

## Article

Symposium 2008:  
Data Collection: Challenges, Achievements and New Directions

### **Searching for the Silver Bullet: Are EIA's Current Tools Adequate for Designing and Testing EIA Business Surveys?**

by Stanley R. Freedman

2009



## **Searching for the Silver Bullet: Are EIA's Current Tools Adequate for Designing and Testing EIA Business Surveys?**

Stanley R. Freedman<sup>1</sup>

### **Abstract**

This paper addresses the efforts of the U.S. Energy Information Administration to design, test and implement new and substantially redesigned surveys. The need to change EIA's surveys has become increasingly important, as U.S. energy industries have moved from highly regulated to deregulated business. This has substantially affected both their ability and willingness to report data. The paper focuses on how EIA has deployed current tools for designing and testing surveys and the reasons that these methods have not always yielded the desired results. It suggests some new tools and methods that we would like to try to improve the quality of our data.

Key Words: Survey design, Survey testing, Business surveys

### **1. Introduction**

There is an old-time American radio program, later turned into a 1950's television show that featured a masked Texas Ranger who galloped around the countryside on his white horse righting injustices and coming to the aid of those in trouble. The most important tool of his trade was a gun belt of silver bullets that felled his enemies with a single shot.

Much as the masked man of the old west was expected to sense trouble, cure the problems and right off victoriously into the sunset all in the space of a 30-minute theatrical production, so are survey methodologist expected to find problems with data collections, right the wrongs of years of neglect, the first time out and with little more than a few weeks before a survey goes into the field. In effect, they are searching for that silver bullet that will right all wrongs and cure all ills with one well-placed shot.

The Energy Information Administration (EIA) has an active program of questionnaire design and testing. Techniques employed include focus groups, pre-survey design visits, cognitive testing, and record keeping studies. EIA has also implemented new methods of collection such as fillable electronic spreadsheets and Internet data collection. We have had to do this in an environment of decreased budgets, staffing and the retirement of key survey personnel. The difficulties are compounded by ever changing survey frames and decreased turnaround time for fielding data collections. EIA is finding that the tools we are using have not worked as well as they once did for producing quality surveys the first time out of the gate.

The paper focuses on how EIA has deployed current tools for designing and testing surveys, and the reasons methods have not always yielded the desired results. It documents our efforts to find the "silver bullet" to improve our data quality in the face of diminishing resources and shortened time frames.

---

<sup>1</sup> Stanley R. Freedman, Director of Survey Methods, Energy Information Administration, 1000 Independence Avenue SW, Washington, DC 20585. stanley.freedman@eia.doe.gov

## **2. About EIA and the energy industry**

EIA is the United States agency responsible for collecting information about the production, transportation, distribution and consumption of energy. We conduct about 65 surveys, including weekly, monthly, quarterly, and annual collections. All but 2 or 3 surveys are mandatory collections. Our collections are characterized by populations of several hundred or thousands of respondents, rather than tens of thousands of respondents. EIA also produces forecasts of energy supply, demand, and prices. Policy makers to inform decisions, and energy analysts to explore the impact of proposed policies use our data. They are the primary source of information that moves energy markets.

For the past 10 years, EIA has undertaken an aggressive program of reexamining its data program at least once every three years. There are several reasons for this. First, the energy industry is changing very rapidly. Deregulation of petroleum, electricity and natural gas has significantly changed the way these industries operate. Changes in business operations necessitate changes in business record keeping practices that in turn impact on our ability to collect high quality data. Deregulation has also brought new actors to the energy stage. These populations need to be surveyed to provide a complete picture of energy markets. Finally, the fast pace of industry change and the recent emphasis on environmental issues have accelerated the pace of new data requirements being placed on EIA.

The remainder of this paper discusses the tools and techniques used by EIA to keep up with changing data requirements and assure that we are collecting valid and accurate data.

