virginia**

## Participants and Attendees

**Patricia White**, National Science Foundation  
Workshop Organizer

**Frank Scioli**, National Science Foundation  
Workshop Moderator

**Andrew Beveridge**, City University of New York Queen's College

**Robert Bell**, National Science Foundation

**Suzanne Bianchi**, University of Maryland

**Norman Bradburn**, National Opinion  
Research Center

**Linda Carlson**, National Science Foundation

**Mark Chaves**, University of Arizona

**Paul Ciccantell**, National Science Foundation

**James Davis**, National Opinion Research Center

**Cheryl Eavey**, National Science Foundation

**Barbara Entwistle**, University of North Carolina - Chapel Hill

**Jeremy Freese**, University of Wisconsin

**Kevin Gotham**, National Science Foundation

**Peter Granda**, Inter-university Consortium for  
Political and Social Research (ICPSR) University of Michigan

**Edward Hackett**, National Science Foundation

**Michael Hout**, University of California, Berkeley

**Brian Humes**, National Science Foundation

**Ronald Inglehart**, University of Michigan

**Jon Krosnick**, Stanford University

**David Lightfoot**, National Science Foundation

**John Logan**, Brown University

**Robert Mare**, University of California, Los Angeles

**Peter Marsden**, Harvard University

**Douglas Maynard**, University of Wisconsin

**Leslie McCall**, Northwestern University

**Daniel Newlon**, National Science Foundation

**Steven Nock**, University of Virginia

**Philip Paolino**, National Science Foundation

**Gregory Price**, Jackson State University

**Tom Smith**, National Opinion Research Center

**Lynn Smith-Lovin**, Duke University\*

**Steven Ruggles**, University of Minnesota\*

**Roberta Spalter-Roth**, American Sociological Association

**Frank Stafford**, University of Michigan

**Steven Tuch**, George Washington University

\* Submitted paper; did not attend meeting.

## Executive Summary

The National Science Foundation (NSF) began supporting the General Social Survey in the early 1970's and has continued to do so with a grant to the National Opinion Research Center (NORC) in 2005 to complete the 2006 and 2008 surveys. The GSS, along with the Panel Study of Income Dynamics and the American National Election Studies, is one of the three major infrastructure projects supported by the Division of Social and Economic Sciences. The GSS is an important survey data resource used by sociologists and other social scientists for research and teaching.

On May 2-3, 2007, the Division of Social and Economic Sciences convened a workshop on *The General Social Survey (GSS): The Next Decade and Beyond* at the National Science Foundation (NSF) in Arlington, Virginia. The Sociology Program funded and organized the workshop. The purpose of the workshop was to hold a conversation with the social science community about how best to maintain and enhance the GSS as NSF moves forward with plans for a recompetition for the conduct of the 2010 survey and beyond. Workshop participants were asked to assist the NSF in completing an assessment of the GSS as a major NSF infrastructure investment and to “assure the best use of NSF funds for supporting research and education.”

The workshop brought together twenty-one experts in survey research methodology and scholars with intimate knowledge of the GSS. These participants prepared background papers and discussed major areas of the GSS, including operations and governance, survey topics and content, conceptual and methodological innovations, collaboration and integration with other major national and international survey programs, outreach to the social science and other scholarly communities, dissemination of data and results dissemination, and other broader impacts. These background papers are included in the second section of this report. The workshop discussions and papers provided advice and generated recommendations to NSF on how to improve and strengthen the GSS as it moves into the next decade.

The report is organized in three major sections – *Background of the GSS* (and the Workshop); a descriptive *Overview of the GSS*; and a list of workshop *Recommendations*. The first section provides a brief discussion of the context from which the idea for the workshop emerged, the purpose of the workshop, and its role in the 2010 GSS recompetition. The second gives an overview (including a chronology) of the major components of the GSS. This overview covers the purpose and development of the GSS, current GSS design, survey development, the GSS core, recent innovations, cross-national research, experimentation and methodological research, and data dissemination and usage. The final section lists workshop recommendations, organized around three topics—(1) The Basic GSS, What to Maintain and Continue; (2) Governance, Survey Administration and Funding; and (3) Data Collection, Dissemination and Outreach. Additionally, the report Appendix includes the workshop agenda and papers submitted by the workshop participants.

## Summary of Recommendations

Workshop participants provided a number of insightful recommendations to NSF on how best to improve and strengthen the GSS as it moves into the next decade.

### The Basic GSS- What to Maintain and Continue

*The GSS should remain a nationally representative survey of attitudes and intergroup relations and continue to monitor trends in attitudes and behaviors.* It should maintain and further enhance its openness to the addition of new topics and questions by the social science community.

*Continue to administer the GSS on a biennial cycle.* Since time is the critical independent variable in administering the GSS, it should maintain the biennial schedule.

*Develop plans to maintain the panel design.* The panel component is a valuable addition to the GSS and allows for the measurement of individual change over time on many variables, particularly changes in attitudes, allowing the direct study of temporality and causation. The panel study also allows GSS users to study issues that could not be studied with the earlier, purely cross-sectional survey design.

*Maintain GSS participation in the International Social Survey Program (ISSP) data collection program.* The ISSP is an important vehicle that can be used to study important social processes in a comparative perspective by both examining differences across societies and changes within the current 43 nations over time.

*Continue a timely and comprehensive release of all data.* Data from the GSS, both the core and topical modules, should continue to be made available to all users at the same time and as soon as possible after collection and data cleaning.

### Governance, Survey Administration and Funding

*Strengthen GSS governance and the role of the Board of Overseers.* Significantly enhance and strengthen the role of the GSS Board in evaluating the content of the survey, including the core and topical modules.

*Enable the Board of Overseers to participate more fully in the selection and development of GSS topical modules.* Topical modules provide a vehicle for innovation and should remain part of the future GSS, but the Board of Overseers should have a stronger voice in evaluating topical modules for scientific merit.

*Provide support for the GSS Board of Overseers to hold a “module” competition.* The quality of the GSS would be further enhanced if funds were provided to the GSS Board of Overseers to allow them to manage competitions to determine the addition of modules and questions to the

GSS. This competition would be open to social scientists who could propose topics and questions to add to the GSS.

*Explore ways to realize cost savings in survey administration.* The GSS should explore alternative modes of administration in order to reduce per-interview costs.

*Seek innovative modes of administering the GSS.* There is considerable value in maintaining the face-to-face CAPI interview process. The future GSS, however, should examine incorporating other modes of administering the survey, including a leave-behind questionnaire and internet surveys by providing internet access and even laptop computers to allow those in the sample to respond to a series of internet surveys over time.

*Study the impact of the split-ballot technique on data analyses.* The split-ballot administration that gives each respondent two-thirds of the core may create significant problems for analysis. The GSS should study the impact of this aspect of survey administration and, in particular, explore ways to eliminate the split-ballot of the core items.

*Share the funding of the GSS across other social science and education fields.* The GSS is a valuable survey data resource for disciplines other than sociology, with less than 50 percent of GSS users identifying themselves as sociologists. The GSS is heavily used as a teaching and research tool at both the undergraduate and graduate levels. NSF is encouraged to seek joint funding of the GSS more broadly across the social sciences at NSF.

## **Data Collection, Dissemination and Outreach**

*Make the GSS core more transparent to the user community and allow for changes in core items.* The content of the core and how it has evolved over time should be made clearer to all potential users of the GSS. The GSS should sketch out the core set of items, but the core should be allowed to evolve over time via the interaction of the GSS Board, GSS PIs, and the user community. Also, the possibility of enlarging the core, including bringing back time series questions that were deleted in the early 1990s, should be explored.

*Explore ways to enhance the data collected.* The GSS should explore the possibility of oversampling minority groups in order to allow comparisons between groups. It should also consider collecting and making available to users much more “paradata/metadata,” including interviewer characteristics, the mode of administering the interview, the number of calls made to obtain the interview, characteristics of the home and neighborhood, respondent reaction times for answering questions, and spatial identifiers. These data would be released subject to full protection of respondent confidentiality, which in some cases will include special protections and restricted access as all major data projects now do.

*Experiment with digital-recording of interviews.* Explore ways to digitally record interviews using the computers that interviewers use to conduct the interview. Such recordings would enable studies of “how” the interview is done; these studies would help improve data quality and encourage the integration of qualitative and quantitative research.

*Provide opportunities for experimentation.* The GSS should encourage the embedding of both methodological and substantive experiments in the survey.

*Develop “targeted” GSS dissemination activities.* The current NSF funding model does not directly address the issues of data access and dissemination. Targeted NSF funding is needed for the dissemination of data, including providing web access and increasing ease of access to the data for research and teaching purposes.

*Upgrade and create more user-friendly modes of dissemination.* The current modes of dissemination are confusing and badly outdated. The future GSS should include a state-of-the-art dissemination system that is updated frequently, including a user-friendly website, a high quality index, easy to use search tools, thorough documentation about survey procedures, questions, and data files, simple data downloading in a variety of formats, accessible technical assistance, and simple on-line analysis tools for users.

*Expand the type of data products that are publicly available.* In addition to disseminating the complete dataset, the future GSS should explore making other data products available to users, such as datasets of only core items, each topical module, and each minority group. Also, any data linked to GSS data by users should become part of the GSS publicly available data, including the computer code necessary to replicate analyses.

*Secure a separate budget for data dissemination.* Proposals for the future GSS should include an extensive plan for data dissemination with an adequate budget, whether dissemination is to be done in-house or via a third party. Past NSF grants have not included adequate funds for outreach and dissemination.

*Develop cooperative relationships with other major NSF surveys.* Both the principal investigators and the Boards of other NSF-funded surveys, in particular the Panel Study of Income Dynamics and American National Election Studies, should work cooperatively to share information and technologies. They should also consider greater coordination and integration, such as common data access mechanisms and repositories, as well as using common measures for demographic variables.

## Table of Contents

<b>Executive Summary.....</b>	<b>4</b>
<b>An Overview of the General Social Survey.....</b>	<b>10</b>
<b>Current GSS Design.....</b>	<b>11</b>
<b>Survey Development.....</b>	<b>11</b>
<b>The GSS Core.....</b>	<b>13</b>
<b>Recent Innovation in the Survey.....</b>	<b>13</b>
<b>GSS Panel Component.....</b>	<b>14</b>
<b>Spanish Language Translation.....</b>	<b>14</b>
<b>Cross-National Research.....</b>	<b>15</b>
<b>Experimentation &amp; Follow-up Studies.....</b>	<b>15</b>
<b>Data Dissemination and Usage.....</b>	<b>15</b>
<b>Recommendations.....</b>	<b>16</b>
<b>The GSS- What to Maintain and Continue.....</b>	<b>17</b>
<b>Governance, Survey Administration and Funding.....</b>	<b>17</b>
<b>Data Collection, Dissemination and Outreach.....</b>	<b>18</b>
<b>Appendices</b>	
<b>Appendix 1: Workshop Agenda.....</b>	<b>20</b>
<b>Appendix 2. Workshop Papers.....</b>	<b>23</b>
<b>Andrew Beveridge, The General Social Survey and Its Impact in Sociology and Other Social Sciences.....</b>	<b>24</b>
<b>Suzanne Bianchi, The GSS, the ANES and the PSID &amp; GSS Methodology Comments Prepared for the NSF Workshop on Planning the Future of the GSS.....</b>	<b>27</b>
<b>Norman Bradburn, Thoughts on the General Social Survey.....</b>	<b>31</b>
<b>Mark Chaves, The General Social Survey: Innovation and Dissemination Comments for NSF Workshop on the GSS.....</b>	<b>33</b>
<b>Barbara Entwisle, Hard Choices: Reflections on the Design of the General Social Survey.....</b>	<b>36</b>
<b>Jeremy Freese, Commentary for GSS Workshop, Methodology &amp; Technological Innovation and Cyberinfrastructure.....</b>	<b>40</b>
<b>Peter Granda, Best Practices in the Dissemination of Survey Data.....</b>	<b>43</b>
<b>Ronald Inglehart, The GSS and International Surveys: Issues and Opportunities.....</b>	<b>47</b>
<b>Jon Krosnick, Thoughts on the GSS Recompetition.....</b>	<b>50</b>
<b>Robert Mare, Operational Aspects of the GSS from the Standpoint of Board of Overseers.....</b>	<b>57</b>
<b>Douglas Maynard, Issues of Data Quality and Data Generation: The General Social Survey and Ethnomethodology/Conversation Analysis.....</b>	<b>61</b>
<b>Leslie McCall, Review of the Content of the GSS.....</b>	<b>65</b>

<b>Steve Nock, Conceptual and Methodological Innovations &amp; Contribution of the GSS to Sociology and Its Broader Impacts.....</b>	<b>67</b>
<b>Gregory Price, The General Social Survey: Contributions to Economics And Recommendations for Future Dissemination.....</b>	<b>70</b>
<b>Steven Ruggles, Review of Web-Based Dissemination of the General Social Survey.....</b>	<b>73</b>
<b>Lynn Smith-Lovin, GSS Content and Innovations.....</b>	<b>77</b>

## Background

The National Science Foundation (NSF) began supporting the General Social Survey (GSS) in the early 1970's, funding the first survey in 1972. It has continued to do so with a renewal grant to the National Opinion Research Center (NORC) in 2005 to complete the 2006 and 2008 surveys. The GSS, along with the Panel Study of Income Dynamics (PSID) and the American National Election Studies (ANES), is one of the three major infrastructure projects supported by the Division of Social and Economic Sciences. The GSS is funded and managed by the Sociology Program, which allocates approximately twenty percent of its annual budget to the support of the GSS. The GSS is a major survey data resource used by sociologists and other social scientists for research and teaching, and in these roles adds tremendous value to the conduct of basic social science research.

The National Science Board in a Resolution Concerning Competition, Recompetition and Renewal of NSF Awards (NSB 97-224) affirmed strong support for the principle that “expiring awards are to be recompeted unless it is judged to be in the best interest of U.S. science and engineering not to do so.” The Sociology Program recognizes the important scholarly and scientific achievements and accomplishment of the GSS, but also its challenges for the future. To assess how to best maintain and enhance this important data resource, while also positioning it to take advantage of innovations in survey research and information technology, the Program convened a workshop—*The General Social Survey: The Next Decade and Beyond Workshop on Planning for the Future of the GSS*—on May 2-3, 2007 to solicit advice from the social science research community. The major focus of the workshop was to discuss methodological and substantive challenges of the GSS in 2010 and beyond and to ensure the best use of NSF funds for supporting social science research infrastructure.

The workshop brought together twenty-one experts in survey research methodology and scholars with intimate knowledge of the GSS. These participants prepared background papers and discussed major areas of the GSS, including operations and governance, survey topics and content, conceptual and methodological innovations, collaboration and integration with other major national and international survey programs, outreach to the social science and other scholarly communities, dissemination of data and results dissemination, and other broader impacts. These background papers are included in the second section of this report. The workshop discussions and papers provided advice and generated recommendations to NSF on how to improve and strengthen the GSS as it moves into the next decade.

## **An Overview of the General Social Survey\***

The General Social Survey (GSS) has provided a wealth of data on contemporary American society for approximately 35 years by measuring social change and trends and constants in attitudes, behaviors and attributes of the adult population. The GSS is a regular, ongoing interview survey of U.S households conducted by the National Opinion Research Center. The mission of the GSS is to make timely, high-quality, scientifically relevant data available to social science researchers. The GSS is a personal interview survey and collects information on a wide range of demographic characteristics of respondents and their parents; behavioral items such as group membership and voting; personal psychological evaluations, including measures of happiness, misanthropy, and life satisfaction; and attitudinal questions on such public issues as abortion, crime and punishment, race relations, gender roles, and spending priorities. Since 1972 the GSS has conducted 26 in-person, cross-sectional surveys of the adult household population of the U.S. Interviews have been conducted with a total of 51,020 respondents. The 1972-74 surveys used modified probability designs and the remaining surveys were completed using a full-probability sample design, producing a high-quality, representative sample of the adult population of the U.S. The GSS has a response rate of over 70 percent above that of other major social science surveys and 40-45 percentage points higher than the industry average.

### **Current GSS Design**

The basic GSS design is a repeated cross-sectional survey of a nationally representative sample of non-institutionalized adults who speak either English or Spanish. Subsampling of non-respondents is done to limit survey costs while maintaining a nationally representative sample. Each GSS formally includes an A sample and a B sample. The preferred interview mode is in-person interviews; however, a few interviews will be done by telephone in the event that an in-person contact cannot be scheduled. Each respondent is asked the replicating core of socio-demographic background items, along with replicated measurements of sociopolitical attitudes and behaviors. Many of the latter are measured by way of a “ballot” design such that each item is answered by a random 2/3 of each sample. Each GSS sample (A and B) includes an International Social Survey Program module (ISSP). Each sample is also asked to respond to several topical modules that may be supported by NSF or others, but are no longer supported by the basic grant from NSF. Some of these topical modules, however, extend across both samples in a given GSS survey.

---

\* This section summarizes materials provided by the GSS PIs in funded proposals, reports, and commentary prepared for the workshop.

## A Selective Chronology of the General Social Survey\*

1972	First GSS; subsequently conducted almost annually until 1993
	Initial sampling design was block quota/modified probability
	Many replicating core items measured on rotation design (2 years on and 1 year off)
	Board of Advisors established; remained in existence until 1983
1975	Shift to full-probability sampling design
1977	Board of Methodological Advisors added; operated until 1983
1982	First African-American oversample
	Bilateral cross-national collaboration with ALLBUS initiated
1983	Board of Overseers established
1985	Expansion of topical modules initiated
	First International Social Survey Program module; topic was the role of government
1987	Second African-American oversample
1988	Split-ballot rotation system for replicating core items adopted
1991	First auxiliary study (National Organizations Study)
1994	Shift to two-sample design with target N of 3000
	Major reduction in size of replicating core to accommodate more topical modules
	Family mobility module and affiliated sibling study
1996	Segments of replicating core on race, gender, religion refreshed to alter series tracked (until 2003)
1998	First National Congregations Study
2000	Clergy and Congregational Attendees studies
2002	Shift to computer-assisted personal interviewing (CAPI) methods
	Second GSS-linked National Organizations Study
2004	Subsampling of non-respondents design adopted
2006	Target population expanded to include Spanish-speaking adults
	Third sample added to accommodate coverage of topical modules
	Baseline wave of data collected for first GSS panel, spanning 2006-2010
	National Voluntary Associations Study underway, linked to 2004 data
	Second National Congregations Study now in progress

### Survey Development

The GSS has six components, which include a replicating core, topical modules, cross-national modules, experiments, re-interviews, and follow-up studies. The replicating core makes up one-third of the GSS and the topical and cross-national modules the other two-thirds. Experiments are conducted in both the core and supplemental modules. Re-interview and follow-up studies are completed through additional data collections. The replicating core is the part of the GSS that the Sociology Program has continually supported over the past 35 years. The “core” consists of questions that regularly appear on the GSS. The contents of the core are periodically reviewed by the grant principal investigators and Board of Overseers (consisting of a multi-

\* Prepared by Peter Marsden for the workshop.

disciplinary group of scholars with expertise in survey and other social science research methodologies) to insure that the content remains relevant. Core questions are updated when deemed necessary by the Board of Overseers. The core is about one-third demographic information and two-thirds items that capture attitudes and behaviors. GSS content is wide ranging with 5084 variables overall in the 1972-2006 file. The topical modules are used to introduce new topics not previously investigated by the GSS and to cover existing topics in greater detail with more fully specified models. The concept for a module may originate with the PIs, the Board, or other scholars. Many prominent scholars help develop topical modules. For example, 74 researchers representing many fields from 48 universities and research institutes have served on the GSS Board and 253 social scientists in a dozen disciplines from 147 institutions have participated in the design of the first 28 topical modules.

