

SAMPLING ISSUES, SURVEY ADMINISTRATION, AND DATA ANALYSIS

The In-Person Interviews and the Frontline Survey were administered in late 1998 and early 1999, following my return to academia, about five years into the procurement reform effort, to procurement employees in 19 local buying offices.¹ In selecting offices, I chose, first of all, ones where contracting was a relatively important activity, in terms of dollars spent, absolutely or as a proportion of the agency's budget, and/or of the importance of contracting to the agency's mission. Thus, for example, procurement is important along both these dimensions for the Defense Department but not for the Department of Justice. Choosing organizations for which contracting was important also was likely to produce ones with a relatively large number of procurement employees, which reduced administrative costs of conducting the research. In addition, organizations were selected to provide a good mix of defense and civilian organizations (and within the Department of Defense of organizations buying weapons systems and commercial items), a mix of Washington and non-Washington based offices, and a range in the degree to which agency-level procurement leadership supported reform. Very much at the margin (for only three of the 19 offices), an office might have been selected over similarly situated other ones because I had a personal association with the local office head, which made it easier to gain cooperation; this effect should be seen as insignificant, however, because at senior levels I knew just about all the relevant officials, and at the local offices, there were a number of office heads whom I didn't know personally.

To gain buying office participation, I first approached the senior agency procurement official at the agency to gain permission to conduct the research, including the Frontline Survey. All the officials approached agreed, though in some cases they made it clear a local office head would need to give approval to cooperate at their office. All local office heads contacted did so.

This method of selecting respondents creates a danger of bias. Although in choosing offices to participate, efforts were made to include a range of organizations along several dimensions, this was a kind of convenience sample rather than a truly random one.

The sample might instead have been selected in two other ways.

(1) One possibility would have been to draw a random sample of all individual frontline procurement employees in the government. There is a computerized list of all federal employees available through the Central Personnel Data File administered by the Office of Personnel Management. It is possible to get a sort by job classification (1102 and 1105 are the classifications for procurement people), with agency affiliation (at the sub-cabinet level, so there would be information not just for the Department of Agriculture but also for the Soil Conservation Service within the Department of Agricultural) and workplace addresses. This idea was easy to reject for a number of reasons. A random sample of individuals would not have permitted any workgroup-level or organizational leadership-level analysis except through self-reports, since there would typically only be one person (or at most two or three) from any given organization, and there would have been too many organizations to allow the In-Person Interviews used to provide data for the Frontline Survey about local leader attitudes. Indeed, in such a random sample, a relatively large proportion of the respondents would have worked in extremely small offices of anywhere from one to a handful of procurement employees, such as local offices of decentralized agencies such as the Fish and Wildlife Service or the Soil Conservation Service (although exactly what percentage would be in such small office is unknown, since the government compiles no statistics of the percentage of employees by size of workplace). Sending out mail questionnaires to individual respondents would have produced dramatically lower response rates than were obtained by the method used, especially for a survey of the length used.

(2) A more promising possibility might have been to draw a stratified random sample of organizations to participate in the survey, where an organization's probability of being chosen for the survey would depend on its size relative to the size of the entire procurement workforce. By choosing organizations to participate and not individuals, the possibility of looking at organization/workgroup influences would have been retained. And the aim would have been to have the surveys conducted at the workplace, so as to deal with response rate concerns. However, this method would have been difficult to implement and would have created a number of problems. First, Office of Personnel Management data can be sorted only by agency and "duty station" (which might be a city, county, or state, depending on how many people work for the organization and how the individual agency chooses to code employees). In many cases, it would have been difficult to distinguish among contracting organizations performing very different functions at one location, particularly for large organizations such as military bases (where there are different contracting operations, each with its own leadership, buying weapons, base services such as grasscutting, research and development, IT, etc.). If an organization had several locations in one "duty station" (most likely for agency offices in Washington or where the duty station was listed as a state, or possibly in some cases a county), it would be impossible to distinguish among them. Second, for the stratified random sample of organizations itself to have been statistically more acceptable (which would be the purpose of this exercise in the first place), a significant number of organizations would have had to be chosen. To achieve the cooperation of these many organizations, it would have been necessary to work with every cabinet department to gain top-level approval, followed by efforts to get agreement at lower levels from a much larger number of local leaders. This probably would have produced refusals to participate, as well as much greater uncertainty, due to lack of personal connections, about

whether survey instructions would have been followed and an acceptable response rate achieved. Thus, what might probably have been minor gains in representativeness would likely have been overwhelmed by dramatically increased administrative costs to organize the research, combined with refusals to participate and lower response rates. (Bewley 1999: 16 also discusses the tradeoff between sample randomness and data quality in his interviews with employers about why they didn't cut wages during a recession; he also decided to sacrifice – considerably more than I did – randomness for quality.) And a random sample of any sort would have still produced a significant number of respondents working in small organizations and in agencies where procurement was not an important part of the agency's activities.