## **3. Tools and methods**

### **3.1 Stakeholder consultations**

EIA uses consultations for both identifying new data requirements and informing respondents of proposed survey changes. It is included in this discussion of survey design techniques because it is the oldest method being used at EIA and is used by all of our surveys. Stakeholders include industry groups, data users, and respondents. This consultation takes the form of group meetings, presentations at industry gatherings, and public notices. It is primarily a communication tool for gathering data requirements and informing respondents of what is planned. It does little to inform the process of the mechanics of forms design.

### **3.2 Respondent site visits**

In the past few years, EIA has taken another approach to working with respondents, informing them about survey changes and new data requirements. EIA staff visit respondents at their business locations to test if surveys make sense to them and conform with their business practices and recordkeeping.

One version of visits is pre-survey design visits. Pre-survey design visits are meetings with respondents or potential respondents conducted at their place of business. Visits are conducted at the very beginning of the survey design/redesign process are meant to educate EIA about the industry segment being surveyed.

These visits are designed to understand how the industry operates, what kinds of data respondents generally keep, and whether these data correspond to the concepts that EIA is interested in measuring. Respondents are asked for their input to determine what concepts are feasible to measure, how they would measure these concepts, and if EIA's survey design process is far enough along what kinds of data they keeping related to EIA's data requirements.

These visits have been very successful and yield a lot of useful information used in the forms design process. They are the best way EIA has found, so far, of communicating with respondents. They have the added benefit over other tools of providing quality "face time" with respondents, understanding their problems, and making them feel genuinely appreciated for their efforts. Of course, the number of companies EIA can reach this way is very limited and the costs in time and resources are high.

Traditional cognitive testing of surveys is another form of respondent site visit. They focus on a respondent's ability to understand concepts, answer questions, follow survey instructions, and navigate through the form. The technique can be either a concurrent think-aloud protocol or a retrospective interview, but the emphasis is on the respondent's interaction with the survey instrument. Business surveys have another important aspect to them not covered in traditional cognitive interviewing. A person is responding not for themselves, but as the representative of a company. They are a "middle-man" between the questionnaire and a company's records. As such, it is important for them to understand the correspondence between what EIA is asking, and a company's business records.

Record keeping studies are currently a popular form of respondent contact at EIA. Visiting a respondent's place of business to review plans for a new survey or major changes in an existing survey can be very useful in improving the final design of a questionnaire. These visits generally involve discussions with a potential respondent on the following topics:

- Does the respondent keep the data that the agency wants?
- How closely does the establishment's record keeping correspond to the required survey data?
- How compatible are these record keeping systems with the agency's collection instruments?
- How difficult will it be to provide the data in the time period needed by the agency?

They are most often used on existing surveys to validate data and correct reporting problems. It is not unusual for EIA to conduct between 15 and 30 of these type visits to correct reporting problems. They are important for business surveys because of the obvious reliance on record keeping practices in establishment surveys.

### **3.3 Electronic Data Collection – From paper to pixels**

The technology by which EIA interacts with its respondents has changed significantly in the past 30 years. Initially, paper copies of EIA surveys were mailed to respondents and returned to EIA the same way. Results were either tabulated by hand or more likely double keypunched on Hollerith cards and read into a central mainframe computer for processing. A lot of EIA's interaction with respondents in those days was "Where is your survey form?" and the expression "it's in the mail" became popular. It was not until weeks after a survey was submitted that a respondent ever heard from EIA about its submission.

A variation of the mail back of paper forms is the use of fax to return surveys to EIA. It is a small technological leap, but one that proved useful to respondents as they approached filing dead lines.

The first big technological change came with the appearance of desktop computers both at EIA and at its respondents' facilities. EIA began to develop software that could be installed on personal computers, and when opened, an electronic representation of a survey form would appear. The first such effort was the Petroleum Data Reporting Option (PEDRO). The respondent would fill in the form and electronically transfer the file to EIA. This system also ushered in the first application of edits performed by respondents prior to data submission. One drawback to this technology is that software must be installed on the respondent's computer. Given the varieties of hardware configurations and versions of operating systems, installing software can sometimes be hazardous. It requires a help desk at EIA to work with respondents to iron out problems when they occur. It has also become increasingly common that respondent information technology departments do not want to install "outside" software on their networked computers. It increases the opportunities for computer crashes and the introduction of viruses into their systems.