## **The GSS Core**

The GSS is a 90-minute in-person interview. Forty-five minutes of the GSS are devoted to the core items, 15 minutes are devoted to the questions selected as part of the International Social Survey Program (ISSP), and 30 minutes are allocated to topical modules that are funded by sources other than the main NSF grant for the GSS. The socio-demographics in the core are administered to all respondents and most attitudinal and behavioral measures are administered on split-ballots with background items and a range of replicated measures of sociopolitical attitudes and behaviors. The core questions are administered as a split-ballot, with each respondent answering two-thirds of the core questions. Topical modules are focused on a wide range of substantive issues and are funded by a variety of government agencies, private foundations, and other organizations. These funds supplement the main NSF grant funding and allow the GSS to fully fund the 90 minute survey and collect a larger sample than would be possible with only NSF funds.

The GSS has a “replicating core” that emphasizes collection of data on social trends through exact replication of question wording over time. Core items fall into two major categories— socio-demographic/background measures, and replicated measurements on social and political attitudes and behaviors. Many of the latter items appear on three GSS “ballots”, each of which is administered to a random two-thirds of most GSS samples. Over the course of the project, there have been many changes in replicated items. In addition, there are “quasi-core” items that are repeatedly funded by other sources. In addition to the planned trend items included in the GSS core, there are other data series that arise through unplanned repetition of topical modules and the replication of ISSP items.

## **Recent Innovations**

The 2006 GSS has two major ongoing innovations. First, it serves as the baseline sample for the new GSS panel component, with a sub-sample of cases to be re-interviewed in 2008 and 2010. Second, the GSS core was translated into Spanish and administered in either English or Spanish as needed in 2006. This practice will continue in 2008 and 2010.

## GSS Panel Component

The GSS now includes a panel study component for the first time in order to allow the direct observation of change over time in the same individuals. The GSS switched from a repeating, cross-section design to a combined repeating cross-section and panel-component design. The 2006 GSS was the base year for the first panel. A sub-sample of 2006 GSS cases (most likely about 2000) will be selected for reinterview in 2008 and again in 2010 as part of the GSS in those years (see Table 1). The 2008 GSS will consist of a new cross-section of about 2000 plus the 2006 reinterviews. The 2010 GSS will consist of another new cross-section of about 2000, the second reinterview wave of the 2006 panel cases and the first reinterview wave of the 2008 panel cases. The 2010 GSS will be the first one to fully implement the new, combined design. In 2012 and later General Social Surveys, there will likewise be a fresh cross-section, wave two panel cases from the immediately preceding GSS, and wave three panel cases from the next earlier GSS.

### **Rotating, Three-Wave Panel/Cross Section Design**

Samples	2006 n=	2008 n=	2010 n=	2012 n=
New Cross Section	3000	2000	2000	2000
Initial Reinterview Target	2000	2000	2000	2000
Expected Wave 2 Completed Interviews	N/A	1500	1500	1500
Expected Wave 3 Completed interviews	N/A	N/A	1200	1200
Total Sample	3000	3500	4700	4700

## Spanish Language Translation

In 2006 the GSS added Spanish to its standard, English-language version. This translation allowed the GSS to expand its target population to adults living in U.S. households able to be interviewed in either English or Spanish. The addition of the Spanish language interviews “notably increased the number and proportion of Hispanics in the GSS.” In addition, the composition of the Hispanic population changed in several notable ways. The adding of Spanish-language interviews shows that Hispanics are notably less assimilated than indicated in the previous English-only samples and also differ on several other demographics. This variation across demographics is often, but not always, linked to the differences in level of assimilation across the language-use/ability groups. The analysis of non-demographics further indicates that Hispanics often significantly differ across language-use/ability groups.

Further analysis of the differences across language-use/ability groups focusing on the English and Spanish bilinguals identifies a few items on which language effects may be occurring. These will be explored further by building Spanish-wording experiments into the 2008 GSS.

## **Cross-National Research**

The GSS first spurred cross-national research by inspiring other nations to develop data collection programs modeled on the GSS, including Allgemeinen Bevölkerungsumfragen der Sozialwissenschaften (ALLBUS) in Germany, the British Social Attitudes Survey, the National Social Science Survey in Australia, the Polish GSS, and the Japanese GSS. Second, it joined with these and other programs to form the International Social Survey Program (ISSP), a collaborative program of comparative survey research. The fundamental goal of ISSP is to study important social processes in a comparative perspective by both examining differences across societies and changes within countries over time. Since 1984, ISSP has grown to 43 nations, which includes the founding four-- US, Germany, Britain, and Australia-- plus Austria, Belgium, Brazil, Bulgaria, Canada, Chile, China, Croatia, Cyprus, the Czech Republic, Denmark, the Dominican Republic, Finland, France, Hungary, Ireland, Israel, Italy, Japan, Latvia, Mexico, the Netherlands, New Zealand, Norway, the Philippines, Poland, Portugal, Russia, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Turkey, Uruguay, and Venezuela. Data from ISSP modules on the role of government, social networks and support systems, social equality, the family, work orientation, religion, the environment, national identity, and citizenship are available from various national archives and the Inter-university Consortium for Political and Social Research (ICPSR) in the United States.

## **Experimentation and Re-Interview, Follow-up and Methodological Studies**

Since its inception, experiments have been included as part of the GSS and the GSS has completed numerous re-interviews and follow-up studies. Experiments are an integral part of the GSS program of methodological research, and dozens of studies have been completed as part of the replicating core and topical modules to examine most aspects of survey methods. Re-interview studies have included both methodological and substantive studies and the GSS has also served as the source for follow-up studies of employers, voluntary associations, religious organizations/leaders, and family mobility. In a continuing effort to improve data, the GSS conducts methodological research on topics such as survey error, sensitive topics, sample-frame comparability, third-person effects, contextual effects, the measurement of race and ethnicity, item non-response, cross-national comparisons, and network measurements.

## **Data Dissemination and Usage**

Since 1972 the Roper Center for Public Research has been the original point of deposit for GSS Data. Upon release, data users may immediately secure data on a CD from the Roper Center. Data, however, are also simultaneously released to ICPSR and the University of California at Berkeley Survey Documentation and Analysis (SDA) Archive where, after additional preparation, data are made publicly available for download and analysis. GSS data are also distributed by over half a dozen archives around the world. In 2007 the GSS received support from NSF to update their website. The funds are being used to create a GSS data dissemination portal which will allow researchers and the general public to browse the GSS 1972-2006

cumulative data file, run basic statistical tabulations on-line within their web browser, and select variables of interest for download into a selected statistical package format for additional analysis.

GSS data are made available to a broad-based user community. Currently, there are over 150 different versions of GSS/ISSP. Documentation comes in five major forms-- the GSS Cumulative codebook, a SPSS system file, *the GSS Report Series*, the GSS Data and Information Retrieval System (GSSDIRS), and, for the ISSP, electronic and hard-copy codebooks in English, data files, and a CD-ROM with codebooks and files, plus copies of original language questionnaires. The GSSDIRS website seems popular with users, having approximately 4,000,000 visits annually. The ISSP website is visited approximately 200,000 times a year. The user community includes researchers, college teachers, university students, business planners, media and public officials. Academic scholars who use the GSS as a data source include sociologists, political scientists, economists, statisticians and survey methodologists, anthropologists, geographers, biologists, engineers, psychologists, criminologists and legal scholars, medical/health researchers, and business administration and management scholars.

The GSS is a well used research tool and there are now 14,000 documented uses, but the detailed categorization has not been updated since 2003. In 2003, the PIs were able to document 8,662 uses of the GSS: 4,862 journal articles, 1,664 books, 1,364 scholarly papers, 568 reports, and 188 dissertations and theses. Most users (82%) were academics with college affiliations. Usage has grown over the years, increasing from 200 per annum in the late 1980s to over 600 per annum in 2003. With the exception of the Census and its Current Population Survey, the GSS is the most frequently used data set in the three leading sociology journals. The GSS has been used about as often as the total of the next six most frequently used data sets combined.

## Recommendations

The workshop participants strongly emphasized that the GSS had made many invaluable contributions to social science, to U.S. policy debates, and to public understanding of the characteristics of U.S. society. GSS time series and cross-sectional data provide the foundation for social science understanding of a wide range of issues, allowing for the tracking of changes in attitudes regarding race and ethnic relations, religious beliefs and practices, and family life. The GSS occupies a unique niche as an effective social science omnibus survey, which it has occupied for about 35 years. Its future strengths lie in being able to continue to operate as a very high quality survey of the behavior and attitudes of a representative sample of the adult U.S. population. The GSS is the “gold standard” for survey research in the U.S. and globally. Thus, the current basic structure of the GSS provides a strong foundation for moving into the future. Workshop participants recommended that the basic structure of the GSS should remain the same. However, they also offered recommendations that focus on how to move forward and enhance the GSS for the future.

## **The Basic GSS- What to Maintain and Continue**

*The GSS should remain a nationally representative survey of attitudes and intergroup relations and continue to monitor trends in attitudes and behaviors.* It should maintain and further enhance its openness to the addition of new topics and questions by the social science community.

*Continue to administer the GSS on a biennial cycle.* Since time is the critical independent variable in administering the GSS, it should maintain the biennial schedule.

*Develop plans to maintain the panel design.* The panel component is a valuable addition to the GSS and allows for the measurement of individual change over time on many variables, particularly changes in attitudes, allowing the direct study of temporality and causation. The panel study also allows GSS users to study issues that could not be studied with the earlier, purely cross-sectional survey design.

*Maintain GSS participation in the ISSP data collection program.* The ISSP is an important vehicle that can be used to study important social processes in a comparative perspective by both examining differences across societies and changes within the current 43 nations over time.

*Continue a timely and comprehensive release of all data.* Data from the GSS, both the core and topical modules, should continue to be made available to all users at the same time and as soon as possible after collection and data cleaning.

## **Governance, Survey Administration and Funding**

*Strengthen GSS governance and the role of the Board of Overseers.* Significantly enhance and strengthen the role of the GSS Board in evaluating the content of the survey, including the core and topical modules.

*Enable the Board of Overseers to participate more fully in the selection and development of GSS topical modules.* Topical modules provide a vehicle for innovation and should remain part of the future GSS, but the Board of Overseers should have a stronger voice in evaluating topical modules for scientific merit.

*Provide support for the GSS Board of Overseers to hold a “module” competition.* The quality of the GSS would be further enhanced if funds were provided to the GSS Board of Overseers to allow them to manage competitions to determine the addition of modules and questions to the GSS. This competition would be open to social scientists who could propose topics and questions to add to the GSS.

*Explore ways to realize cost savings in survey administration.* The GSS should explore alternative modes of administration in order to reduce per-interview costs.

*Seek innovative modes of administering the GSS.* There is considerable value in maintaining the face-to-face CAPI interview process. The future GSS, however, should examine incorporating other modes of administering the survey, including a leave-behind questionnaire and internet surveys by providing internet access and even laptop computers to allow those in the sample to respond to a series of internet surveys over time.

*Study the impact of the split-ballot technique on data analyses.* The split-ballot administration that gives each respondent two-thirds of the core may create significant problems for analysis. The GSS should study the impact of this aspect of survey administration and, in particular, explore ways to eliminate the split-ballot of the core items.

*Share the funding of the GSS across other social science and education fields.* The GSS is a valuable survey data resource for disciplines other than sociology, with less than 50 percent of GSS users identifying themselves as sociologists. The GSS is heavily used as a teaching and research tool at both the undergraduate and graduate levels. NSF is encouraged to seek joint funding of the GSS more broadly across the social sciences at NSF.

## **Data Collection, Dissemination and Outreach**

*Make the GSS core more transparent to the user community and allow for changes in core items.* The content of the core and how it has evolved over time should be made clearer to all potential users of the GSS. The GSS should sketch out the core set of items, but the core should be allowed to evolve over time via the interaction of the GSS Board, GSS PIs, and the user community. Also, the possibility of enlarging the core, including bringing back time series questions that were deleted in the early 1990s, should be explored.

*Explore ways to enhance the data collected.* The GSS should explore the possibility of oversampling minority groups in order to allow comparisons between groups. It should also consider collecting and making available to users much more “paradata/metadata,” including interviewer characteristics, the mode of administering the interview, the number of calls made to obtain the interview, characteristics of the home and neighborhood, respondent reaction times for answering questions, and spatial identifiers. These data would be released subject to full protection of respondent confidentiality, which in some cases will include special protections and restricted access as all major data projects now do.

*Experiment with digital-recording of interviews.* Explore ways to digitally record interviews using the computers that interviewers use to conduct the interview. Such recordings would enable studies of “how” the interview is done; these studies would help improve data quality and encourage the integration of qualitative and quantitative research.

*Provide opportunities for experimentation.* The GSS should encourage the embedding of both methodological and substantive experiments in the survey.

*Develop “targeted” GSS dissemination activities.* The current NSF funding model does not directly address the issues of data access and dissemination. Targeted NSF funding is needed for

the dissemination of data, including providing web access and increasing ease of access to the data for research and teaching purposes.

*Upgrade and create more user-friendly modes of dissemination.* The current modes of dissemination are confusing and badly outdated. The future GSS should include a state-of-the-art dissemination system that is updated frequently, including a user-friendly website, a high quality index, easy to use search tools, thorough documentation about survey procedures, questions, and data files, simple data downloading in a variety of formats, accessible technical assistance, and simple on-line analysis tools for users.

*Expand the type of data products that are publicly available.* In addition to disseminating the complete dataset, the future GSS should explore making other data products available to users, such as datasets of only core items, each topical module, and each minority group. Also, any data linked to GSS data by users should become part of the GSS publicly available data, including the computer code necessary to replicate analyses.

*Secure a separate budget for data dissemination.* Proposals for the future GSS should include an extensive plan for data dissemination with an adequate budget, whether dissemination is to be done in-house or via a third party. Past NSF grants have not included adequate funds for outreach and dissemination.

*Develop cooperative relationships with other major NSF surveys.* Both the principal investigators and the Boards of other NSF-funded surveys, in particular the Panel Study of Income Dynamics and American National Election Studies, should work cooperatively to share information and technologies. They should also consider greater coordination and integration, such as common data access mechanisms and repositories, as well as using common measures for demographic variables.

## Appendix 1: Workshop Agenda

**The General Social Survey: The Next Decade and Beyond**  
**National Science Foundation**  
**Workshop on Planning for the Future of the GSS**  
**Arlington, Virginia, Room 515, Stafford II**  
**May 2-3, 2007**

### AGENDA

**Wednesday, May 2, 2007**

- 8:30-8:35**                    **Welcome and Introductions**  
Patricia White, Program Director Sociology
- 8:35- 9:00 am**                **Opening Remarks**  
Dr. David Lightfoot, Assistant Director, Directorate for Social,  
Behavioral and Economic Sciences  
Dr. Edward Hackett, Director, Division of Social and Economic  
Sciences
- 9:00- 9:30**                    **An Overview of the General Social Survey (GSS). History of Current  
Activities**  
Tom Smith, National Opinion Research Center (NORC)  
Peter Marsden, Harvard University  
Norman Bradburn, NORC and University of Chicago
- 9:30-10:00**                    **GSS Operations and the Board of Overseers - The role of the GSS  
Board of Overseers; governance and accountability; PI(s) and  
organizational expertise and skill sets; procedures for adding modules to  
the GSS.**  
Michael Hout, University of California, Berkeley  
Rob Mare, University of California, Los Angeles  
Barbara Entwisle, University of North Carolina, Chapel Hill
- 10:00-10:10**                    **Break**
- 10:20-11:20**                    **Content of the GSS - Core and emergent areas of the GSS, adequacy of  
the core, areas that need updating, potential new topical areas, and  
conceptual issues.**  
Barbara Entwisle, University of North Carolina, Chapel Hill  
Leslie McCall, Northwestern University  
Lynn Smith-Lovin, Duke University\*

- 11:20- 12:05pm**      **Methodology-** Sampling, interviewing techniques. survey/item/module development  
                                  Suzanne Bianchi, University of Maryland  
                                  Jeremy Freese, University of Wisconsin  
                                  Douglas Maynard, University of Wisconsin
- 12:05 – 1:00 pm**                      **LUNCH**
- 1:00-1:45**                      **Conceptual and Methodological Innovations**—Geographic Information System component, panel design, embedded experiments, Spanish language translation and other potential enhancements.  
                                  Mark Chaves, University of Arizona  
                                  Barbara Entwistle, University of North Carolina, Chapel Hill  
                                  Steven Nock, University of Virginia  
                                  Jon Krosnick, Stanford University  
                                  Lynn Smith-Lovin, Duke University\*
- 1:45- 2:45**                      **The GSS, ANES & PSID - Opportunities for collaboration and integration**  
                                  Jon Krosnick, Stanford University  
                                  Frank Stafford, University of Michigan  
                                  Suzanne Bianchi, University of Maryland
- 2:45-3:15**                      **GSS and International Surveys**  
                                  John Logan, Brown University  
                                  Ronald Inglehart, University of Michigan  
                                  Steven Tuch, George Washington University
- 3:15-3:25**                      **Break**
- 3:25-4:15**                      **Contributions of the GSS to Sociology and other social sciences and its “Broader Impacts”**  
                                  Douglas Maynard, University of Wisconsin  
                                  Gregory Price, Jackson State University  
                                  Andrew Beveridge, CUNY, Queens College  
                                  Steven Nock, University of Virginia
- 4:15-5:00**                      **General Discussion**
- 5:00**                      **Adjournment**

**Thursday, May 3, 2007**

- 8:30 – 9:45 am**      **Dissemination** - Outreach and data dissemination, web interface and data analysis, user communities  
Gregory Price, Jackson State University  
Mark Chaves, University of Arizona  
Peter Granda, Inter-university Consortium for Political and Social Research (ICPSR)  
Steven Ruggles, University of Minnesota\*
- 9:45–10:30**      **The GSS and Technological Innovation and Cyberinfrastructure**  
Jon Krosnick, Stanford University  
Jeremy Freese, University of Wisconsin
- 10:30—10:40**      **Break**
- 10:40-11:00**      **Insights from the ANES Recompetition and other NSF Projects**  
Frank Scioli, Senior Advisor, Division of Social and Economic Sciences
- 11:00- 12:00**      **General Discussion** - Recompetition of the GSS – Directions and Advice
- 12 noon**      **LUNCH**
- 1:00pm**      **General Discussion (continued)**
- 2:30**      **Adjournment**

## **Appendix 2: Participant Papers**

## The General Social Survey and Its Impact in Sociology and Other Social Sciences

Andrew A. Beveridge  
City University of New York, Queen's College

The General Social Survey (GSS) serves as the *Omnibus Survey* for the entire social science (especially the sociological) community. It has a core set of questions repeated every administration and many topical sets of questions added for a given administration and then retired, sometimes to be used again. It uses the highest quality (read most expensive) methods, including personal interviews and full probability samples, to elicit data from a sample of the adult non-institutionalized population. Many, many researchers use the GSS either as the major focus of their work or as a secondary source of data to put their analyses in context. It has spurred the development of comparative surveys in a host of countries. It has been subject to many methodological experiments, including split-half questions and a wide variety of others. Furthermore, it has achieved a position of the preeminent survey for use in sociology classes in methods that occur early in the career of many undergraduates.