It would appear fair to say that the convenience sample selected is a good sample of procurement employees in large organizations where procurement was rather significant to agency missions. The only two organizations in the government fitting those specifications but not represented are NASA and the Department of Energy (also the Navy among the military services). Based on my knowledge, there is no reason to believe responses would have been different in those organizations. One may reasonably argue that large procurement organizations in agencies where procurement is important dominate any impact that reform will have on government contracting, so that this is a good universe to investigate. However, it should be noted that such offices may differ in systematic ways from small ones and/or ones in agencies where procurement is not important. Large offices, particularly in agencies where procurement is important, are more likely to have an infrastructure for communicating material from top agency and/or systemwide leadership. They are also more likely to receive personal visits from senior leadership from outside the office. This may mean that results from these organizations are biased towards exaggerating the influence of senior agency or systemwide leadership on

individuals on the front lines. (It might be the case as well that in small offices, one's immediate supervisor is more influential than are those in larger offices.) Also, organizations where contracting is an essential function use fewer less-trained purchasing agents, as opposed to more-trained contract specialists. (Purchasing agents were 14.8% of the workforce in 1997, but only 4.8% of the weighted sample.) (Federal Acquisition Institute 1998: 3) So results presented in this research should not be taken as representing small procurement offices and agencies where procurement is unimportant.

The In-Person Interviews

The In-Person Interviews were structured interviews, with pre-established questions, conducted "on location" at each buying office. Those interviewed were the local office head (and often the deputy), as well as some division chiefs (direct reports to the local office head), first-line supervisors, and non-supervisory employees. I conducted some of the interviews, and others were conducted by a research assistant who was an active-duty Army officer. Among respondents corresponding to job categories in the Frontline Survey (non-supervisors and first-line supervisors), supervisors were overrepresented in the In-Person Interviews; 35% of respondents were supervisors, compared to 20% in the Frontline Survey.

Most, though not all, the interview questions were "open ended" ones to which the interviewee formulated a response. In a few cases, questions gave specific response categories. (Local leaders were asked the PRVIEW question, and some questions gave response alternatives such as "to a great extent," "sometime," and "not at all.") The interviews began by asking the interviewee to discuss his or her history and career path in government procurement, which was in significant measure done to "break the ice" and allow the respondent to speak about an "easy" topic. The interview proceeded with various questions about the history of procurement reform

at the office and about the respondent's personal reactions to the reform process, with special emphasis on how they had changed during the five years of reform. Most, but not all, questions were the same for respondents at different levels of the organization. In many cases, for reasons of time (interviews were scheduled one right after another, and some respondents spoke at great length) not all questions were posed.

In coding open-ended responses into categories, I did not follow standard procedure, used in research that is based mostly on analyzing interview data, of having two coders independently place answers into categories and then checking on "inter-rater reliability" in the coding. In my view, candor and spontaneity in interviews are significantly affected if the interview is taped, particularly when interviewees are not used to being interviewed. Therefore, notes on all the interviews were handwritten, which made the interview data (at least in the case of my handwriting) illegible to a third party. Since virtually all the quantitative data analysis was to come from the Frontline Survey, it didn't seem worth it to lose candor and spontaneity in order to allow multiple rater coding. Generally, I believe coding was relatively straightforward; in one instance (in Chapter Four, pp. xxx) using coded data from the In-Person Interviews, I am skeptical about the reliability of the categories, and I note this in a footnote to the text.