Variations of this response principle are fillable PDF files, and spreadsheets. They are created to look like printed forms. The respondent downloads them from a website, fills them in and returns them to EIA, either by email or secure file transfer. They lack respondent side edits but have the advantage of having little or no impact on a respondent's desktop.

The final technological advance to discuss here is the use of the Internet to collect survey data. In the late 1990's EIA began discussing putting surveys on the Internet using browser technology. The events of September 11, 2001 accelerated the process. When anthrax spores were discovered in the mail rooms of government offices, all correspondence coming into EIA was first sent to a facility in the Midwest to be irradiated. Survey forms came into the agency with a soft brown patina, crispy, and very, very late. In addition, companies no longer felt comfortable receiving mail from the government. On October 8, 2001, just 27 days after 9/11, EIA deployed its first survey on the Internet, the State Heating Oil and Propane Program soliciting heating oil and propane prices. Respondents logged on to an EIA server, opened up their browsers, completed the survey and securely submitted the data to EIA over the Internet. The system had a number of key features:

- Respondents were able to interrupt their sessions and return to the form without losing what they had entered.
- Respondents could print copies of their forms and save their submissions on their own desktops.
- The form took full advantage of the power of the computer for navigation and data entry.
- Context sensitive help was available to respondents.
- Respondents received confirmation of successful data submissions.
- Data were transmitted electronically and securely to EIA and eliminated the need for data keying.
- Respondent side edits were incorporated into the system along with the ability of respondents to provide electronic explanations of data that did not pass edit.

EIA was always been interested in reducing the steps involved in data capture and processing and making the whole process easier for its respondents. This system accomplished that very well. But more importantly, EIA continued to push more of the burden of editing to respondents. The Internet, coupled with the ability to query historical data bases hopefully, from EIA's point of view, would produce error free data from respondents and eliminate the need for respondent contact at all.

The respondents liked the new Internet system because it made it eliminated paper, facilitated data entry, and they didn't need special software to submit data. Most importantly, the Internet system reduced the number of phone calls they received from EIA. EIA survey managers liked the new system because it eliminated data keying, increased timeliness of respondent submission, and importantly, reduced the number of phone calls EIA had to make to survey respondents. The trend was to reduce the amount of Agency – respondent contact. The system was so popular that at this point about 50% of EIA's survey respondents are completing surveys over the Internet. EIA is currently working on an enterprise wide Internet data collection system.

## **4. Selecting the right method**

Below are the steps used at EIA to decide what methods will be used to test surveys.

### **4.1 Analyze existing respondent contact records**

For existing surveys the first step is to examine and analyze respondent contact records if they exist. Usually these records contain the reasons for callbacks and data changes that are made before a survey is closed out. If these records do not exist, interviews with survey staff are conducted to get a sense of where the problems areas are on a survey. Combine the two if time permits. Survey staff who are actually dealing with respondents frequently have the best intelligence on where problems exist.

### **4.2 Analyze the data**

This is part of the validation process, but will also yield useful information in identifying anomalies in reporting, strange cyclical patterns, and other peculiarities that may indicate problems with reported data. Look for unexpected changes in time series, and departures from anticipated results.

### **4.3 Review edit failures and corrections to these failures**

Are the same items on your survey failing edit each survey cycle? Are there particular respondents or classes of respondents that are failing edit and require callbacks time and time again?

### **4.4 Has the industry you are surveying recently changed?**

Changes in an industry are often followed by changes in record keeping practices, changes in company structure or changes in business practices themselves. All of these will lead to misreporting and inaccurate data.

### **4.5 Listen to your data users**

The best way to assess whether your surveys are doing a good job of measuring intended concepts is to solicit feedback from your data users. They are the ones conducting the complex analyses and producing forecasts. They are in an excellent position to point out problems with the data, and are rarely reluctant to do so. These problems usually have their roots in survey design issues.