The impact of the GSS on sociology and social science has been massive, enduring and irreplaceable, but the fact that it is an omnibus survey, originally developed in the late 1960s and early 1970s also accounts for its significant limitations. Many of which, in my opinion cannot be remedied by any methodological or redesign "fix." At the same time, the core uses of the survey remain very relevant; even some 35 years after the GSS first went into the field. Put another way, some of the most recent proposals for the continuation of the GSS, seems to call for changing the leopard's spots, so that the GSS in certain ways matches some of the animals now in the survey jungle, for instance the National Educational Longitudinal Survey, the Early Childhood Longitudinal Surveys from Birth and from Kindergarten, the Survey of Adolescent Health, the National Longitudinal Survey of Youth, the National Survey of Drug Abuse, the Chicago Neighborhood Survey, etc.

The fundamental strength and weakness of the GSS is that it is a full probability sample of roughly 3,000 United States adults and asks them a wide array of questions on a large number of topics, some of which remain the same and some of which change from administration to administration. This makes it possible to do a number of things that no other survey instrument allows:

- 1) Track changing patterns of behavior over time, e.g. sexual activity, religious affiliation, etc.
- 2) Track changing patterns of societal attitudes over time, e.g., political views on a wide array of subjects, including redistribution, role of women, religion, etc.
- 3) Compare responses from the GSS to responses to similar surveys from other nations through ISSP.

- 4) Relate the core variables to answers to questions on a wide variety of the specific topical modules, added to the survey from administration to administration.

At the same time, the GSS has significant limitations, which means that other topics and other sorts of analyses are more properly investigated by other surveys and other data collection efforts. The fundamental limitation in the GSS is its small sample size of 3,000 and the fact that it is (and was designed to be) a cross-sectional survey of the United States population. This fact means that the GSS is not as good an instrument for the following:

- 1) Assess patterns for sub-groups either geographically or by particular race of ethnic group. When one compares the GSS with the CPS (about 110,000 respondents in 78,000 households) for instance, or with the Census (14 million) or the American Community Survey (about 3 million released for 2005) it becomes plain that the GSS asks more and perhaps better questions than standard Census surveys, but of many fewer respondents. Analyses for specific cities, areas or neighborhoods are not possible using the GSS.
- 2) Track change at the individual level over time. Even with the retrospective and prospective panels proposed for the GSS it is the case that the GSS would not have anywhere the number of respondents that more topical longitudinal surveys have, such as the ECLS-K (roughly 20,000 usable cases for five waves), the NELS-88 (about 25,000 cases), ADDHEALTH or the other large-scale longitudinal surveys. These surveys, and others, have the drawback for generalization to the whole population of being limited to one or a few cohorts and focused on a given topic, e.g. education, health. However, for these specific (and important) uses the GSS is no match.
- 3) Bringing in contextual or spatial variables. The lodestar study here, of course, is the *Project on Human Development in Chicago Neighborhoods*, which has rich contextual data collected by surveys in a sample of neighborhoods. Though the addition of some contextual variables from the US Census might be useful, it is also the case that contextual variables should also include a spatial and other component. Furthermore, the same limitation discussed above applies here: since it is a sample of the entire United States population, the effect of contextual variables in a given location may turn out to be somewhat difficult to grasp. A multi-multi-million dollar study such as the Chicago neighborhood study may be necessary to analyze the effects of context on a variety of outcomes, including the correlates of crime.

Thus, the niche for the GSS is as an effective social science omnibus survey. It is the niche that it has occupied for some 35 years, as the type of research done in social science radically changed. New methods for data collection and analysis were developed, including a wide array of methods to analyze longitudinal and contextual data. Many new survey operations were launched to collect exactly the sort of data revolving around a given topic. It is not surprising that those planning the GSS would want to participate in these newer approaches. Nor is there anything fundamentally wrong with attempting to do so. However, given the heritage of the

GSS it is especially important that it continue to function within its own niche and provide the sorts of data that it is best at providing: a very high quality survey of the behavior and attitudes of a representative sample of the adult United States population.

In a similar vein, over the years the GSS has conducted a series of experiments with the instrument. It is also important that any such experiments be directly related to the core mission of the survey.

I realize the temptation, after 35 years of conducting such a survey, to try to add bells and whistles: e.g. rotating panel, contextual variables, and more methodological experiments. My recommendation would be to judge the continued support for the GSS in terms of its core mission and not to deviate very much from that mission, since in that role it has been invaluable to the social sciences, both in terms of research and education, and also in terms of dissemination of social science to the wider world.

In my own work with the *New York Times*, we have used the GSS on numerous occasions, including the following:

1. To assess changing views of inequality
2. To look at the social status and attitudes of various religious adherents, especially evangelicals
3. To see if attitudes of those living in small towns are different from those living elsewhere
4. To understand who defines themselves as middle class
5. To look at changing views of women's role, as well as a number of other uses

The GSS is without peer in allowing one to assess attitude and behavior changes and relate them to demographic or other characteristics. It staked out this role in 1972 and this should be its role in the future.

## The GSS, the ANES and the PSID & GSS Methodology Comments Prepared for the NSF Workshop on Planning the Future of the GSS

Suzanne Bianchi  
University of Maryland

### The GSS, the ANES and the PSID

If the only thing that unites ANES, GSS, and PSID is that major funding comes from NSF, this is not a strong basis for interaction. However, if there are common problems, some mechanism for cooperation and exchange may be desirable. Currently, the GSS Board includes Jon Krosnick, an ANES PI, and Bob Schoeni, a PSID PI. This facilitates communication among the three surveys but is not a long term arrangement.

There are a number of issues where sharing information across surveys can be beneficial to the GSS.

1) **Panel Design and Recontact.** As the GSS undertakes a panel, it is useful to have regular contact with other surveys that do panel data collection like the PSID. IRB issues about recontact of panel members, for example, was an issue that became apparent largely because Bob Schoeni is currently on the GSS Board and had encountered similar issues in the discussions surrounding recompetition of the PSID.

2) **Website Enhancements.** Another area where communication across the surveys is useful and has benefited the GSS is in areas such as website design. For example, at one of the GSS Board meetings, Jon Krosnick demonstrated features of the ANES site to GSS Board members.

3) **Contextual Information Appended to the GSS.** PSID had developed rules and procedures for use of restricted access files, such as those with contextual information appended, that could be modified by GSS.

4) **GSS Expertise in Attitudinal Measurement.** Benefits of contact can also flow from the GSS to other surveys. For example, surveys like PSID, which in the past has not had a great deal of attitudinal information especially on non-economic topics, can benefit from the long history of item development in the GSS.

I serve on both the GSS and PSID Boards and there are differences that seem important to highlight.

The PSID, like the GSS, faces the challenge that NSF does not provide sufficient funding for the conduct of the survey. The PSID's funding diversification strategy has been quite different from the GSS, however. The PSID has managed to get sizable funding from NIHD and from NIA. The NICHD funding was largely secured by adding on a new, in-depth assessment of PSID

children (The Child Development Supplement to the PSID). With this strategy, the PSID PIs have maintained control over the coherence and intellectual content of the whole survey.

An intergenerational panel survey like the PSID may be better able to attract NIH dollars but GSS might consider whether something more akin to the PSID strategy might enhance the scientific value of the GSS. The GSS panel may eventually make this more likely, though the panel may have to extend for more than four years to attract funding from sources such as NIH.

The PSID faces pressures to diversify the PIs involved – to be more inclusive of researchers beyond the PSID PIs. In many ways, the GSS has accomplished much greater PI diversification through its selling of topical module time. There is a balance that must be struck between control of content and openness and the potential innovation that might come from greater inclusion of the broader research community.

Another difference between the PSID and the GSS is in staffing. The PSID seems to have more staff dedicated to the survey than GSS. The PSID also has rotated in a new co-PI in recent years with Bob Schoeni. It retains continuity with the continuation of Frank Stafford as PI but also seems to have given attention to succession. This is an important issue for the continuance of a high quality GSS as well.

## **GSS Methodology**

From 1972 through 2006, the GSS has monitored trends in U.S. attitudes, behaviors and attributes and provided high quality data on social change in the U.S. A key issue is therefore how to continue the time series replication of core items for monitoring change while also being responsive to new developments in the survey research field and attentive to costs and benefits of engaging in the “status quo” versus innovations in data collection methodologies. Following is a list of some of the issues that this workshop might address:

1. **In-person Interviews.** The use of in-person interviewing in the GSS is a unique feature of the survey, as many other surveys have moved to telephone interviewing. Use of in-person interviews with respondents no doubt keeps GSS response rates high relative to surveys that use other modes. In-person interviews also may enhance the goal of replication of the time series, as changes in mode may introduce noncomparability in the trends in the core items on the GSS. However, in-person interviews are costly. The case for strict adherence to in-person interviewing techniques is especially strong if it can be demonstrated that comparability would be compromised, data quality would decrease, and response rates would decline with any other mode of data collection.

In this workshop, as in the workshop conducted before the recompetition of the ANES, it would seem important to discuss mode of interview. The notice of recompetition must include specifications to potential applicants about this important issue. An outcome of the ANES workshop was the recommendation to NSF to address the issue of “The value of maintaining primarily face-to-face interviewing of the core component.” Is the use of in-person interviewing

absolutely required to ensure the integrity of the GSS? How much or how little mode experimentation should NSF encourage in the recompetition notice?

2. **GSS Panel.** In 2006, the GSS inaugurated a prospective panel. Cases from the 2006 GSS will be targeted for reinterview and reinterviewed cases in 2008 will be slated for reinterview a second time in 2010, creating a 3-wave panel covering a four-year time period. The GSS has never had a panel before and it will no doubt take a few years to evaluate how worthwhile the GSS panel is and whether the current design is optimal. Hence, a design issue for the recompetition is whether or not the recompetition notice will specify that anyone submitting a proposal must maintain the panel design that was implemented with the 2006-2008 survey. Or, will the recompetition allow latitude to rethink the introduction of the panel into the GSS? A further complication that the panel design adds, as became clear at the Spring 2007 GSS Board Meeting, is that 2006 GSS respondents were told they might be recontacted by NORC. Any other survey firm does not have permission for recontact.

3. **Assessment of Nonresponse and Data Quality in the GSS.** The GSS Board has been especially interested in having information recorded by interviewers or about the conduct of the interview, including record of calls, appended to the GSS data set in order to permit methodological work on data quality and nonresponse bias. A more complete set of variables about the actual contact with the household will be available in 2006 than has previously been the case. Record of call will not be appended because, unless interviewer training is enhanced, the recording of contact attempts is not currently systematic enough across interviewers to be coded and usefully analyzed. One issue that workshops participants may wish to discuss is the amount and type of information about the interview that should be routinely included on the GSS data file.

4. **Maintaining Innovations.** There have been a number of innovations in recent GSS survey rounds that need to be continued. For example, the GSS innovated in sub-sampling nonrespondents for follow-up in the 2004 survey. This seems to have worked well, represents an innovation that comes from access to leading sampling statisticians and survey operations at NORC, and represents an important dimension on which to assess those who enter the recompetition. Another innovation was the introduction of a Spanish language version of the GSS for the first time in 2006. This no doubt enhances the representativeness of the GSS sample. If enhancements such as this are considered critical to maintaining the quality of the data collection, these would need to be requirements for future GSS data collections.

5. **ISSP Modules.** The GSS has included ISSP modules and the GSS Board and current PIs have been active members of the ISSP community. To the extent that this involvement enhances the scientific usefulness of the GSS, this component of the GSS program would also be an important feature of the GSS recompetition notice.

6. **Topical Modules.** Currently, NSF funding does not cover the full cost of conducting the GSS. Additional resources are raised by selling time on the survey in the form of topical modules. On the plus side, this allows a broad range of researchers to have access to a high quality survey and it encourages proposal development to fund content on the GSS. On the negative side, this means that scientific coherence and content do not always drive the decisions

about the questions that are included on the GSS; the content of topical modules is a compromise between available funding for content and scientific merit of the content. Also, considerable PI time is spent fund raising. The GSS Board provides feedback on items and modules, with ultimate veto power over poor quality modules. However, there are limits to this oversight, given the need to fund the survey. Also, often the funding of modules comes in so late that there are data quality and methodological implications. For example, there is often inadequate time to do cognitive testing of new survey content before the pretest.

**7. Methodological and Substantive Experiments in the GSS.** Some GSS Board members have expressed concerns that GSS engaged in more methodological innovation in earlier years than it does currently. In addition, in the workshop before the ANES recompetition, it was pointed out that more methodological experiments (e.g., question wording experiments) than substantive experiments are done in the ANES. This is probably also true of the GSS. If it is the case that methodological experimentation is declining and there is little experimental manipulation that addresses substantive issues, this may be especially problematic at this juncture. There is growing interest outside the field of sociology (e.g., in behavioral economics) in experimentation and in the feasibility of embedding experiments in surveys like the GSS, with cross-sections of the population, rather than with college students in laboratory settings. Should innovations along these lines be encouraged in the resubmission notice?

**8. Data Accessibility.** The ease with which data can be accessed is extremely important. Currently, work to enhance the GSS website is underway. Continued improvement in data access and in the GSS web presence must be part of the mandate for any future GSS proposal.

## Thoughts on the General Social Survey

Norman M. Bradburn  
National Opinion Research Center

The General Social Survey is the most important and cost effective infrastructure program funded by SBE. It is important to the entire social science community for four principal reasons:

1) The GSS is the means of measurement for “perishable” social phenomena such as attitudes and values that are basic data for the study of society and social change. Much as telescopes provide the means of measuring celestial phenomena at specific dates that form the basic data for astronomy, the GSS is means by which we measure social phenomena at specific times. The measurements must take place at regular intervals, if change is to be observed, because once the time period is past, the phenomena cannot be observed. They are in this sense “perishable.” Without an instrument like the GSS, it would be impossible to study scientifically many of the most important aspects of social change.

The goal of providing these basic data imposes some requirements on those responsible for its well-being. Funding must be large enough to insure that the survey is conducted at the highest level of scientific standards. Funding must also be sufficiently long term to insure continuity of leadership, organization, execution and dissemination of the data. Governance must effectively represent the best scientific thinking and the broad user community without regard to disciplinary boundaries. Methodological work must be continually conducted to adapt to changes in language and social conditions and to take advantage of newer technologies while maintaining comparability of measurements.

2) The GSS provides a “gold standard” against which other data collections can be evaluated. It functions as a standard in two ways; 1) it provides national data against which surveys on special populations can be compared when the same questions are used; and 2) it provides a methodological standard which allows surveys that have to make compromises in the quality of data collection to make estimates of possible biases resulting from these compromises.

3) The stability and cumulative nature of the GSS design allows for the study of small groups in the population that would not otherwise be possible at any reasonable cost. These features have resulted in a large number of analyses that go beyond the principal goal of measuring social change and have enabled a vast literature on social processes that greatly enhance the cost efficiency of social research. (See for example the work of Erzo Luttmer, a political economist, on attitudes toward social welfare as a function of the proportion of local recipients from their own racial group recently reported in the N. Y. Times).

4) The GSS has served as a model for the development of similar survey programs in other countries, thus leading to the formation of an international consortium that enlarges the value of the GSS as an instrument of social research. It is now an integral part of the International Social Science Program (ISSP) and could become an even more important in the

development of international comparative social science if it is expanded to a North American GSS and more closely allied with the European Social Survey.

# The General Social Survey: Innovation and Dissemination

## Comments for NSF Workshop on the GSS

Mark Chaves  
University of Arizona

### Innovation in the GSS

The GSS has a spectacular and still growing record of innovation. Maintaining the capacity to innovate conceptually and methodologically should be a high priority for the next decade and beyond.

There are two senses of “innovation.” Innovation can mean introducing something completely new to the world, or it can mean changing previous practice in ways that, while perhaps not inventing something completely new, refresh an enterprise and keep it on the cutting edge. The GSS’s record of innovation is impressive in both of these senses.

The GSS’s record of innovation in the first sense includes inventing the very concept of an ongoing *general* social survey, helping to inspire and create the ISSP, advancing measurement validity and reliability through question-wording and context experiments, producing new substantive knowledge on too many subjects to name, developing time series that are available nowhere else, and facilitating the generation of new, high-quality samples at other levels of analysis (e.g. employing organizations, religious congregations, voluntary associations).

The GSS has from the beginning been innovative in the second sense as well. Most recently, significant innovations to the GSS include adding Spanish-language interviews (in 2006), introducing a panel supplement to the cross-sectional time series (the first re-interviews of panel members to occur in 2008), and implementing a new two-stage sampling design that more efficiently uses available resources to achieve a high weighted response rate (in 2004). The GSS is not the first national survey to incorporate these elements, but their recent incorporation illustrates well the GSS’s proven and ongoing capacity to adapt in ways that keep it at the forefront of survey research on American society.