In choosing respondents for the In-Person Survey, I asked local office heads to provide a random group, say by taking the appropriate number of individuals from an alphabetical list and giving me those with last names starting with letters at the beginning of the alphabet. However, I was concerned these instructions were not typically followed and that local office heads in many or most cases tended to choose predominantly respondents who were sympathetic to themselves and/or to procurement reform (perhaps to please me). Examination of the pattern of responses to certain questions in the In-Person Interviews confirmed this fear. One of the interview questions

was, “Do you recall when you first became aware that the new Administration in Washington was making a push for acquisition reform?” The follow-up question was, “Do you recall what your initial reaction was.” For non-supervisors, 52% of respondents answered that their initial reaction to reinventing government was positive (coded as “positive” or “positive, but thought it would be hard to achieve”). This compares with only 21% of non-supervisory respondents² answering “strongly agree” or “agree somewhat” to a statement on the Frontline Survey: “My initial reaction to the National Performance Review was very enthusiastic.” Further supporting the view that respondents were disproportionately selected from among supporters of reform was that this disproportionate initial positive reaction was much less among supervisors. Since I specified units within each office where I wanted to interview people, the local office head had less ability to select supervisors for the interviews, and in smaller offices, most or all first-line supervisors were interviewed.³

Finally, I was concerned that people whom I interviewed myself would be hesitant to express criticisms of procurement reform to my face. I therefore compared answers to questions about whether the respondent before 1993 would have regarded the traditional system as “broken” and about initial reactions to the reinventing government effort for respondents interviewed by me and by my research assistant. Results appear in Table One, divided into first-line supervisors and non-supervisors. The differences were small, although they were present, and in the predicted direction, limited to non-supervisors (for whom the status difference with me was greater than for supervisors).

It should be noted that I seldom use results from the In-Person Interviews in ways where the overrepresentation of real, or apparent, reform supporters, or of supervisors, is relevant.

Where it is, or might be, I note this in the text; in one situation (see footnote xxx in Chapter Three), I weighted the raw data.

The In-Person Interviews were conducted prior to completion of the questions being used in the Frontline Survey, so that insights gathered from those interviews could (and did) influence the content of questions in that survey.

The Frontline Survey

The Frontline Survey was a fixed-format written questionnaire. It began with questions about the extent and nature of the respondent's government and procurement experience (such as what year they began working in contracting for the government) and with demographic questions. The PRIMPACT question was followed by several feeling thermometer questions, starting with the PR question. Then respondents were asked about attitudes towards reform of coworkers, supervisors, and other leaders. They were then asked to rate a list of 19 reforms (ranging from "the credit card" to "increasing the use of past performance in source selection and contract administration") on a 1-7 scale in terms of how central they were to reform, how far their own organization had come in implementing them, and whether they personally thought they were a good idea. Most of the rest of the survey consisted of fixed-format questions using a five-point Likert-type scale, asking the respondent's opinion (ranging from "agree strongly" to "disagree strongly") on various statements. In all, the Frontline Survey was very long, consisting of several hundred questions and taking 60-90 minutes to complete.

The survey sample consisted of all non-supervisory employees in each of the 19 buying offices, along with first-line supervisors. The survey was administered after visits to these offices, which meant I had developed a personal connection with each local office head and with some supervisors. Copies of the questionnaire were mailed to each local office head, along with

instructions for administering the survey, which was to occur at the office. Instructions called for administering the survey, if possible, at an “all hands” meeting, so maximum participation could be obtained. Both the instructions on each questionnaire, and the text of instructions to be read to participants by the person conducting the survey, emphasized both respondent anonymity and the point that there were no right or wrong answers to survey questions.⁴ Respondents did not put their names on the surveys, nor did surveys contain any identifying numbers (since, unlike a mail questionnaire, no repeat mailings to non-respondents were necessary). Instructions directed respondents to put their surveys into a manila envelope provided them.

The questionnaire was professionally printed, with a two-color cover (crimson red and tan) including the Harvard seal, the identification, “HARVARD UNIVERSITY, John F. Kennedy School of Government,” and my name and academic title.

The instructions provided appear to have been generally followed, but not always. In some organizations, surveys were distributed individually to respondents, and they were asked to return them to a supervisor. For all the offices, however, individual questionnaires were returned to me in the sealed manila envelopes. Two of the buying offices, which had common leadership and each of which had a very large number of procurement employees, established an overall quota that half their employees would participate in the survey, and then required supervisors to produce an appropriate number of volunteers. Union locals at two of the offices asked that employee participation be voluntary. At one of those two offices, management made no real attempt to encourage participation, and the response rate was extremely low, under 10%. The other did encourage participation, producing a somewhat higher, but still modest, response rate. For other offices, response rates were close to 100%, allowing for sickness or other absences

when the survey was conducted. 88% of respondents completed the entire survey, no small feat given its length.