## **5. Case studies**

EIA is still in the early stages of research for categorizing which methods work best for which types of survey problems. The interrelationships are complex and rarely straightforward. We are not at the point where we have even preliminary generalizations. Rather, I will provide some illustrative case studies.

### **5.1 Case study #1 – “I take great store in what I reported last time”**

One of EIA’s older natural gas surveys had been designed in the period prior to the deregulation of that industry. The survey remained unchanged. Survey managers and data users began to see an increased disparity between supply and disposition balances. All the usual suspects were checked including the processing systems and edit routines. No light was shed on the problem. The survey methodologists were asked to look at the issue and decided to see if the change from a highly regulated industry to a deregulated one were causing any misreporting. After a series of site visits at which sections of the form were cognitively tested, we discovered a substantial proportion of the respondents were continuing to report commodities on an equity basis rather than on a custody basis. As is often the case, respondents used previous submissions to guide their reporting and were ignoring changes to instructions and section titles on the form. We probed further and found that their record keeping practices would support the change from equity to custody if we could get them to recognize what they were supposed to be reporting.

We chose cognitive interviewing as our research method because we believed that the respondents did not notice or understand the changed reporting requirements.

EIA was able to fix this particular reporting problem by modifying the form and instructions to more clearly reflect the changed reporting requirement. More importantly, we sent email notifications to all the survey respondents informing them of the changed requirements.

### **5.2 Case study #2 – “Measuring the un-measurable”**

About six years ago, EIA was asked to report on a new concept in the electricity area. Manufacturers that use heat or steam in their production process has begun to use some of that excess heat and steam to spin turbines and produce electricity before the steam was put back into the production process. This meant that some manufacturing facilities were now also electric power producers. They used the electricity within their own facilities and in some instances actually sold the power back to the electricity grid. EIA was interested in the amount of electricity that was generated, and how much fuel was used to generate electricity vs. fuel used in the manufacturing process. EIA

met with industry groups that assured the agency that these manufacturers could differentiate the fuel used for electricity production and even provided EIA with some language and terminology for collecting the information.

When the survey was fielded, many respondents did not understand what EIA was asking for and those that did could not report the data. EIA survey methodologists were again called in to help. We spoke to about 15 companies on the phone and realized that this was a problem not just of understanding, but also of record keeping and plant operations. We scheduled over two-dozen site visits with a large variety of manufactures. We focused on 3 areas:

- Did the companies understand the concept we were trying to measure
- Did they measure this concept at their facilities, and if they did
- Did they keep this data in their record keeping systems?

What we discovered was the companies did not understand the terminology we were using, but once we explained what we wanted to collect and why, they understood the concept. They did not measure fuel for heat and steam and fuel for electricity because it was not physically possible. All the fuel entered the plant through one input valve. They did not measure the excess steam being used to generate electricity since it was to them waste heat now being put to productive use, but did not change the amount of fuel they consumed for manufacturing. Nor were they particularly interested in trying to measure it.

Our first approach was to try to get the companies to produce an engineering estimate for EIA, but this did not prove feasible to have all the respondents use the same method, since all the manufacturing facilities were different. We finally settle on having the companies provide some basic data and having EIA produce aggregate estimates based on standard engineering principles.

This was a very complex measurement problem that could not be adequately addressed with cognitive studies alone. Cognitive studies alone would have misled EIA into thinking we could solve the problem with better question wording or more complete instructions. We needed to combine the cognitive studies with record keeping studies and reviews of plant operations before recommending an acceptable method.

## **6. Developing criteria**

As mentioned above, EIA is still in the early stages of categorizing which methods work best for which types of survey problems. The interrelationships are complex. We are beginning to see that simple cognitive methods work best when:

- Industries are stable
- Concepts are directly measured at the plant or facility
- These measurements are directly transferred to a company record keeping system.

More complex methods, particularly record keeping studies, combined with cognitive studies, need to be employed when:

- There are major shifts in industries or business practices
- Concepts are not directly measured with meters, gages or valves
- Establishments must estimate, adjust or otherwise alter their basic record keeping to answer survey questions.