The highest priority innovations for the next decade of the GSS are, I think, two initiatives described by the PIs in the background documents for this workshop:

1. Increase the capacity for cross-national research using the GSS, including augmenting the ISSP and creating the North American Social Survey. Cross-national research using surveys of individuals within countries is an increasingly fruitful strategy for advancing knowledge on many fronts. The GSS should continue to be a leader in international efforts to further enhance our capacity to conduct cross-national research of this sort.
2. Add geographically-based contextual data. Enhancing our ability to locate GSS respondents geographically and socially will advance knowledge of the correlates and causes of the many attitudes and behaviors measured by the GSS.

To these top-priority innovations, I might add four others to consider in the coming years:

3. Gather data from other members of GSS respondents' households to further deepen our knowledge of the social contexts in which respondents live.
4. Collect biomarkers from GSS respondents. The GSS could be a leader in the increasingly important effort to integrate biological, behavioral, and social/cultural levels of analysis.
5. Continue to improve measurement on the GSS by implementing cognitive pretesting of all new items.
6. Pending the outcome of initial tests, implement the multi-level, integrated, database approach described in our background materials.

The core mission of the GSS is to gather high-quality data on the attributes, attitudes, and behaviors of people in the United States in order to document and explain trends and constants in those attributes, attitudes, and behaviors. Candidate innovations should be prioritized by their potential to significantly advance that core mission. In my view, enhancing our ability to integrate different levels of analysis—whether by placing GSS respondents in their geographic and social contexts, placing U.S. data in a cross-national context, or examining biomarkers along with attitudes and behaviors—seems the most fruitful way to advance that core mission given the current state of sociological knowledge, methodological sophistication, and technical possibility. This is why my top four priorities for future innovation all involve efforts to enhance our ability to integrate different levels of analysis.

The recompetition notice should make clear that future GSS leaders will be expected to push the GSS forward in these important ways. More generally, and whatever specific innovations are pursued, future leaders should possess a demonstrable capacity for building on the GSS's established legacy of innovation and leadership in survey research on American society.

### **Dissemination of the GSS**

GSS data and results have been disseminated remarkably widely. The background materials document a staggering level of productive use of GSS data by researchers, teachers, students, journalists, and government officials: it is the third most highly used dataset (after the Census and the CPS), with almost 9,000 published uses, 188,000 datasets downloaded, 19 million visits to GSS/ISSP web sites between 1999 and 2003, 90 data extracts distributed with textbooks, and 250,000 students annually enrolled in courses that use the GSS. It is clear that GSS dissemination efforts have been wildly successful and, faced with numbers like these, it is tempting to say that the recompetition notice should simply state that GSS leaders should be prepared to continue current dissemination efforts without any change in current practice.

Still, I think it is fair to point out that internet access to GSS data and documentation is not state-of-the-art. GSSDIRS may have been state-of-the-art when it was developed in 1999-2000, but resource constraints have prevented updates of that system, and the GSS consequently has fallen behind the curve in internet dissemination. For example, there is no one definitive, up-to-date

GSS web site where a user can find all GSS data and all relevant documentation. New GSS data sets are first distributed through the Roper Center, from whom users must purchase them. The data sets are deposited at ICPSR and made available through other data archives within a few months, but this arrangement means that there is a several month period during which new data are available only via purchase from the Roper Center. To mention another example, the “GSS Methodological Reports” button on the GSSDIRS site produces a list of papers, but the full text of most of these papers is not available via the site. I think many users of GSS data would now expect to be able to read all GSS Methodological Reports at the click of a mouse.

The need to bring the GSS up to speed in internet dissemination—and keep it there—is made more urgent by the increasing complexity of the GSS. The subsampling design implemented in 2004, the panel data that will be introduced in 2008, and the potential addition of contextual geographical data all make it even more important that users have the easiest possible access to technical documentation as well as to the most recent data.

I understand that GSS staff already is working to develop a new, fully current project web site to be launched some time this year. That’s great. Still, as the history of GSSDIRS illustrates, *keeping* a web site current in both function and content is as much of a challenge—perhaps more of a challenge—than creating it in the first place. Perhaps, then, the point here is simply that the GSS recompetition notice should make clear that, in the coming decade and beyond, NSF expects the GSS to *stay* at the forefront of internet dissemination of data and documentation, and it will provide the resources to make that possible.

## **Hard Choices: Reflections on the Design of the General Social Survey**

Barbara Entwisle  
University of North Carolina, Chapel Hill

The GSS is a biennial personal interview survey of American adults consisting of a mix of sociodemographic background items, a rotating set of core attitudinal items crucial to the measurement of social trends, International Social Survey Program modules, and other large and small modules proposed by outside scholars with outside funding. It is a highly worthy project, and the current model is one of success. The current model is also the product of a series of compromises made over the history of the project. I think that it may be helpful at this point to consider aspects of the design from this perspective and to ask whether there might be other approaches with equal or greater appeal.

The GSS started out in 1973 as an annual survey. Other than 1979 and 1981, it continued as such until 1994, when it shifted to a biennial survey, but with a twist. Rather than a separate survey fielded annually, two surveys were packaged together and fielded at the same time in a single year. This design change resulted in substantial savings (training costs, listing expenses, etc.), and the GSS has continued as a biennial survey ever since. Should it? This question needs to be considered in relation to the speed of the changes the survey is intended to capture, some of which are quick and others not. It might be time to consider the survey in terms of its components and fielding them according to different strategies.

It is my understanding that the continuation of the biennial strategy is largely driven by ISSP obligations. The ISSP designs a topical module for use each year, which must be fielded that year or the next. In the current design, one of the ISSP modules is assigned to one of the two GSS samples, the other ISSP module to the other GSS sample, thus meeting the ISSP requirement. I am highly committed to the goals of the ISSP. I think it is essential that the GSS continue to participate in this international program. However, if the ISSP requirements could be met in some other way, if the components of the current GSS could be packaged somewhat differently, then it would be possible to consider other schedules for the administration of the non-ISSP portions of the GSS.

The GSS is a personal interview survey. This is one of the features of GSS design that gets the most scrutiny, because the costs of personal interviews have risen dramatically over the years. What is the rationale? Personal interview surveys are the “gold standard.” Coverage is better than for telephone surveys because households without telephones are included. Further, personal interview surveys generally have response rates that are higher than telephone surveys. Indeed, the GSS has achieved response levels that are the envy of the field, historically as high as 80% and recently at the 70% level. Importantly, the number, difficulty and complexity of topics that can be covered in a personal interview survey is greater than in a telephone survey. I remain convinced that a personal interview survey has a lot to offer the GSS project. I do, however, wonder whether this approach is needed for all of the components. For instance, the ISSP

modules might be fielded in some other way. Participating countries already use a variety of approaches so changing the mode for the ISSP modules is no threat to the international program. Opportunities might be taken to innovate in the administration of these modules. If the ISSP modules were removed from the personal interview, and fielded in some other way, it would no longer be necessary to field the non-ISSP portion of the GSS every two years. Note that I am not suggesting that ISSP be removed from the GSS—just that there might be more flexibility with respect to the various elements of the GSS.

The GSS aims to survey a representative sample of American adults. Until very recently, this was operationalized as English-speaking adults 18+ living in households in the US. In 1972, when the first survey was fielded, there was a good correspondence between English-speaking household population and the US population as a whole. However, this correspondence weakened over time as waves of immigrants entered the country (consequent to major legislative changes enacted in 1965). The omission of Spanish speakers was a particular problem. Funds for including them have been made available starting with 2006. I am very pleased by this change and feel strongly that Spanish translation should continue to be supported in the future. A major strength of the GSS is its focus on trends and changes at the population level. If we want to understand how American society now is similar to or different from American society ten years ago, twenty years ago, or thirty years ago, we need to include all of the relevant groups.

With respect to content, there is a major tension between continuing to ask questions (including response categories) the same way and making changes to question design to reflect advances in survey methodology. The GSS has emphasized the former for the purposes of replication and to guarantee comparability over time. As the GSS PIs have said on numerous occasions, “If you want to measure change, don’t change the measures.” My views on this depend on the component of the data collection. Specifically, I think that sociodemographic measures should be updated and response formats revised as needed. Why? It seems to me that the value of these variables for the GSS is (1) for identifying subgroups for trend analysis and (2) as background or predictor variables in analyses of core items or variables in the modules. Given these purposes, there is less need for questions and response categories to remain unchanged. If my goal were to track trends in the sociodemographic composition of the population, I would turn to some source other than the GSS. I am more sympathetic to the need to keep things the same when it comes to the core attitudinal items than for the sociodemographic background variables. Although there continues to be a need to review and update core attitudinal items, I think that the argument for leaning in a conservative direction is much stronger for these items.

With respect to the core, I would like to revisit a design change that was made some years ago, i.e., the introduction of the rotating core. In the rotating core, items are divided across three ballots—A, B, C. Each respondent is given two of the three ballots: AB, BC, or AC. Across two samples of roughly 1500 each, this yields responses for 1000 respondents on each core item, a sufficient number for tracking trends. However, this makes the data much more difficult to use for cross-tabulations and regression analysis. Prospective users look at the codebook and see that 1000 cases are available for analysis for a set of variables of interest to them. What they do not realize until trying to do their analysis is that the actual number of cases available may be 1000, but may also be 0 (listwise deletion with one item each coming from A, B, and C). I have seen even fairly sophisticated users flummoxed by this. This is a negative from the standpoint of

one of the important contributions of the GSS: the education of sociology students.

The modules comprise the final component of the GSS. The modules make it possible for the GSS to be topical, timely, theory-driven, and innovative in its content and methodology while at the same time achieving its other goals. As such, they are a vital component of the GSS and should be continued. How these modules are funded deserves some thought, though. NSF originally funded their development and implementation, but this has not been true for a while. In fact, externally funded modules have become a critical source of the funds needed to field each wave of the GSS. This is a burden for the PIs and adds a lot of uncertainty to the planning process. It also puts the Board of Overseers into an uncomfortable position. (What is their role with respect to fundraising?) I would like to see at least some of the topical modules centrally funded. Perhaps there could be an open competition for space in these modules sponsored by NSF, with proposals evaluated by the Board of Overseers.

In sum, the GSS can be seen as the historical outcome of decisions and compromises made over many years and for many reasons. I think it is time to revisit some of these. How important is it to field the survey biennially? My understanding is that participation in the ISSP is a major factor driving periodicity. Would it be possible to consider an alternative approach to fielding the ISSP? What is the most appropriate periodicity for the core items? If the GSS continues with a personal interview approach for its core content, but at less frequent intervals, would it be possible to ask all of the respondents all of the core items? Would the funds stretch to cover the development of specialized modules? What might be done to update the sociodemographic items (including response categories)? The current model is clearly one of success, but thinking about the future, other approaches should also be considered.

<b>GSS Structure</b>						
<i>Sample X (n=1500)</i>			<i>Sample Y (n=1500)</i>			<i>Time</i>
Sociodemographic Core						45 minutes
Rotating Core (each respondent gets 2 of the following 3 ballots):						
Ballot xA	Ballot xB	Ballot xC	Ballot yA	Ballot yB	Ballot yC	
ISSP- Year 1			ISSP- Year 2			15 minutes
Module 1			Module 2			15 minutes
Mini-Modules 1			Mini- Modules 2			15 minutes

(adapted from a graphic distributed by Mike Hout in spring 1999).

# Commentary for GSS Workshop Methodology & Technological Innovation and Cyberinfrastructure

Jeremy Freese  
Harvard University  
and  
University of Wisconsin-Madison

## Methodology

The General Social Survey is an unquestioned national resource and has become this as a result of the excellent stewardship and continual initiative shown by its principals. As someone who cares about social research methods, my principal complaint about the General Social Survey is perhaps that in some ways it is a victim of its own success, at least within sociology. By which I mean that its very prominence can invite problems of use that result in a number of projects that use the GSS being regrettable from a scientific standpoint in various ways. While I recognize that large-scale data collection efforts are often considered having as their purview being simply getting the highest quality data they can, and pathologies of use are not their problem, I do wonder whether more can be built into the structure of the GSS that would improve the quality of work being done from an analysis side as well as from the data collection side. Three points:

1. Two closely connected problems that I see with the practical use of the GSS is that its familiarity to sociologists leads it to be pressed into service for questions for which other data sources exist that would be better to use and that, despite the mostly descriptive emphasis on documenting trends that was emphasized in the materials we were provided, sociologists sometimes seek to use the GSS to speak to causal questions in ways that cross-sectional surveys are not well-designed to speak to. In this respect, I find special reason to be very excited about three of the developments described in the materials: (1) the continued engagement with ISSP, which will allow theoretical discussions of large-scale causes of social trends to benefit from drawing implications about the timing of changes in different nations (although this is limited by whatever ways the ISSP has unique rather than shared content across years, and so more repetition in the ISSP would be desirable); (2) the development of a panel design to the study, although I would like to know more details about this (which will be desirable to allow inferences about intervening events -- for example, changes in marital status -- especially given the wise decision to attempt to gather three points in time on respondents); (3) the possible expansion to include more administrative record matches (which can have the advantage of extending information about respondents across time). I commend these efforts. In this respect, anything NSF could do to specifically promote use of these components, whether in funding specific projects or in funding infrastructure that would make this data as easy to use as the regular GSS cross-sections are now, could be valuable for improving the overall quality of work done using GSS.

2. Especially given increasing emphasis in the behavioral sciences on interdisciplinary work and on being able to describe "mechanisms" and "pathways," the lack of much good measurement on

psychological constructs in the GSS seems like it might post a threat to its value to interdisciplinary contribution going forward. Such measurement is more commonly conducted by SAQ rather than in-person interviews. The Health and Retirement Study, for instance, has invested now in collecting psychosocial measures using a leave-behind questionnaire. The lack of any kind of leave-behind SAQ with GSS has always been somewhat puzzling to me, and seems especially the case now that respondents will be impaneled. For that matter, even an Internet-administered SAQ for those respondents who have Internet access (even recognizing the limitations of such samples then being limited to only respondents who have Internet access), would provide more strength for getting at psychological moderators and mediators of causal processes than what can presently be done with GSS.

3. In economics, there has been increasing prominence of methods of causal analysis that are based less on structural modeling than on identifying 'natural experiments' within the data, especially those that derive from identifying discontinuities in causes and from identifying sources of exogenous variation (instrumental variables). A less obvious implication of this movement is perhaps the interest of the GSS in making available as much so-called paradata (if I am correct in understanding this term to refer to data on the details of fielding and other characteristics of cases), administrative data matches, and geography as possible, even if its analytic implications are not readily apparent, as such data may have uses for being able to support kinds of inferences along these lines that those running the survey might not themselves anticipate.

### **Technological Innovation and Cyberinfrastructure**

My points are about cyberinfrastructure for users of the General Social Survey. As a broader cyberinfrastructural matter, the General Social Survey's web presence is notoriously confusing; in my own effort this weekend to search for materials, there appear to be two different GSS pages at NORC and two different kinds of GSS pages at ICPSR, which link to different materials. Given what a valuable national resource it is, I am unclear why GSS's web presence seems so much less well-developed than either NES or PSID and worry especially that this may impose a hindrance to investigators outside sociology.

1. *Regarding data availability, when supplemental data collection is done.* The "Generation of Data" material lists some GSS Reinterviews and Auxiliary Data Collection projects. Some of these I was already aware of; some of which I was not. I searched for information online about obtaining these data and came away frustrated and empty-handed. Are these data systematically available to the research community? If not, why not? Can anything be done to make information about the existence of these data more widely known and the data themselves more readily accessible?

2. *Regarding analysis of confidential data.* The advancement of the GSS as a scientific tool I think will require it to be collecting and making more data available to researchers than what disclosure requirements can readily allow to circulate publicly. I would be interested in knowing more about any strategies GSS has regarding the release of data that cannot be made available publicly; indeed, I'm only vaguely and anecdotally aware of what policies are now and was not able to locate anything on the Web describing it. (In this respect, I have \*far\* greater enthusiasm

for solutions that make data available on a secure basis to be analyzed using any statistical software than solutions that have attempted to develop specialized statistical packages for online use, especially when the latter do not allow analysis to be conducted in any way that provides a record of steps that is subsequently replicable.)

3. *Regarding data agreements.* As the materials we were provided emphasize, one of the great strengths of the GSS is that it is open and public. I have on multiple occasions now checked findings of a paper using GSS either as a reader or reviewer of the GSS by re-analyzing GSS data myself. One of the great advantages of the Internet is that it allows a considerable increase in the transparency it allows for research. Economics has taken advantage of this transparency by having now enforced rules in its most prominent journals that code and data for analyses are to be made available to the maximum extent possible at the time of publication. Sociology, despite once been at the forefront of initiatives for sharing materials, has lagged behind. A costless way for the General Social Survey to show leadership in this area would be to ask its researchers to show the same kind of attitude toward the desirability of openness and public character of work that the GSS itself exemplifies. Namely, as part of its agreement for users downloading the data, the GSS can inform users of the expectation that users will deposit software code sufficient to replicate results from GSS analyses in the ICPSR Publications-Related Archive, or another public archive that accepts code, at the time of the publication of any articles that use the data. In doing so, the GSS can expand its teaching mission as the availability of code for inspection allows new researchers to see first-hand the procedures by which published work was done, and in so doing gain lessons for their own practice.

## **Best Practices in the Dissemination of Survey Data**

**Peter Granda  
Inter-university Consortium  
for Political and Social Research (ICPSR)  
University of Michigan**

### **Preservation**

Preservation is an important part of the data life cycle, allowing for long-term access to valuable digital materials. Digital materials are best protected by having multiple copies stored at off-site locations. An ideal preservation storage situation includes a minimum of six off-site copies of digital materials undergoing regularly scheduled back-ups. In addition to this redundancy, the media on which the digital materials are stored require ongoing refreshment. Data and documentation formats should also be software independent. An organization with an effective preservation strategy makes an explicit commitment to preserving digital information by:

- Complying with the Open Archival Information System (OAIS) and other digital preservation standards and practice
- Ensuring that digital content can be provided to users and exchanged with archives so that it remains readable, meaningful, and understandable
- Participating in the development and promulgation of digital preservation community standards, practice, and research-based solutions
- Developing a scalable, reliable, sustainable, and auditable digital preservation repository
- Managing the hardware, software, and storage media components of the digital preservation function in accordance with environmental standards, quality control specifications, and security requirements

### **Disclosure Analysis**

Any plan to disseminate survey data must include very specific procedures for understanding and minimizing the risk of breaching the promise of confidentiality that is made to respondents at the time of the survey. Appropriate disclosure risk analysis involves both practical and statistical steps that attempt to identify cases and variables that might be recognizable to an intruder, or matched with external databases. Once those cases and variables are identified, the survey can be evaluated. In virtually every case, the data can be masked in various ways that make it possible for public use data to be distributed, usually through a Web-based system. Sometimes these masking procedures reduce the usefulness of the data for analysis, in which case it is appropriate to create less-thoroughly masked versions that can be distributed under restricted use contracts or made available within a research data center or “enclave.” The key goal of disclosure risk analysis and processing is to ensure that the data have the greatest potential usefulness while simultaneously offering the strongest possible protection to the confidentiality of the individual respondents.