This method of administering the survey was chosen to make it feasible to obtain response rates dramatically in excess of mail questionnaires. Furthermore, it would have been impossible to undertake a survey of anywhere near this length by mail, so this method allowed collection of an extremely rich dataset.

Having the survey originated by a researcher who was widely known as himself the leader, while in government, of the reform effort, as well as having the survey administered at the respondent's office (since just about all local office heads were, to one degree or another, supportive of reform), created the risk of a "political correctness" bias towards answers supportive of reform. Steps were taken to counteract this, through anonymity, use of the sealed manila envelope, and the academic connection.

It is plausible to believe "political correctness" may have been something of a problem in responses, and that mean responses on the PR and PRIMPACT variables were therefore somewhat high. Some empirical test of this worry is possible. There was variance in the sample in the extent to which respondents' immediate supervisors and office heads supported reform. If the responses to these questions reflected answers the respondent believed his or her bosses wanted rather their real beliefs, this would be reflected in exaggerated correlations between the respondent's self-reported attitude towards reform and the attitudes towards reform of his or her immediate supervisor and/or office head. In fact, data analysis of determinants of attitudinal support for reform shows small, or not even statistically significant, effects of leader attitudes on the attitudes of frontline employees. It is hard to reconcile these results with the conclusion that there was a significant upward bias in the score for PR caused by a "political correctness" bias,

especially since there are reasons to hypothesize that local leader attitudes might genuinely influence the attitudes of respondents. This does not exclude the possibility that some respondents might have given more positive responses to please the (absent) researcher, but it is hard to develop a story for why such a phenomenon would be important if seeking to please actual bosses wasn't. Also, the only losses any "political correctness" bias would be expected to produce would be to bias upwardly mean responses to those variables, not to affect regression coefficients. In all, the benefits of access my previous government work provided surely outweighed problems any "political correctness" bias might have created.

In addition, there were special problems involving three of the 19 offices. At the two where supervisors were asked to produce a quota of volunteers, reform supporters may have been overrepresented among both volunteers (self-selected) and "volunteers" (selected by supervisors), the former because supporters might have been more eager to participate in a survey being conducted by a leader of the reform effort and the latter because most supervisors were themselves pro-reform. Also, one of the two offices where virtually no employees responded to the survey (because of union insistence participation be voluntary combined with lack of management encouragement to participate) was the agency selected specifically because its senior procurement official was relatively skeptical towards reform, in order to have a significant number of responses in the survey from frontline employees in such an organization.

Regarding the organization with a skeptical senior procurement official and few responses, it should first be noted that any bias this absence introduces cannot be enormous because, in addition to this one official, there was only one other senior agency procurement official in the government at the time of the survey who was as skeptical about reform (and that official served in a department where procurement was unimportant in terms of number of

employees, dollars spent, and mission significance), and one other (also at a minor cabinet department in terms of procurement importance) almost, though not quite, as skeptical. The absence of this office, out of the 19, in results reported for most of the regression models (there were too few respondents to make it a good idea to include respondents in analysis involving organization or workgroup-level variables, so these few respondents were excluded) does mean one cannot test for the possible influence of having an organization-level leader more hostile to reform than any of the other senior leaders in the offices in the survey. And, if having such a leader had an effect on frontline attitudes towards reform or reform-oriented behaviors, absence of this office from the results would bias mean values for these variables somewhat. Similarly, in the two organizations with quotas/"volunteers," there might have been a slight upward bias in mean scores for PR and PRIMPACT. In neither case would one expect an effect on coefficients in the regression models.

The sample, compared to the universe of procurement employees, considerably overweights the Defense Department. Of the procurement workforce population, 26% work on Defense Department weapons systems, 42% work in the Defense Department doing other than weapons systems, and 32% work for civilian agencies.⁵ Of the sample, by contrast, 72% worked on Defense Department weapons systems, 16% in the Defense Department doing other things, and 12% for civilian agencies. Partly because Defense Department procurement employees are more highly educated and more male than civilian ones, the sample was also more educated and male than the population.⁶ (57.5% of the population has a college degree or higher, compared with 75.3% of the sample; 40.1% of the population is male, compared with 67.3% of the sample.)

This bias created two potential problems. First, Defense Department procurement employees on the whole were less sympathetic to reform than civilian ones (the average score on PR was 67.0 for Defense Department and 73.2 for civilian employees). They also thought it had influenced the way they did their jobs somewhat less (1.82 mean score on PRIMPACK versus 1.77 for civilian employees). The unweighted sample thus would give less accurate mean values for a number of relevant variables. Second, it is at least plausible that relationships between some predictor variables and dependent variables might differ, between Defense and civilian employees. Thus, for example, the relationship between education and other variables might differ in an environment where most people had college degrees versus one where fewer did. And there might also be differences between Defense Department employees working on weapons systems and all other employees. People working on weapons systems work in a unique procurement environment, dealing with large contractors in a concentrated industry selling only to the government, with long-term contracts (so that for many years an employee might be working on a single contract with a single contractor), and with often-strained relationships with industry. This situation might plausibly affect all sorts of relationships between predictor and dependent variables, such as, hypothetically, the relationship between the extent one thinks that reform had made it easier to choose best value contractors and one's attitude towards reform (since many of these employees are seldom involved in choosing new contractors), the relationship between one's views of contractors and one's attitudes regarding reform (attitudes towards industry might be more salient for people working with defense contractors than for others in causing other attitudes), or the relationship between coworker's attitudes and one's own attitude or behavior (since people in these offices are more likely to work in longterm teams with the same fellow-employees). If there are different relationships

between predictor and dependent variables among Defense and civilian employees, an unweighted sample would create coefficients that were biased estimates of the entire population.

For these reasons, the sample was weighted to correspond to percentages of Defense weapons systems, Defense other, and civilian procurement employees in the population.⁷ As a result, the percentage of the sample with a college degree (62.2%) and percentage male (54.3%) came closer to the population as well. All data analysis is for the weighted sample.

¹ The local offices and agencies of which they were a part, in addition to participating in the In-Person Interviews and the Frontline Survey, were asked to provide various statistics about themselves, such as on employee downsizing.

² Unweighted sample, since the In-Person Interview sample is similarly unweighted.

³ The result is frequency distributions that are the opposite of those from the Frontline Survey, which showed first-line supervisors to be more pro-reform than frontline respondents.

⁴ The initial instructions on the survey form stated: “Almost all the questions...ask for your opinions about various topics. Most of the questions ask your opinions about acquisition-related issues. Others are questions about your job in general. And some of the questions ask about your likes and dislikes as a person, not related to acquisition or to your job. This is not a test! There are no right or wrong answers. ... You should not put your name on either the surveys or the envelope.” (emphases in original).

⁵ These statistics were developed using several sources. Overall government numbers, including numbers by agency (which allowed a distinction between Defense and non-Defense), along with numbers on education levels and percentage male (as of September 30, 1997), came from Federal Acquisition Institute (1998). Numbers for Defense Department procurement employees working on weapons systems (based on the employee’s command (as of March 1998) were provided by Alex Sinaiko of the Defense Manpower Data Center. Excluded from Defense Department numbers were procurement employees working for the Defense Contract Management Command (since renamed the Defense Contract Management Agency), a unit in the Defense Department, which is engaged only in contract administration, involving specialized functions and not represented in the sample. Information on the Defense Contract Management Command, including education levels and gender distribution, was provided by Scott Blank. Governmentwide gender and education percentages were calculated excluding Defense Contract Management Command employees.

⁶ Those at larger procurement offices, and in organizations where procurement is more important for the agency’s mission, are also likely to be more highly educated than those at smaller ones.

⁷ This method is not without problems. If one sees this sample as being a sample of employees in relatively large procurement organizations in agencies where contracting was important to how the agency does business, then the appropriate weighting might involve the population of procurement employees in such circumstances. It is likely, for example, that procurement employees working on weapons systems constitute a larger percentage of procurement employees in “large organizations where contracting is important to the organization” than of all procurement employees. However, even if one allowed subjective judgments, for the purposes of weighting, which

agencies had contracting as an important part of their mission, there are no data on the size distribution of procurement offices either by agency or even governmentwide. It should be noted, however, that weighting the sample does not dramatically change any regression results.