## **Documentation Processing**

High quality metadata is essential to effective data use, and adopting the Data Documentation Initiative (DDI) XML standard for metadata offers several advantages. First, all information that the analyst needs is available in a core document, from which other products, such as setup files, can be produced. Second, the XML file can be viewed with Web browsers and lends itself to Web display and navigation. Third, because the content of each field of the documentation is tagged, the documentation can serve as the foundation for extract and analysis programs, search engines, and other intelligent agents. Finally, preparing documentation in DDI format at the outset of a project means that the documentation will also be suitable for archival deposit and preservation. DDI XML should ideally be generated by the CAI system used to collect data.

With rich DDI markup, instrument documentation can be presented so that users can track the logic of the questionnaire. Also enabled is a bank of all questions ever asked in multi-year studies, years they were asked, differences in question wording, etc. This approach permits linking to the documentation of related surveys -- for example, those conducted in other countries -- with variable text viewable in the native languages, so that analysts can study relationships among all of the survey items.

## **Data Processing**

An effective data processing strategy focuses on the production of data files which will provide optimal utility for researchers. Processors must perform a series of steps to ensure the integrity of public-use files. Such steps include: a thorough investigation of any wildcode or inconsistent responses, the standardization of all missing data values, reformatting any variables to maximize storage capacity, and the creation of complete and concise variable and value labels which will provide researchers with clear descriptions of their analytic results. The format of the data files should permit access through a wide variety of statistical packages all of which will produce the same results no matter how complicated the analysis requested.

## **New Data Products**

Data producers and archives should consider producing ancillary files for those data collection efforts which cover multiple waves of respondents or several geographic areas. Special subsets of data which take advantage of the longitudinal richness of long-term collections provide unique opportunities to study important social, political, and economic issues from different perspectives particularly with regard to the changing characteristics of the sampled respondents.

## **Finding Aids**

Finding aids are critical to a Web-based system. A robust search engine is needed to query the fielded metadata so that the user can find variables of interest efficiently. The search should also run against a study's bibliography so that there is two-way linking enabled between variables and publications based on analyses of those variables. Displaying full text of the publications whenever possible is also essential to realize the full potential of the online research

environment. Dedicated staff should be continuously searching journals and online databases to discover new citations.

### **Types of Dissemination**

The data producer must make every effort to make all public and restricted data and documentation files available to the research community through secure and predictable channels. The producer may decide to provide their own access but should also send copies to a trusted digital repository for permanent preservation should the producer decide to cease such services in the future.

Providing optimal utility for researchers means that data producers and archives produce a variety of products for their varied constituencies. To address the needs of those who seek to do intensive statistical analyses with particular software packages, processors should produce setup files and ready-to-use ‘portable’ files in SAS, SPSS, and Stata. To address the needs of policymakers and those who are browsing for new data sources, seeking summary analytic information, or may want to download specific variables quickly, producers and archives can create tools within the Web-based system to permit online analysis, subsetting, and access to full documentation.

### **Training and Outreach**

It is very important that major survey research products reach out to the user community effectively in order to ensure that they receive the greatest possible use. The most straightforward way to reach out is to develop an effective on-line presence and to ensure that the data are easily located and acquired, and that metadata and bibliographical citations are also available. Beyond that, effective outreach usually includes three activities. First, data producers often organize workshops or conferences soon after the data are released to bring early users together to discuss important preliminary results and to ensure both that the data are used effectively and that any problems with the data are recognized and corrected. Second, data producers often hold training workshops to ensure that novice users have a chance to learn about the data from experts and especially from the data production team itself. Longitudinal data and repeated cross-sectional data are particularly challenging to analyze without specialized instruction and training. These training courses can be brief half-day or one-day sessions at the time of professional meetings, or they can be three- or five-day sessions in the summer (or during the academic year) with a more detailed focus. Finally, data producers send representatives to important professional meetings with a display “booth” where staff from the project can describe the data, distribute documentation and sample data, and encourage researchers to make use of the data.

### **User Support**

Rounding out such a Web-based system is easy access to user support through phone, email, online chat, user forums, and tutorials. All user questions should go into a database that tracks them and creates an accumulating knowledge base, which can also serve to generate Frequently

Asked Questions. Tutorials, some of which may be offered in video format, can be used to provide help in using the data, the online analysis system, and the major statistical software packages. The user forums provide the foundation for an online community of researchers and students who can discuss their experiences using data and learn from each other.

## **The GSS and International Surveys: Issues and Opportunities**

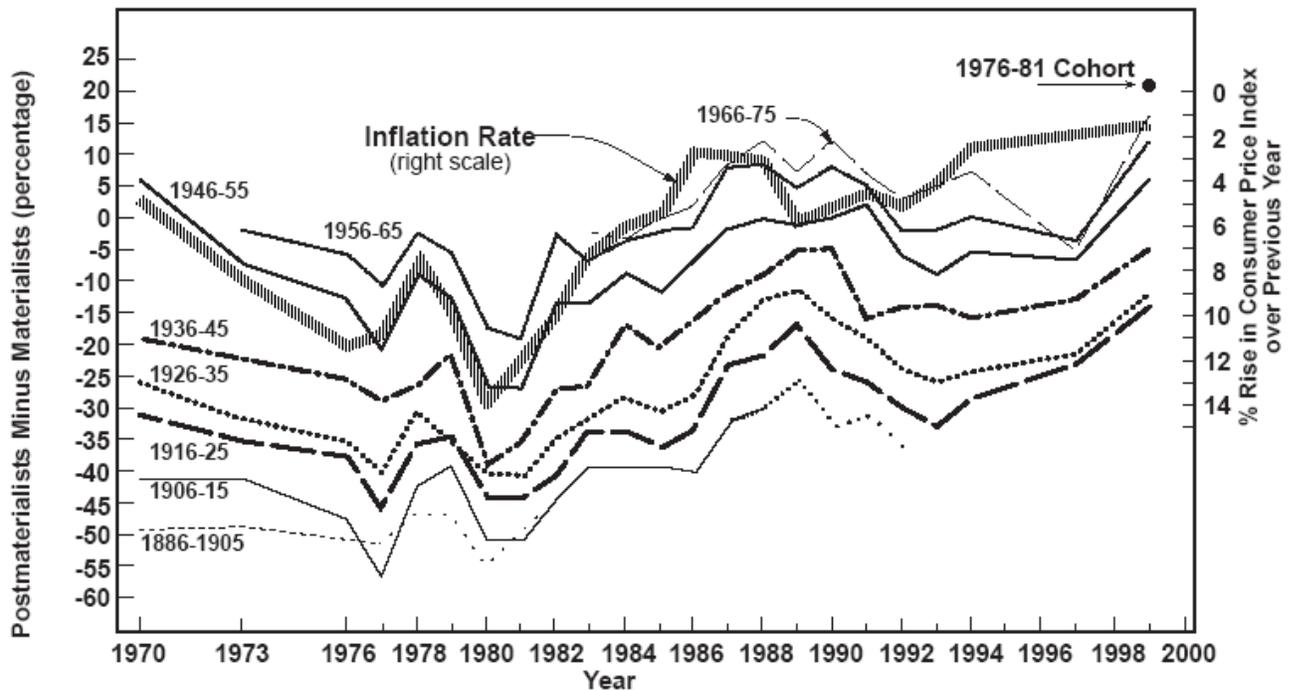
**Ronald Inglehart  
University of Michigan**

The GSS plays a key role in international survey research. In connection with the ISSP, it has produced cross-nationally comparable data measuring a wide variety of important social concerns. These data have been used extensively by social scientists, students and decision makers, and in connection with the emerging North American Social Survey, the GSS has an opportunity to play an even more significant role.

Although the data produced by the GSS has been a valuable resource for social scientists throughout the world, its value could be enhanced substantially through better coordination with other major cross-national survey research programs. There are several reasons why this is true. The most important one is that in order to carry out analyses of social change that can lead to conclusive findings, one needs much more frequent measures of key variables than are now being gathered.

As the GSS proposal to the NSF notes, most important processes of social change take place through intergenerational population replacement, which is almost always accompanied by period effects. Unless one has numerous and frequent measures of such variables, extending over a long period of time, it is almost impossible to distinguish between life cycle effects, cohort differences and period effects. Ideally, such variables should be measured at least once a year; the Euro-Barometers measure a number of variables that are of particular interest to the European Commission twice each year. As the figure below demonstrates, this has made it possible to identify a process of intergenerational value change that has major political and social consequences—the shift from Materialist to Postmaterialist value priorities—but which is complicated by period effects linked with economic fluctuations. Without frequent replication of the relevant questions, this type of analysis would be impossible—but it remains very much the exception rather than the rule.

Although numerous cross-national survey programs now exist, many of the key variables in terms of both theoretical importance and empirical explanatory power, are only measured sporadically—making it difficult or impossible to determine whether one is dealing with long term trends or situation-specific fluctuations. Better coordination of this research could result in agreements to replicate key measures in successive waves of several different programs, providing the type of data base that is essential for analysis of important social changes.



**Figure 4-2. Cohort Analysis with Inflation rate superimposed (using inverted scale on right): % Postmaterialists minus % Materialists in six West European societies, 1970-1999.**

Source: Based on combined weighted sample of European Community surveys carried out in West Germany, France, Britain, Italy, The Netherlands and Belgium, in given years, using the 4-item Materialist/Postmaterialist values index.

1049031

Another strong reason for better coordination of cross-national survey research is that it would enable the GSS and its partners to triangulate their findings across a broader range of societies. Both the ESS and the ISSP research programs are carried out mainly in high-income democracies, complemented by a handful of middle-income countries, almost no low-income countries and virtually no non-democratic societies. In order to analyze some of the most important processes of social change, from secularization to the spread of gender equality to democratization, it is necessary to have a wide range of variation on the dependent variable. The World Values Survey will have covered more than 90 societies by the end of its fifth round of surveys, at the end of this year. These surveys include more than a dozen low-income countries, large numbers of both lower and upper middle-income countries, a score of ex-communist societies and a number of non-democratic countries, extending across every major cultural zone, including a dozen Islamic societies. More effective coordination between the GSS, its partners and the World Values Survey would enhance social scientists' ability to learn how key variables from the GSS function in a wider range of economic, political and cultural settings.

For this purpose, I suggest that it would be useful to establish a working group from the GSS, ISSP, ESS and WVS that would be entrusted with the task of identifying a limited number of key variables that have been used and validated in the respective surveys, and that show promise of

providing scientifically valuable returns if they were more widely replicated—and reaching agreement to do so in coming waves of each survey. Systematic coordination and cross-fertilization of this kind would enrich the surveys and provide even more useful material for social scientists around the world, especially for the study of social change.

## Thoughts on the GSS Recompetition

Jon A. Krosnick  
Stanford University

As the recompetition of the GSS is inaugurated, it is irresistible to pause first to reflect on the many great successes of the project to date. A remarkable number of time series tapping numerous aspects of Americans' attitudes and behavior patterns have accumulated during the last 30 years, and this treasure trove of data has been mined by thousands of scholars who have produced thousands of publications as a result. News media coverage of GSS findings has been continuous, and the data have thusly frequently informed public debates on a wide range of topics. The GSS has served as a training tool for countless undergraduates and graduate students who have been introduced to quantitative social science by illuminating patterns in the survey data. And the data being used in all these ways are of the highest quality: obtained via face-to-face interviewing with enviably high response rates. All of this has been possible importantly due to NORC's commitment to excellence in general and to this project in particular, doing what it has taken at crucial times to maintain the quality of the data collection efforts.

Fabulous. So one might imagine that we should stay the course – keep things going just as they are and continue to equip researchers with valuable tools. And I endorse that notion in general. But in this memo, I will offer some suggestions about how to take a terrific project and make it even better. This seems like a perfect opportunity to consider such innovative possibilities.

Some of the themes I'll raise below have been the focus of discussions among the GSS Board of Overseers for years. And some of them are being addressed now. But because none of these issues have yet been fully addressed, it seemed worthwhile to bring them to the attention of the workshop attendees to perhaps stimulate discussion and influence the forthcoming call for proposals.

### **More Staff**

Whereas the ANES and the PSID have numerous staff members carrying out a range of activities, GSS has typically been staffed by a single principal coordinator: Tom Smith. Tom has been remarkable in this role, overseeing all aspects of the project while maintaining a vigorous schedule of seeking additional funding to augment NSF support and also writing a stream of important publications on public opinion, American life, and survey methodology. Continued achievements into the future by the person playing Tom's role would be even more substantial if he or she were to be supported by at least two additional full-time individuals. Through my experience running the ANES, I have learned that even our staff of 7 is not sufficient to keep up with all the activities of merit for the project as quickly as would be ideal. I am not fully informed about the PSID staff, but I gather that it too is considerably bigger than one person. Realizing the potential of the GSS would be significantly enhanced by additional staff carrying out the range of activities done by the ANES and PSID staffs, including some I will outline below.

## **More Coordination with Other NSF Infrastructure Survey Projects**

Although there are many obvious distinctions among the three NSF infrastructure survey projects, there are some striking similarities as well. Needless to say, all collect and disseminate survey data. All need to document how those data were collected. All need to document past uses of their data by building bibliographies. All need to design new questionnaires with the input of many scholars. And all need to be evaluated in terms of the quality of the data they are producing.

Given these commonalities, it is striking how little coordination and even cross-conversation takes place among the teams coordinating the three projects. I serve on the Board of Overseers of the GSS, as does Bob Schoeni from the PSID. And Suzanne Bianchi, a member of the GSS Board, chairs the Board of Overseers of the PSID. But that's the extent of coordination and collaboration among the projects. Not much.

One thought might be to suggest that the PIs and staffs of the three projects hold at least an annual meeting for at least a couple of days, so they can share their activities and insights and perhaps achieve economies of scale by carrying out joint activities, as well as learning from one another's insights.

One simple illustration of the potential for economy of scale is website development. The three projects all have websites, and they all provide basically the same information. But they are organized in very different ways, guided by very different implicit philosophies of user navigation. Perhaps a single web designer could design and maintain all three projects' websites. Perhaps such coordination would be a nightmare, but perhaps not. Perhaps after some discussions to establish common interests and assumptions, a single, optimal website design could be achieved. I can tell you that the ANES website is constantly being tweaked based on our own intuitions. We can and should do better at this.

## **Augmenting the PI Team with a Younger Member**

My experience as co-PI of the ANES has been nothing short of overwhelming. Admittedly, we set out to be innovative in many ways all at once, substantially broadening the scope of the project and collecting data in many new ways. But still, it's hugely time-consuming, and we're all learning a great deal as we carry out this work.

With this sort of substantial responsibility in mind, it seems wise to see to it that each infrastructure survey is directed by a team of PIs who represent multiple generations of scholars, so that the benefits of learning experiences can be passed from generation to generation as the studies continue for many decades in the future. The PSID PIs have just this sort of structure. In the case of the ANES, both co-PIs are from the same generation, and we have been expanding our pot of wisdom by frequently seeking advice from more senior scholars, some of whom have served in ANES leadership positions in the past. And we are beginning conversations with University of Michigan associate professor Vince Hutchings to increase his involvement with the leadership of the project in coming years.

The GSS would benefit in similar ways from multi-generational team of PIs. If a new PI is to be added to the team, the physical location and disciplinary expertise of that person should be considered carefully. In my experience, cross-country collaboration in running a large project is definitely possible, but I think it works in our case because I was “born and raised” at ISR at Michigan, and my PhD advisor was PI of the ANES when I was a graduate student. So I knew how the project worked from the inside. If there are such individuals who could serve as GSS PIs and who are not at NORC, great. But if not, it may make sense for a new, younger PI to be housed at NORC and the University of Chicago. It may also make sense to consider the possibility that a new PI might be from a discipline other than sociology, to broaden the disciplinary focus of the study even more than it is now.

### **More Methodological Transparency**

We are now in an era of constantly changing survey methodology. No longer can large infrastructure projects settle on a single methodological approach to collecting their data and implement it year after year. Instead, rising costs, increasing insights, and growing challenges mean that survey researchers must constantly be rethinking their approaches, keeping up with the literature, and updating their approaches continuously as new knowledge is gained about best practices.

The GSS has certainly been doing this. Two notable examples of recent changes in methodology are (1) the move to double sampling, and (2) the practice of carrying out some interviews by telephone instead of face to face. Both of these innovations were implemented to achieve a single goal: to maximize response rates. But as far as I know, it is not easy for users to learn the details of how and when and why these methodological changes were made. Furthermore, I am not aware of any work done yet to evaluate whether these methods were in fact effective at achieving their goals and what effects they may have had on GSS data quality. If the community of users were more aware that these methodological changes had been implemented and were informed about the details of how they were implemented in each year, research might be inspired to evaluate the methods’ effectiveness and impact.

Another need for methodological transparency involves the questionnaires used for the GSS interviews. As far as I know, it is not possible for users to obtain copies of these questionnaires. That means that users cannot know the exact sequence in which the questions were asked. Given the large literature documenting the impact of question order, it seems important to equip analysts to know the order in which the questions were asked. Needless to say, providing readable versions of CAPI questionnaires is a challenge, but it is a challenge worth taking on for the GSS user community.

There are many more aspects of the procedures for data collection that should be made fully public for users. All interviewer training materials should be public, because the content of interviewer training can influence the substantive results of a study. I personally believe that such materials should not be proprietary, because analysts need to know this information in order to fully understand the meaning of the data they’re analyzing. Furthermore, providing full details on the procedures of data collection can allow the Board of Overseers and interested

outside scholars to spot strategies that may be suboptimal in light of ongoing methodological advances in the survey research community.

Another striking illustration of the need for more transparency is the set of core questions that are repeated in the survey year after year. As I understand it, no single document exists listing the core questions. This should be produced to make it easy for users to understand what is and is not in the core and to make proposals for changes in the core.

### **A Better Website**

As I have hinted above, the GSS website is sub-optimal. Users should be able to type “GSS” or “General Social Survey” into Google and go to a single webpage that provides a boatload of information on the survey, as is true for the ANES and the PSID. I understand that NSF has recently provided financial support for the creation of such a website, which is terrific and long overdue. I am hopeful that the ANES and PSID websites can be models, illustrating the sorts of information that should be provided to users.

### **Methodological Leadership**

For decades, some of the most important papers on survey methodology were generated using GSS data, many of them authored by Tom Smith. Tom set a standard of productivity and creativity that was very important for the field at that time. A substantial portion of those papers addressed questionnaire design issues, especially question wording and order effects. Tom has continued to produce methodology papers, but (and here, I’m guessing) perhaps the amount of time and energy he has to spend recruiting module sponsors to balance the project budget leaves less time to do such scholarship.

The methodological work Tom did was especially important because it was done on the best data collection platform available (face-to-face interviews with nationally representative samples), and it was done at essentially no cost, because the experiments could be incorporated in ongoing data collections. Thus, GSS data yielded both substantive and methodological insights and advances, the latter essentially for free. And the methodological work presumably informed the writing of new questionnaire items for the GSS and other surveys as well, to maximize reliability and validity.

The absence of such work in recent years has led the GSS Board of Overseers to initiate an effort to re-invigorate this long-standing GSS tradition. At the annual meeting of the American Association for Public Opinion Research in May, 2007, the Board will host a meeting to invite AAPOR attendees to submit proposals for question design experiments to be incorporated in future GSS surveys. The session will review the history of such experiments in the GSS and will provide an overview of the core questions, with which such experiments might be done in future rounds of the GSS.

Another potential route for methodological leadership is in the study of unit non-response. With increasing concern about non-response in surveys, it would be useful to improve the collection and dissemination of information with which methodologists could conduct analyses to assess

whether respondents are systematically different from non-respondents. To do this, the coversheet used by interviewers could be expanded considerably, so they collect hundreds of pieces of information about the sampled dwelling units and their surroundings and the people and activities observable from nearby. This information can be collected before any contact with the household takes place, so the information is obtained on participating and non-participating households alike. This would allow comparisons between them. And since NORC has the addresses of the selected households, it may be possible to obtain some public records information on them, such as whether residents of those houses have voted in recent elections (according to official public government databases). Indeed, it may be possible to link the GSS to confidential data on the households made available via Census Data Centers to augment the available records.

All this would do a great deal to equip researchers to study survey non-response, but only if a mechanism is developed by GSS to disseminate this information to qualified scholars in secure ways. ANES does this through its SPAR (special access requests) procedure. GSS might consider implementing and advertising a similar mechanism.

It would also be worthwhile to consider other ways in which to establish the GSS as a test bed for understanding and optimizing survey methodology, including on issues of questionnaire design, interviewing, interviewer training, interviewer selection and hiring, interviewer supervision, response rates, open-ended text coding, and more.

### **Cognitive Interviewing**

At least ten years ago, the federal survey establishment enthusiastically embraced the notion that pretesting survey questions could be done better than by conventional methods alone. A conventional pretest involves having interviewers ask a small group of respondents the questions, and then asking the interviewers to describe any problems they had. But in recent years, we have come to recognize that this process can fail to uncover significant problems with questions that could be remedied through rewriting. Two particular techniques have been developed to identify such problems: behavior coding and cognitive pretesting.

Until now, the GSS does not subject its new questions to any pretesting procedures other than conventional pretesting. But at its most recent meeting, the GSS Board decided to devote some funds to allow a very small scale trial run of cognitive pretesting of a small set of items, to be carried out by NORC staff. This sort of pretesting requires time, so it will set the schedule back a bit. And it requires a considerable investment of funds that have not been explicitly budgeted in past project grants.

But the fact that GSS has not been doing this pretesting distinguishes it from the vast majority of major federal survey projects. And from my experience, this is a handicap for the GSS, because my experiences with cognitive pretesting have uniformly yielded insights with compelling face validity pointing to needed changes in question wordings to prevent misunderstandings or misinterpretations. It may be worth considering institutionalizing this practice at GSS.

## **Questionnaire Review**

When GSS datasets are released, it is not easy for users to distinguish questions that were in modules paid for by outside investigators from other questions. And in order to balance the project budget, there is considerable pressure to maintain a steady flow of questions funded outside. Many of the proposed questions are reasonably designed according to the principles of optimal measurement. But others are not. In some cases, proposers wish to maintain question wordings to be consistent with their uses in prior surveys in order to track changes over time. But when questions are to be asked for the first time, there is no reason not to improve their design if possible.

Yet the GSS does not have a practice of reviewing these proposed questions and suggesting best practices improvements to the proposers for them to consider. In some cases, the proposers may be resistant. But in other cases, the proposers may be open to suggestions and even grateful for them. And if such improvements are made prior to fielding the questions, the entire user community may benefit from the improved questions.

Interesting, a considerable number of questions in the core are designed in suboptimal ways. In order to maintain continuity, we might want to stick with those wordings. But it is also possible to implement “splicing”, whereby half of a new sample is asked the old version and the other half is asked a new version. If this splicing is done for a couple of rounds of data collection, it may be possible ultimately to shift exclusively to the new wording while equipping analysts to connect trends lines across the splice. It may be worth considering this practice.

## **More Aggressive Marketing Efforts**

Although the GSS webpage includes a sheet of instructions on how to propose a new module of questions, I am not aware of regularly implemented marketing efforts to notify broad communities of academic social scientists about this opportunity. This may be worth doing via email, newsletter announcements, and even paper mail, to enhance the likelihood that academics will fill the questionnaire with broadly valuable questions.

We have been doing this sort of thing with the ANES, including via public competitions soliciting proposals for questions to be included in the questionnaires at no cost to the proposers, and we have gotten many excellent suggestions, indeed considerably more than we could accommodate. Perhaps the GSS could consider doing the same.

## **Technological Innovations**

Because the GSS is done with CAPI, it would be possible to turn the laptop around to face the respondents and make use of computer software to innovate in two ways: (1) present visual or auditory stimuli to which respondents react, and (2) measure reaction time when making various sorts of reports. One particularly active area of research these days using these approaches is in the measurement of racial prejudice, an area in which the GSS has a long history of excellence. Social psychologists have developed measures like the Implicit Attitude Test that present visual stimuli and measure reaction time in order to measure attitudes in ways that skirt direct self-

reporting. The GSS might consider making best use of its CAPI approach by incorporating such measurements.

## **Conclusion**

In conclusion, the GSS is and has been a wonderful project and has contributed tremendous riches to social science. In the next round, there are various ways in which NSF can enhance the value of the GSS even more. I hope the comments above are helpful in suggesting some possible directions for innovation and improvement.

## **Operational Aspects of the GSS from the Standpoint of Board of Overseers**

Robert D. Mare, GSS Board Chair (2004-2008)  
University of California, Los Angeles

The General Social Survey (GSS) is supported by funding from the National Science Foundation, supplemented by payments from users who contribute “modules” of questions of varying length and complexity. NSF funding is awarded on a five-year cycle and supports the bulk of the survey core. Survey operations are carried out by NORC under the direction of the GSS PI’s, Drs. Tom Smith, James Davis, and Peter Marsden.

### *Mission of GSS Board of Overseers*

The GSS Board of Overseers is an independent body of scholars who serve NSF and are entrusted with providing review and oversight of the GSS. As set out in the “Charter of the Board of Overseers of the General Social Survey,” the Board’s mission is to, in consultation with the principal investigators and the GSS staff, review the work and develop plans and budgets of the GSS; advise and consult with the PIs in developing proposals to agencies or foundations; in consultation with the principal investigators and representatives of funding agencies, approve priorities and the allocation of time in the survey instrument (including the balance of continuity and new areas of inquiry); approve the questionnaire proposed by the GSS staff; take other steps to enhance the scientific value of the GSS, such as recommending to the GSS research on issues of measurement and validity and undertaking its own studies to assess the quality of the GSS data.

### *Make-up of Oversight Board*

The Board is made up of between 9 and 15 scholars chosen to represent the GSS user community with respect to academic discipline, substantive area of research, methodological expertise, institutions, and geography and in a way that is consistent with the NSF’s affirmative action goals. Although Board members are selected carefully, full representation of all constituencies on a small board is seldom possible at any given time.

Over time, however, the Board tries to ensure that important constituencies are not excluded for a long period. The Board has traditionally been dominated by sociologists, although, at any point during the past five years, between one and four Board members have been political scientists. The Board currently has one economist. The Board typically includes expertise on race-ethnicity, politics, stratification, gender, family, religion, health, crime, law, demography, and survey research methodology. Efforts are made to include persons who have had experience with other major survey data collection efforts and with the GSS as researchers, teachers, and module developers. In recent years, the Board has included persons who have served on the oversight boards or as investigators for NSF’s other two long standing social surveys, the American National Election Survey and the Panel Study of Income Dynamics. Affiliates of the University of Chicago or NORC are not eligible to serve on the Board.

Board members typically serve four-year terms. Officers of the Board include the Board Chair, who serves a two-year term (renewable for a second term), a Board representative to the International Social Survey Program (ISSP), and the Chair of the Board's Long Range Planning Committee. New Board members are selected through election by the current Board, subject to final confirmation by NSF staff. The Board comprises persons who have varying lengths of tenure. Because the Board selects its own colleagues and successors, it must weigh the desirability of recruiting highly qualified and compatible members against an inward-looking tendency to select persons who resemble themselves too closely. For the most part, the Board has successfully maintained a balanced approach to this issue. Further details about the Board's membership rules are provided in the Board Charter.

### *Board Activities*

Most of the Board's work occurs during and around its regular semi-annual meetings with the PIs and representatives of NSF, typically Pat White of the NSF Sociology Program and Ed Hackett, Director of NSF Division of Social and Economic Sciences. The Board's specific activities depend in part on when meetings occur relative to the two-year cycle of planning and implementing the GSS. In a meeting during the later stages of data collection (e.g., Fall 2006), the Board receives and comments on reports supplied by the PIs, the NORC Vice President, and representatives from field staff. NORC representatives attend these Board meetings, make an oral presentation, and field questions from Board members about response rate, cost, data quality, interviewer performance, and other operations issues. In meetings that take place shortly after data collection is complete (e.g., Spring 2007), more attention is given to longer range issues.

In all meetings, but especially those leading up to a new survey, the Board devotes considerable attention to review of topical modules that have usually been proposed by third parties (non-PIs, non-Board members). Between meetings, selected Board members carry out regularly schedule activities, such as attendance at ISSP meetings, or ad hoc Board business. Recent examples of the latter include a review of possible items on the measurement of gay identification, development of a module design to explore the potential of the new GSS panel, and sychometric analysis of the GSS vocabulary test WORDSUM. The Board also receives reports from the PIs about funding the upcoming survey, plans for the ISSP, and other logistical or substantive issues that the PIs bring to its attention. The Board's Long-Range Planning Committee usually meets the day prior to the regular Board meeting. This committee identifies longer run issues that require the Board's or the PIs attention, including enduring procedural problems and periodic review of core questionnaire items.

### *Module Development and Review*

Topical modules originate from university-based academics with funding from government or foundation sources; researchers in government agencies; from PI Tom Smith in collaboration with a funding source; and from Board members themselves. The GSS is an attractive vehicle for modules because it is a face-to-face interview with a comparatively high response rate, it contains substantial complementary demographic and attitudinal content in the core part of the survey, and its regular schedule provides the opportunity for longitudinal collection of module items. Module proposers often approach the PIs or, less often, the Board with ideas for modules

before they obtain funding. Proposers confer with Tom Smith about cost, time estimates, and timetables. The Board reviews these initial proposals – usually in the form of a written prospectus although sometimes on the basis of a short oral report by Smith. The Board comments on the scientific merit, technical quality, and feasibility of the proposals and encourages or discourages further development. Smith conveys the Board’s reactions to module proposers. Proposals that are given a “green light” by the Board for further development are, subject to the developers’ success in securing funding, further reviewed for substantive and technical merit by the Board when specific questionnaire items have been written. In principle, the Board can accept or reject a module at this time. In practice, when necessary, the Board typically supplies critical advice for further improvement, which Smith conveys to the proposer.

### *Funding of the Board*

The Board is supported by a segregated part of the NSF core grant. The Board’s funds pay for the costs of semi-annual meetings, GSS pretests, and board initiated studies.

### *The Problem of Innovation*

The procedures described above generally work well for both the PIs and the social science community. Nonetheless, the Board, the PIs, and NSF are constrained by the exigencies of funding and scheduling a regular large scale survey that is largely devoted to replicated measurement. These constraints create ongoing operational problems.

- The PIs must complete a fully funded survey in timely fashion every two years. A large fraction of GSS funding comes from modules funded by sources other than the NSF core grant. The PIs are constrained to accept modules that are technically competent but may not be of general scientific interest, or that may, in the Board’s judgment, be of marginal technical quality. The PIs and Board must balance the need for funds against the scientific quality and significance of the survey.
- The time between when module questionnaire items are submitted for Board review and deadlines for pretesting and CAPI programming is short. Requests by the Board for further revision may be impractical if a module is to be included and its financial contribution secured in time for a given Survey. This limits the Board’s ability to influence survey content and quality. Requiring proposers to submit initial and final module plans farther in advance of final deadlines may discourage proposers and jeopardize key funding sources.
- The GSS should balance replication in substance and method with substantive and procedural innovation. The PIs are constrained by their funding, which provides limited room for major innovation. The PIs tend to make their greatest efforts at innovation when they apply to renew the core NSF grant. The outcome of this application largely determines the scope of innovation for the next five years. For example, long overdue innovations such as the panel and Spanish interviewing only appear in the 2006 GSS after repeated unsuccessful past efforts by the PIs to secure core funding for them. The Board, however, while firmly committed to replicated measurement and mindful of the PIs circumstances, tends to push for scientifically valuable changes without regard for the five-year funding cycle.

The Board's impulses at times do not jibe with the priorities of the PIs who, understandably, are concerned with replication, deadlines, and budgets. Some low cost (yet highly desirable) innovations have occurred mainly through the Board's efforts at persuasion. For example, following extensive discussion between Board and PIs, the GSS will now release interview "process" data (e.g., numbers and types of contacts, interviewer characteristics, etc.) on the public file. This innovation required no supplementary funding. Other innovations, such as a state of the art website for public access, while long a Board priority, is only now being implemented because of the PIs success in obtaining supplementary NSF funding. Yet other costly innovations, such as cognitive pretesting of GSS items remain under discussion between PIs and the Board.

## Issues of Data Quality and Data Generation: The General Social Survey and Ethnomethodology/Conversation Analysis

Douglas W. Maynard  
University of Wisconsin—Madison

Ethnomethodology (EM) and conversation analysis (CA) may seem unlikely bedfellows with survey research but recent developments in both areas indicate otherwise. On the survey side, the possibility of digitally recording CATI interviews, as was done in the 2004 round of the Wisconsin Longitudinal Study (WLS), means a low-cost way not only for storing raw interview data in the form of the talk transpiring between interviewers and respondents but also for enabling inquiries such as EM and CA that investigate structures of talk and collaborative practical actions. On the CA and EM side, there has long been a concern with how scientists, including social scientists, do their work. In the social sciences, such a concern dates to Garfinkel's inquiries regarding record-keeping, coding, interviewing, the use of evidence, and the like. The concern has always been with the "how" of these tasks and the social orderliness involved in accomplishing them.

More recent studies have especially deployed CA to study interaction in the survey interview and to analyze the "how" in terms of tacit and taken-for-granted practices or talk-based mechanisms by which participants—interviewers and respondents alike—assemble the data that emerges from their relatively brief and anonymous encounter. In connection with studying the "how," I invite the reader to consider three things. First is to explore a transcript illustrating practices of talk and interaction and how they affect administering a question and obtaining an answer. Second is the question of why the study of interaction matters to improving the quality of survey data. And third is to suggest the continued importance of possibilities for crossing the line between qualitative and quantitative research.

### (1) An Illustration

Conversation analysis is concerned with the organization of interaction. There is an *interactional substrate* to the conduct of the interview whose properties can affect how well the interview is performed in the standardized way it is supposed to go. Anyone who has listened to an actual interview, for instance, knows that interviewers often depart from the script they are supposed to read, and sometimes this violates standardization. That interviewers do so often can be ascribed to their *skill* or *competence* as interactants rather than the lack professional expertise. Consider an example from the Current Population Survey, and specifically a question that asks what kind of enterprise the respondent's place of work is. (The respondent in this example works for an insurance company. Although we will consider subsequent talk as well, the focal question is at lines 4-5, and the answer of initial interest is at line 7 (FI = female interviewer; MR = male respondent):

## CPS Interview 007c (Normalized transcript)

FI: And what kind of business or 1 industry is this?  
2 MR: The insurance industry.  
3 (7.0 seconds silence) ((typing))  
4 ->FI: Is this business or organization mainly manufacturing retail trade  
5 wholesale trade or something else.  
6 (1.0)  
7 ->MR: It's a service industry  
8 (1.8)  
9 FI: So it'd be under?  
10 (2.0)  
11 ->MR: Well it wouldn'- sh'wouldn't be manufacturing or retail or (0.9) or  
12 anything like that it's (0.7) I don't know how- I don't know what  
13 you'd (.) classify it.  
14 ->FI: Under something else.=  
15 MR: =Yeah:  
16 (1.0)  
17 FI: And what kind of work do you usually do at this job that is (.)  
18 what is your occupation.

When the respondent, at line 7, answers the question, he does not choose one of the categories he has been given, and instead offers a new category. There are good interactional reasons for this answer. These reasons have to do with ways that listing things in talk conditions how a recipient processes a final item. Here the interviewer lists particular categories of business organizations, and ends with a generalized term. That way of listing can indicate to the respondent that, if the others do not apply, then he should name a category that does. That is, the practice here is that the respondent treats “or something else” as an *invitation* to complete the list with another, particularized category relevant to his situation, rather than as an utterance containing a response category in its own right that he can choose.

And now notice how the interviewer deals with the respondent’s answer. He produces a probe at line 9, in a neutral way that fits the canons of standardized interviewing. This probe asks the R to use categories already mentioned rather than adding to them. So far, so good, except that at line 10, the respondent delays for two seconds, and, at lines 11-13 further indicates trouble with the question by denying the relevance of the categories so far named. In an expression of uncertainty, he announces that he doesn’t know “how” or “what” the classification would be. To that characterization of the response we need to add an analysis about its fundamentally interactive character. In terms of *sequencing* in conversation, the turn is a practice called “reporting” that solicits guidance or help by implicating the relevance of its recipient gathering an upshot of the report. In other words, it invites the *interviewer* to produce a candidate answer; note how the interviewer at line 14 deals with the respondent’s utterance by proposing one of the original categories as a possible answer. The respondent's agreement ("Yeah," line 15) accepts the proposal and ends the verbal exchange such that the interviewer can record the answer (which is probably what the silence at line 16 indicates) and move on to the next question (lines 17-18).

### (2) Why Interaction Matters to Improving the Quality of Survey Data

The example illustrates how interaction matters and how analysis of interaction can be consequential for survey design and implementation in at least two ways. For one, although there is a sophisticated literature on question wording, it is usually about semantics or how different words referring to the same thing can produce different response distributions. There is also a literature on “context effects” but that usually means the relation of an item on a questionnaire to preceding items. Conversation analysis examines words in relation to the very local sequential context of their production—their relation with other parts of an utterance, and also how an utterance is prospectively and not just retrospectively contextual. An utterance works as a social action to occasion a specific responsive action from its recipient. Research has yet to incorporate CA findings about talk to see how, in addition to being cognitive phenomena, different forms of questions operate as interactional items and, as such, may affect response distributions.

Another way that interaction matters for survey design and implementation can underline the matter of *competence* on the part of both interviewer and respondent and its effects on the data. In the example, if we were to focus only on the respondent’s hesitations and uncertainty in answering, he might be regarded as rather inept at parsing a straightforward question. What we know about conversation, however, suggests that he engages a practice that participants regularly use to solicit inferential upshots from their co-participants. The practice, as mentioned, is called *reporting* and it’s what we do, for example, when someone invites us to a movie and we say, “Sorry, I have to work tonight.” That is a report that solicits the inference from the inviter that the answer is “no.” Reports strongly compel responsive inferencing, and the recipient of a report may go beyond cognitive deduction to do what the interviewer does in the example, and that is to offer a candidate upshot for the speaker of the report to confirm. Once again, that is a matter of interactional competence. Interactional competence often runs up against and often takes precedence over acquired skills for official, standardized interviewing. It sometimes looks as if the interviewer is being unprofessional because of the departure from standardization—more concretely, in this instance she seems to be unskillful in the way she should do a neutral probe. But the interviewer here appears perfectly competent as a conversational participant. The point is that we need more investigations regarding how the tension between these two forms of competence or skill (interactional vs. standardized interviewing) affects data quality. When interviewers probe incorrectly, is it because they are prone to violate standardization, or are they giving precedence to interactional practices? If interviewers avoid doing the interactionally appropriate moves because they must follow procedure does this affect subsequent answering on the part of the respondent? For example, does it raise the frequency of item nonresponse? To be able to investigate such matters means obtaining audio or (preferably) video recordings of the interview and being able to study them with such tool sets as EM and CA.

### **(3) Crossing Over: Qualitative and Quantitative Research**

Four years ago, NSF conducted a “Workshop on Scientific Foundations of Qualitative Research.” Qualitative researchers were called together and discussed, among other matters, the crossovers between qualitative and quantitative research—“hybrid” relationships, the “serious

use of both kinds of methods to analyze central processes” (rather than one being the handmaiden of the other), and spanning “case-oriented and variable-oriented” research.\* In this crossover, the integrity of both kinds of research can be preserved. Such crossover research is important as an end in its own right and as a means toward addressing the fundamental problems of improving the quality of research of many kinds and in many domains. At Wisconsin, using the Wisconsin Longitudinal Survey and its digitized telephone interviews, we are engaging in collaborative hybrid studies both to better understand cognitive measurement and to find out what the effects on participation may be when taking into account conversational practices for soliciting participation while controlling for respondents’ propensity to engage in the interview. Crossing over is hard work but it can be facilitated by available technologies of recording that, in turn, allow for research on real and actual social processes and practices. By making hybrid, crossover, and case-oriented research more possible, the proposal for the GSS to continue its basic mission of gathering data on American society, comparing the U.S. to other societies, and making high quality data accessible to scholars and students, can be enhanced.

---

\* See Charles C. Ragin, Joane Nagel, and Patricia White (2004). Workshop on Scientific Foundations of Qualitative Research. Sociology Program and Methodology, Measurement & Statistics Program in the Directorate for Social, Behavioral & Economic Sciences. On the web: <http://www.nsf.gov/pubs/2004/nsf04219/nsf04219.pdf>

## **Review of the Content of the GSS**

Leslie McCall  
Northwestern University

As a relatively new user of the GSS, I have found it to be an enormously important and well-documented resource for understanding social attitudes and behavior. After studying patterns of wage inequality using census data, I became interested in the question of whether Americans were aware of rising income inequality. I turned to the GSS because it was the best available data source for examining this question.

It is only within this very positive spirit, then, that I offer suggestions for possible improvements in the GSS. My suggestions are based on my experience studying a topic that has not been central to the GSS core or topical modules (but, rather, to the ISSP topical modules). Therefore, the issues I raise may not be shared widely by other users. Nevertheless, I hope they will be useful for discussion, or reconsideration as the case may be.

### **Linkages among Core and Topic Modules**

One of the key advantages that the GSS has over other surveys is its scope (across time and subject matter). Even if the GSS replicates topics asked in other surveys (e.g., NES, WVS), it may have other variables and modules that those surveys do not, and it may have them over a longer time period. This provides a unique opportunity to analyze relationships among domains.

For example, although the NES has an extensive battery of questions on policy preferences, its questions on egalitarianism are worded in generic terms, making it difficult to examine attitudes toward the specific issue of growing income inequality. The GSS, on the other hand, is less focused than the NES on policy preferences but has the advantage of having fielded four cross-sections of questions on income inequality. However, the questions on policy preferences are not perfectly aligned, temporally, with the questions on income inequality in the GSS. Another example concerns questions about changing economic conditions for individuals and their families. Such conditions could be an important influence on attitudes about income inequality, but these questions were not asked in one of the four years that the income inequality questions were asked. And, as explained below, each available year of data is critical.

It is probably too much to ask for “perfect alignment,” so my more general suggestion is for greater attention to potential linkages among different topical modules and between elements of the core and the topical modules.

### **Studying Social Change: Cyclical/Structural Changes Versus Gradual Cohort Replacement**

My sense is that gradual social change related to cohort replacement represents the dominant approach to studying social change among researchers using the GSS. For example, the 2004 NSF proposal for the GSS states that “most social change in attitudes is slow, steady, and cumulative, explained (in decreasing order of importance) by a) cohort-education turnover models, b) episodic shocks (e.g., wars and political scandals), and c) structural changes in

background variables” (p. 6). This assessment is no doubt true of the topics that have been studied extensively with GSS data (e.g., attitudes on social issues) and much social change does clearly transpire in this way.

However, should the GSS also facilitate analysis of social change related to episodic change and structural cycles or transformations? Again, income inequality provides an example. In understanding whether, and if so why, attitudes about income inequality shift, most researchers would argue that background conditions should be important, such as actual trends in income inequality and the business cycle, or knowledge of increasing inequality and structural economic transformations. If this is the case, timing is extremely important. For example, questions asked during the downturn as well as the upturn in a business cycle would be important for determining whether the business cycle influences beliefs about income inequality. The fact that the questions on income inequality were asked in 1992, for example, and inexplicably again in 1996 (because the Social Inequality Module of the ISSP was not asked in any other country besides the US), was extremely fortuitous. These years occurred during important parts of the business cycle, including a downturn.

Again, my general point here is not that the GSS provide questions on awareness of the business cycle (as the NES does) or that it try to time questions according to the US business cycle, which is impossible. Rather, it is to consider the degree to which some topics might be more sensitive to timing—because of their more episodic and cyclical nature—and thus in need of more regular or frequent intervals of replication. A related issue is the extent to which the GSS can act to replicate topical modules that are time sensitive but do not have advocates for replication, assuming that the overseers deem the issue of significant social and scientific importance.

### **Theoretical Background for Development of Questions and Question Wording**

I suspect that there are a lot of users like myself from fields outside of social psychology (and perhaps inside social psychology) who are skeptical at first of many of the GSS questions (and their wording). Such scholars will not make productive use of the GSS despite significant interest in the topics covered by the GSS. I think this pertains to some of the questions about income inequality, in fact. It may be that the GSS does not need more users, but I think it would elevate the quality of research if documentation were available that explained the theoretical rationale of questions, especially those in the topical modules. Such information should be available from the proposals written for new questions. Perhaps these could be made public in some form.

### **Current Adequacy of the Core**

My only suggestion here would be, if at all possible, to reduce the number of questions on some topics that take up a substantial share of the core. I defer to experts in the respective fields, but it seems that there are a lot of core questions on religion and family relationships, for example, that could be condensed without loss of information.

# **Conceptual and Methodological Innovations & Contribution of the GSS to Sociology and Its Broader Impacts**

**Steve Nock**  
University of Virginia

## **Conceptual and Methodological Innovations**

The proposal outlines a wide range of innovative changes to the GSS that, in my opinion, are justified, timely, and of great value. Here I comment on the value of those innovations and make suggestions about future developments. I rely mainly on the recently funded proposal as well as supplemental materials provided. I also rely on my own experience with the GSS over many years.

The most obvious innovation is the introduction of a panel component to the GSS. I read an earlier draft of this proposal and was not convinced that the GSS should move in this direction. There are a growing number of excellent longitudinal studies now, and I was uncertain whether the GSS would add value beyond that currently provided by sources such as NLSY, PSID, Add-Health, and the many other surveys that lend themselves to general social-science applications. I am now convinced that this new strategy is valuable, especially because of the retrospective panel proposed. By re-interviewing respondents from earlier rounds of the GSS, panel data would become available immediately to users. Since GSS was not designed as a panel, however, this new design will encounter some problems.

The proposal mentions the need for a panel to facilitate studies of individual changes and transitions. While this is certainly the primary rationale for any panel, the current GSS proposal goes specifically to life-course transitions like entry into cohabitation, employment, marriage, divorce, births, and so on. This may present a problem. The GSS has not asked about age at first marriage (or age at subsequent marriage – variable AGEWED) since the 1994-95 administration. Cohabitation, divorce, retirement, and other life event dates present a comparable problem. With respect to fertility, similar issues arise for all but the first birth. A 10-year retrospective panel, therefore, could not establish marriage cohorts, nor permit research on life-course transitions that require dating of events. These problems reflect the initial design of the GSS that did not anticipate a panel component. I believe, however, that any future innovation proposed must give serious attention to event-history issues. In fact, one of the additional ‘follow-up and auxiliary’ studies currently being considered focuses on partners, and another on intergenerational transfers. Both of these will require considerable detail on the composition and dates of events. Future administrations of the GSS, in short, will need more detail on event histories.

The proposal outlines an ambitious and needed augmentation to the “contextual and geographic data” component of the GSS. The lack of GIS information to the average user has increasingly limited the attractiveness of the GSS. Widespread availability of GIS software that easily integrates Census and non-Census data with geographic information has made surveys without geographic identifiers less attractive to students, especially, and to those researchers who are

interested in context more generally. I am happy to see that the proposal acknowledged the growing request NORC has received for geocode data. I believe this will be a major improvement for GSS generally and will expand its usage, especially among urban and environmental scientists, political scientists, as well as sociologists. As I outline below, future applications of the GSS should allow simple mapping of variables by geographic unit to the extent that cases support it.

Web access to the GSS, while widespread, is rather primitive at present. The current GSSDIRS is out of date and limited in its utilities (especially statistical and graphic). Factfinder (Census), for example, allows the user to conduct very basic elementary analyses, but also produce graphs and maps. In fact, Census provides a model for how data may be analyzed or distributed on the Web. The GSS is still analyzed, most frequently, by accessing a file of raw data (especially for years that are not easily available on the Web except from several university archives (e.g., Berkeley). The average user should be able to conduct basic and moderately advanced analyses on-line without the need to download data. I am not sufficiently familiar with the DDI XML protocols, but assume these will take GSSDIRS to the state-of-the art. This is overdue and represents a very significant enhancement to the GSS.

The Spanish language translation of the GSS is an obvious and much-needed innovation. The proposal was somewhat vague on how the translation is to be done and/or how conventional vernacular English terms will be converted for a Spanish-speaking population that includes Mexican, Cuban, and other Spanish groups (e.g., does “respect” mean to feel equal to someone as in English, or does it mean to look up to someone as in Mexican Spanish?). I presume that NORC has sufficient experience in both language translation and meaning translation to make these accurately. I am undecided about whether additional language translations might be justified, and hope the group might discuss this.

Finally, the proposal outlines two strategies that will encourage greater participation in the ISSP by less developed nations. These make good sense. But it is not clear how it might happen.

### **Contribution of the GSS to Sociology and Other Social Sciences and its “Broader Impacts”**

I believe the proposal adequately and faithfully represents the enormous contribution that the GSS has made both in the academy and elsewhere. As a GSS user for over 30 years, I have seen widespread applications of these data by colleagues (and students) in sociology, political science, economics, and in the federal government. All of these are summarized in the proposal.

The proposal also mentions the widespread use of the GSS in the classroom, noting how several leading research-methods and statistics books include modules that rely on these data. The proposal also mentions how teaching packages are often developed by instructors for using the GSS. My undergraduate students typically produce simple frequency distributions from the GSSDIRS site (or an alternative) in the first week of class. Beyond that, however, the student needs a significant amount of instruction and assistance to conduct meaningful analysis beyond simple tabulations. This is especially so when a student is interested in conducting analyses over time, recoding many variables, creating many composites, and so on. For these applications, the instructor typically must first teach students something like SAS or SPSS, and then provide a

data file for analysis. This works well, but requires significant effort and resources before the student sees much in the way of results.

The PSID has a series of teaching modules that help students access the data, combine or alter them as needed, and produce results. These are found in the “Tutorials: PSID in the Classroom” section of their site. These teaching modules are well designed with objectives, and analysis-results questions that test the user’s understanding. Though the PSID is vastly more complex than the GSS is currently, the addition of a panel component will change that very quickly. Students (and many seasoned researchers) will need more help at that point. And even in the current version of the GSS, such modules would be particularly valuable for educators. How, for example, might a student conduct a birth-cohort analysis of political orientation, and so on? This would be a simple module to develop and would add significantly to the teacher’s resources when moving into longitudinal designs. A similar module on the ISSP would be particularly valuable. My basic point is that while the GSS is unquestionably valuable in the classroom, there is too little instructional information available to facilitate its use.

To the extent that such instructional materials could be provided, the larger impact of the GSS would probably increase. Those who are less familiar with survey research and basic statistics, for example, will struggle with the GSS as it is now. But this is not so for resources such as the PSID where step-by-step instructions guide a novice user (presumably a second or third year undergraduate or senior researcher without the requisite skills) through the steps required for analysis (and includes a test of whether the user did things correctly). I strongly encourage greater development of teaching materials for the GSS, both elementary and advanced.

As the proposal mentions, the items included on the GSS are probably among the most tested available. The design of the questionnaire, likewise, represents the best known strategies. Since the transition from paper-and-pencil instruments, however, it has been difficult to know the entire range of items included in the GSS. The indexing on GSSDIRS is good, but still misses many topics and themes because the variables are not linked to the variety of issues some users might search for. For example, there is no entry for FERTILITY, though there is for CHILDREN. It is this type of index linkage that I am hoping will be included in the newly designed web resource. But more generally, the novice user struggles to understand the range of issues and topics covered in the GSS. This, I believe, limits its broader impact, especially to those who are not students, academics, or trained researchers. I would suggest that something similar to what one finds for the NSFH be included, that groups comparable elements of the questionnaire by broad topic and allows the user to link directly to those variables. Since an Interview-to-Internet and extensive hyperlinks are proposed, I believe it would be worth considering the development of a concise (one or two page) index that links the user to all related materials associated with those broad headings (neither GSSDIRES nor the Berkeley GSS site offers this).

# The General Social Survey: Contributions to Economics and Recommendations for Future Dissemination

Gregory N. Price

## The GSS And Its Contributions To The Economics Literature

Since its inception in 1972, the United States General Social Survey (GSS) has allowed economists to test and explore theories that otherwise would have not been possible. The rich array of data on individual and household characteristics measured in the GSS—particularly those captured through topical modules—have turned out to be a veritable gold mine for economists. A search on "General Social Survey" in the American Economic Association's "EconLit" bibliographic database results in some 76 articles in the economics literature that either cite and/or report results based on GSS data.<sup>\*</sup> An examination of this count shows that GSS data have been contributed across the subfields of economics, including for example public economics, labor economics, and the economics of religion.<sup>†</sup> While the absolute count of articles either citing and/or using GSS data seems low relative to the economics literature in general, it is also instructive to note the quality of the contributions. Of the 76 articles in the economics literature that either cite and/or use GSS data, 14 represent contributions in the top 36 economic journals.<sup>‡</sup>

---

<sup>\*</sup> See itemization of these articles in "The General Social Survey In the Economics Literature" GSSWorkshop Report (Price, 2007). The number of articles may be an undercount, as the search only includes articles in the economics literature that has the word/phrase "GSS" in the article abstract, and many listed articles do not include abstracts.

<sup>†</sup> This author, having research interests in the economics of religion (Granger and Price, 2007), has certainly benefited from the availability of GSS data.

<sup>‡</sup> The top 36 economics journals are based on the criteria of Scott and Mitias (1996) and as amended by (Price, 2005) are: *American Economic Review*, *Econometrica*, *Economic Inquiry*, *Economic Journal*, *Economica*, *Industrial and Labor Relations Review*, *International Economic Review*, *Journal of Business*, *Journal of Business and Economic Statistics*, *Journal of Development Economics*, *Journal of Econometrics*, *Journal of Economic Dynamics and Control*, *Journal of Economic History*, *Journal of Economic Perspectives*, *Journal of Economic Theory*, *Journal of Finance*, *Journal of Financial Economics*, *Journal of Labor Economics*, *Journal of Human Resources*, *Journal of International Economics*, *Journal of International Money and Finance*, *Journal of Law and Economics*, *Journal of Law, Economics, and Organization*, *Journal of Legal Studies*, *Journal of Monetary Economics*, *Journal of Money, Credit, and Banking*, *Journal of Political Economy*, *Journal of Public Economics*, *Journal of Regional Science*, *Journal of Urban Economics*, *National Tax Journal*, *Public Choice*, *Quarterly Journal of Economics*, *Rand Journal of Economics*, *Review of Economic Studies*, *Review of Economics and Statistics*, *Southern Economic Journal*.

As journals in the top of a ranking hierarchy are more likely to be cited (Kalaitzidakis, Theofanis, and Stengos, 2003), it is likely that the relatively low count of economics articles citing and/or using GSS data grossly understate the impact that articles have on economic science overall. In addition, 6 of these articles are listed as working papers of the National Bureau of Economic Research (NBER). As NBER papers have a tendency to be cited in other economics journals, this is yet another source of downward bias in using EconLit counts of economics articles as a measure of the GSS' impact in economics.

### **Some Thoughts on Future Dissemination Strategies**

As a recipient of tax-supported research dollars, GSS data are required to be publicly available at no cost to the user. It appears that to date, GSS data are freely available via the worldwide Web at: (1) Queens College (City University of New York (CUNY))§ (2) Interuniversity Consortium for Political and Social Research (ICPSR) and (3) University of California-Berkeley, Survey Documentation and Analysis (SDA) Archive. Surprisingly, the National Opinion Research Center (NORC)—organization that houses the principal investigators of the current National Science Foundation grant that supports the GSS—does not provide direct access to GSS data. Instead NORC provides links to one of the previously mentioned sites, and to others, some of whom will make the GSS data available in some portable format (e.g. Compact Disc) for a fee.

In general, worldwide web dissemination mechanisms for the GSS seem to warrant a passing grade of "fair". My recommendations for future web-based dissemination strategies are:

- *Data Format and Current Data.* There should be some consideration of introducing an option for data that can be processed with STATA. Increasingly the econometric/statistical software of choice for social scientists is STATA as it is a low-cost and powerful software for exploiting the tools of modern econometric/statistical methods in research. With the exception of the UC-Berkeley site, the web-based GSS download sites only allow for SAS and SPSS options. While an ASCII file is generated that can be imported into STATA with some effort, this is not an easy option for many potential users of GSS data. The web-based sites also have dissimilar end-years for the GSS data.\*\*\* While NORC indicates on its website that the GSS is available through the year 2006, the latest year for which one can freely download GSS data is 2004, and at the ICPSR site—1998. For a fee, one can apparently obtain GSS data through 2006 through the University of Connecticut's Roper Research Center. This is confusing and frustrating. The data should be made available to the public through the latest year possible.

- *Standardize Data Downloads.* Only the UC-Berkeley site appears to allow the user to select particular years of the data. If the others do, it is not apparent. Users of GSS data should be able to easily download variables of interest—particularly those in topical modules—and link them to similar variables in the same year. Currently, this is difficult to implement, and is only possible through the year 2004 at the UC-Berkeley website. The GSS should aim for standardizing data downloading that is similar to the ease of downloading census data at the University of Minnesota's Population Center—the Integrated Public Use MicroData Series (IPUMS).

---

\*\*\* See [http://soc.qc.cuny.edu/QC software/extract.html](http://soc.qc.cuny.edu/QC%20software/extract.html), <http://www.icpsr.umich.edu/GSS/> & <http://sda.berkeley.edu/archive.htm>

Matching particular variables on individuals/households to a particular year is simple through the IPUMS website—something which should be possible with GSS data as well.

- *Provide Users with Updates.* One of the virtues with using IPUMS data is that by requiring users to register, users are constantly provided news about important updates. Currently, the web-based platforms for downloading GSS data do not require user registration, nor is there any way GSS users can learn about updates, unless apparently, they happen to browse the NORC website—which does produce a GSS newsletter. User registration with required email address would remedy this shortcoming and keep GSS users informed about important updates.

## **Conclusion**

Economics has benefited from the GSS, as evidenced by the number of articles in the EconLit bibliographic database that includes "GSS" in the abstract. No doubt, the number of articles that meet this criterion understate the impact the GSS has had on economic science, as many of the contributions using the GSS are in high impact journals and written by highly cited authors. It is plausible that many more economists would use the GSS in their research if the web-based platform for downloading data were more user-friendly. Future improvements to GSS dissemination efforts should aim for such user-friendliness, and the web-based platforms for data access could be made comparable to that for accessing IPUMS data.

## **References**

- Granger, Maury D., and Gregory N. Price. 2007. "The Tree of Science and Original Sin: Do Christian Religious Beliefs Constrain The Supply of Scientists?", *Journal of Socio-Economics*, 36: pp. 144 - 160.
- Price, Gregory N. 2005. "The Causal Effect of Participation in the American Economic Association Summer Minority Program", *Southern Economic Journal*, 72: pp. 78 - 97.
- Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thanasis Stengos. 2003. "Rankings of Academic Journals and Institutions In Economics", *Journal of the European Economic Association*, 32: pp. 644 - 666.
- Scott, Loren C. and Peter M. Mitias. 1996. Trends In Rankings of Economics Departments in the U.S.: An Update. *Economic Inquiry* 34: 378 - 400.

# Review of Web-Based Dissemination of the General Social Survey

Steven Ruggles  
University of Minnesota

The General Social Survey has evolved a decentralized model of web dissemination strategies, with widely varying websites maintained at different locations. The following sections summarize and evaluate each website in turn. I then comment on the dissemination as a whole, and propose some recommendations for future development.

## 1. NORC Websites

<http://www.norc.org/projects/General+Social+Survey.htm>

The project website at NORC provides a summary of the GSS project, contact information, and links to GSS dissemination websites. No data or documentation is distributed from this website.

<http://gss.norc.org/>

This is described as “the main GSS website.” It has no content except for a page of “Frequently Asked Questions” and links to other GSS websites.

## 2. Berkeley Website

<http://sda.berkeley.edu/archive.htm>

This is the data archive page for the Survey Documentation and Analysis (SDA) software developed and maintained by the Computer-assisted Survey Methods Program at the University of California, Berkeley. This is the top dissemination source recommended by the main GSS website, and it provides online tabulation and subsetting of the 1972-2004 cumulative file. The system is usable, but it has a lot of limitations. As the SDA documentation site acknowledges at the outset, “it does not provide full documentation of the dataset.” The documentation component of the website is not integrated with the analysis and extraction facility. The documentation is sparse and difficult to navigate; there is no system for variable search and retrieval, and no facility for adding variables to a basket for downloading. There is no convenient way to determine the available years for any particular question; essentially, finding variables that have a long run of responses is a matter of trial and error. Apparently, not all questions that were asked are available from the SDA website. The Stata command files provided with extracts require extensive massaging before they will run. There is no explanation why the data stop in 2004, since 2006 data should be available.

## 3. GSSDIRS Website

<http://www.icpsr.umich.edu/GSS/>

This is the General Social Survey Data and Information Retrieval System (GSSDIRS) website hosted by ISPSR. The design of the website is extremely dated; the front page prominently warns that it is best viewed with Netscape 4.0 or higher, a browser released over a decade ago. Nevertheless, the website is surprisingly functional. Identifying variables and their availability is far easier than on the SDA website. Users pick variables as they browse the documentation, and then can extract their selections. The documentation clearly indicates chronological availability within each variable description, although there is no easy way to compare availability for broad groups of variables. Although there is no variable search capability, there is a useful subject

classification and it is possible to locate variables without too much difficulty. GSSDIRS does not do on-line data analysis; when one clicks on the analyze button, one is sent to the SDA website. The biggest problem with GSSDIRS is that the data are almost a decade out of date. Although the front page indicates that the system covers the 1972-2000 period, the data do not actually go past 1998.

#### **4. Regular ICPSR Website**

<http://www.icpsr.umich.edu/cocoon/ICPSR/SERIES/00028.xml>

This is the ICPSR web page where you can actually retrieve the full codebook and data for the 1972-2004 cumulative file. The codebook is a 70MB PDF file with 2,390 pages. Access to the codebook or data file requires ICPSR membership and login. Obviously, this format is unwieldy and would pose a serious obstacle for many potential users of the data.

ICPSR also runs its own version of the SDA analysis and subsetting program from this site. It is branded with the ICPSR logo, runs on ICPSR servers, and has a slightly different interface than the SDA website at Berkeley. The ICPSR SDA version of the documentation has no subject classification whatsoever; thus, unless the user already knows the mnemonics of all the variables they are interested in, I believe that the ICPSR SDA version of the documentation would be effectively unusable. As in the case of the Berkeley site, access to documentation is completely separated from access to data. The ICPSR SDA analysis system could be used in conjunction with another source of documentation, such as the massive PDF codebook, but I suspect most users would prefer to simply go to the Berkeley website.

Unlike Berkeley SDA and GSSDIRS, the ICPSR SDA documentation does at least provide basic information about the samples—information that all users should be aware of—in a prominent location on the main documentation page. On the other websites, it is necessary to dig deeply for this information.

#### **5. Roper Center**

[http://www.ropercenter.uconn.edu/data\\_access/data/datasets/general\\_social\\_survey.html](http://www.ropercenter.uconn.edu/data_access/data/datasets/general_social_survey.html)

The only way to obtain the 2006 GSS at this writing is through the Roper Center. Persons at member institutions may obtain a CD by emailing the center; non-members must pay \$400. Since the University of Minnesota is not a member, I did not attempt to obtain a CD. Apparently, Roper has some sort of agreement with NORC that gives them exclusive rights to disseminate the data for a certain period of time. This practice is inappropriate because it impedes data access, and should be discontinued.

#### **Overview**

The decentralized GSS dissemination system is highly inefficient and provides suboptimal dissemination. The greatest problem is that data and metadata are replicated in several locations in widely varying formats. This is a maintenance nightmare; if an error is uncovered, it must be fixed in many different files by staff members at multiple institutions. Needless to say, it is unlikely that most errors are ever corrected. In addition, with four different dissemination pathways doing overlapping work, there is considerable duplication of effort.

Not only does this approach waste scarce resources for social science infrastructure, it also results in inferior service. Of the websites I evaluated only GSSDIRS provided an adequately user friendly—if old-fashioned—interface. Unfortunately, the data and documentation at GSSDIRS is eight years out of date, making it unusable for a majority of research projects. The only way to obtain up-to-date data is through Roper, and except for the small minority at institutions which subscribe to Roper, the cost is high.

## **Recommendations**

GSS is expensive, and the costs can only be justified if it is widely used. Accordingly, NSF should explicitly fund web-based dissemination of GSS. I have three specific recommendations for the call for proposals.

**1. GSS dissemination efforts should be centralized.** This would mean that only one version of the data and metadata would have to be maintained, greatly reducing the costs of corrections. It would avoid costly duplication of effort for software development and maintenance.

**2. A new integrated web dissemination system for data and documentation should be developed using modern software tools driven by standardized XML metadata.** Because GSS is a simple rectangular survey, such a system would be comparatively inexpensive. The chief data access challenge posed by GSS is the sheer number of survey questions, many of which appear only once or twice. Tools to cut through the clutter and select variables of interest should be the highest priority of the data access system. The data access system should allow easy browsing of variable availability and incorporate a sophisticated system for variable search and retrieval.

The current GSS data access systems present all available variables as a simple pick list, but this approach does not work well for a dataset that has so many variables. Users should be able to narrow the variable list according to keyword or subject area; reduce the list to only those variables appearing in every survey year of interest or to expand it to include all variables in any selected survey year; and view simplified pick lists focusing on the most commonly requested variables, as determined through analysis of extract logs. At any point, users should be able to select variables and add them to a data basket. When they are ready, they should then be able to view the basket and either extract data for download or carry out on-line analysis.

**3. GSS dissemination should be separated from survey design and administration.** There are two reasons for this. First, when unexpected expenses arise or budgets are reduced, survey developers are notorious for cannibalizing their dissemination budgets. From the perspective of the agency, however, such rebudgeting is highly inefficient. Second, those with the greatest expertise in survey development are not the same as those with greatest expertise in social science dissemination cyberinfrastructure. Therefore, it would make the most sense to have a separate call for proposals for the dissemination component of the project.

## **Other Ideas for GSS Cyberinfrastructure**

In addition to web-based data access tools, GSS could benefit from web-based training and user support. On-line tutorials could cover both basic issues, such as how to get GSS data into a statistical software package, and more complex analytic issues, such as variance estimation.

GSS dissemination could also benefit from the tools and technologies of the Social Web, known also as Web 2.0, which stress collaboration and sharing among users of web-based services. Such tools are built on the observation that the collective knowledge of users in a community is substantial, and if leveraged properly can benefit the entire community. GSS has thousands of experienced users. Tools and systems that allow users to support one other would mean that less individualized user support would be needed. By promoting interaction among users researching similar topics, research communities can provide intellectual support as well as purely technical assistance.

Here are some ideas for tools that would exploit the social web concept:

- **Wiki-enabled documentation** that would allow users to suggest corrections and improvements to the extensive documentation of the datasets. The user community contains many experts with deep knowledge of specific subject areas, and many are quite willing to share their expertise to help others.
- **Expert Q&A system** where users can pose specific queries. Volunteer experts can answer these questions by starting discussion threads; other users can comment on or clarify an answer, which generates better quality answers. These threads can then be archived and indexed by keywords, allowing users to search old queries before submitting a new one.
- **Specialized research forums** which can bring together smaller groups of users with detailed knowledge on a problem to share their latest developments. These forums would encourage research collaborations among scholars from diverse disciplines who otherwise might not interact.
- **Tools for sharing SAS, Stata, and SPSS code for data manipulation** developed by individual users that could also benefit others. Currently sharing among users is *ad hoc*, with no systematic match-making. A shared repository with a searchable directory would maximize the efficiency of researchers.
- **Tools for sharing curricular materials** based on the same principles as code sharing. The software developed for code sharing can be substantially reused for this purpose.
- **Expert recommendation system** for problems frequently encountered by users. The idea of this tool is to infer interests and requirements of users from their data requests and other activities, and then to recommend datasets, research forums, discussion threads from the expert Q&A, and code based on a ‘match-making’ algorithm. This approach has been very successful in many domains and has been shown to improve user experience and effectiveness.

## GSS Content and Innovations

Lynn Smith-Lovin  
Duke University

### GSS Content

As the wide-ranging usage of the GSS data in social science publications shows, it is hard to anticipate exactly what content will be useful for future scholars. However, it is fairly easy to see areas where the GSS has a “competitive advantage” over other sources of evidence. The *first* is trend studies using items that have been in the core for substantial periods of time. Therefore, I would urge that the impulse for innovation in this new grant cycle not overwhelm the extraordinary productivity generated by the stability of the survey’s design and core items. The *second* is substantive domains that can be usefully related to those core items. Since the survey has more socio-demographic information than most (especially in religion and family composition/background), individual level beliefs, attitudes, values, behaviors and other cultural traits that tend to form strong “niches” in socio-demographic space are useful to measure. Also, items that relate to the “rotating core” of items already measured (with their emphasis on race, gender, religion) are more likely to be productive. The *third* area that is exceptionally useful is the ability to generate data on extra-individual units by asking questions about those units of the respondent. This feature allows an individual-level survey to generate useful data on units like voluntary groups, congregations, families and so on. While the data are typically limited to features that can be easily observed and reported by individuals (usually compositional and behavioral features, rather than attitudinal ones), these applications have allowed the GSS to contribute to areas of social science with which it could not otherwise connect. A *fourth* domain is perhaps less obvious. Substantive domains that are theoretically linked to the social and economic composition of the geographic area in which the respondent lives can be very useful, since the survey’s sampling design facilitates linkage to Census data on the PSUs (more on this in my recommendations for innovation below).

The development of the modular format (allowed the addition of sections at the end of the survey) for adding new content while retaining the stable core was an excellent innovation in the 1990s. This format allowed continual updating of the survey content (especially with regard to domains 2-4 above) while preserving the continuity of the core items. However, the need for a “pay as you go” policy for that modular format has reduced its usefulness. To field a module, researchers must have close to “insider knowledge” about how the survey is put together, obtain funding from other sources based on the assumption that their items will fit in the survey, and accomplish this well in advance of the survey questionnaire being finalized. This funding need has prevented the survey from opportunities that arise quickly (e.g., it was not possible to do a short follow-up to the 2004 NUMGIVEN question in the 2006 survey-- funding would have had to be generated in just a month or so). It also pushes the content of the survey’s innovations toward areas where much funding is available (e.g., health) and away from theoretically core sociological domains (because the NSF Sociology panel typically reacts to GSS module proposals by thinking “we already paid for that”). Funding for module space as part of the core GSS funding is needed, and should be allocated in an open competition among scholars (e.g., the new ANES format for submitting ideas through an on-line commons web site, with openly

announced calls for proposals and deadlines) would broaden both the survey content and the visibility of the module opportunities.

## **GSS Innovations**

As I outlined in my discussion of GSS content, I think that the most useful source of GSS innovation would come from fully funded module time that could be opened to repeated competition from the larger user community. This innovation would provide a continual updating of the survey content each survey year, while preserving the core trends and maximizing the “value added” nature of the added material (because that would be a primary criterion for module selection).

Given the length of the core survey, substantive innovations should concentrate on items that can contribute a great deal while being measured in just one or two items. (A prototype item already in the survey is the single item health measure, which does a remarkably good job of assessing physical well-being with a single simple question.) We should resist the impulse to use other survey’s excellent measurement of some domains as a guide (e.g., the greater economic sophistication of the PSID or the greater depth of political content in the ANES). Indeed, these are probably domains to *avoid*, because other surveys do them more justice than the GSS ever could.

Three suggestions come to mind. One is adding a simple wealth item. While wealth could not be comprehensively measured in this type of survey, a question like “imagine that you sold everything that you own, including real estate, cars and your personal possessions—after you were done, and paid your debts, would you have money left over or not?” One could then follow with a rough assessment of how much debt or wealth would be left. While very rough, this item would supplement income measures as an indicator of economic well-being, separating those with negative net worth from those with more substantial assets. One could imagine similar simple questions that might get at “safety net” issues like health insurance, pension/retirement savings, etc. Again, the effort is not to do a serious assessment of these complex issues, but to separate the respondent population into those who have some coverage or none. This would help the survey connect with the growing literature on inequality that suggests that income is a highly incomplete indicator of financial well-being and stability.

The second suggestion is a direct analogue to the simple question on physical health, only related to mental health. Health researchers are developing such a question, and it could expand our understanding of the socio-demographic sources of well-being.

The third suggestion is actually a restoration rather than a true innovation. The voluntary association questions (MEMNUM and its components) were taken out of the core in the mid-1990s, just when interest in this type of social capital/weak ties was beginning to take off. It was asked in the network/voluntary association module in 2004 (although in a very different location). This set of items might be worth restoring in order to re-establish a trend, since it is of interest to scholars in several disciplines.

Finally, I propose that the geographic information from the relevant Census be linked and released with the survey data at the lowest level of geographic unit that privacy concerns will allow. Currently, scholars must request the linkage to create this data set (which is available for MSA/PSU levels, but requires additional funds at the county or Census tract level). If the geographic context information were released with the survey data (which would, of course, require additional funding for processing), there would be an explosion of theoretical interest in the impact of social/economic context on individual outcomes.

These innovations may sound like very minor things. They are. This is in keeping with my general belief that the GSS's primary contribution has been through its stability of content and design, together with its ready availability to a wide social science user community. If we could fund module space that would allow continual innovation, I believe that the core survey should not "chase" social science innovation, but rather allow that fashion to rotate back through the survey content. Who would have guessed 10 or 15 years ago that values would again be gaining traction in the social science literature? Who would have thought that whether or not one was willing to vote for a qualified woman for President would be so relevant in 2008? Given the time and care that it takes to field the GSS, any attempt to chase theoretical fashion will inevitably result in a survey that is speaking to yesterday's hot topics.