Building an Alternative Response Process Model for Business Surveys

Diane K. Willimack and Elizabeth Nichols, U.S. Census Bureau[1]
Diane K. Willimack, U.S. Census Bureau, Washington, D.C. 20233

Key Words: Cognitive response model, business surveys, establishments, record-keeping

Introduction

The cognitive response model for survey response, usually attributed to Tourangeau (1984), generated an entire field of research to study and reduce measurement error by evaluating and improving survey questions. While the bulk of this research has been undertaken and reported relative to household and general population surveys, this model has also taken hold in establishment surveys. However, characteristics of establishments as respondents have challenged the direct applicability of the model, and has resulted in some modifications (Edwards and Cantor, 1991).

This paper suggests further modifications to the basic four step cognitive model in the context of establishment surveys. Although additional steps were suggested by our research on the survey response process in large multi-unit companies (Sudman et al., 2000), we find that reinterpretation of the literature lends additional support to our proposed model. This paper will review various enhancements to the basic cognitive model, place them in the context of establishments, and provide supporting arguments for our proposed model. We will also raise several unresolved or seemingly conflicting issues that remain.

The Basic Cognitive Model

Tourangeau's original cognitive response model consists of the following four steps:

1. Comprehension: Understanding the meaning of the question.
2. Retrieval: Gathering relevant information, usually from memory.
3. Judgment: Assessing the adequacy of retrieved information relative to the meaning of the question.
4. Communication: Reporting the response to the question, e.g., selecting the response category, editing the response for desirability, etc.

Adding the Encoding Step

Eisenhower et al. (1991) precede the basic four steps with the encoding process, referring to the "knowledge to answer survey questions." They claim that the potential for measurement error begins with encoding, that is, how knowledge and memories are stored and utilized. They suggest that encoding contributes to measurement error in surveys in two ways:

- Memory formation: The manner in which memories are formed affects their retrieval.
- Proxy response: Memories concerning others are likely stored differently from memories concerning oneself.

Edwards and Cantor's (1991) Cognitive Response Process in Establishment Surveys

Edwards and Cantor (1991) adopt this five-step model (Tourangeau's four core steps plus encoding) for establishment surveys, making a few enhancements. Their modifications appear in ***bold italics***:

1. Encoding in memory / ***Record formation***
2. Comprehension
3. ***Source decision: memory or records***
4. Retrieval / ***Record look-up***
5. Judgment
6. Communication

As the establishment analogue to "encoding in memory," record formation is explicitly added to the first step of the response model. With two potential sources for answering survey questions – respondent's memory or establishment records – Edwards and Cantor (1991) add the source decision step to their model and modify the cognitive retrieval step to include record look-up.

Sudman et al. (2000) Hybrid Response Process Model for Establishment Surveys

Based on exploratory research on statistical reporting processes used by thirty large multi-unit companies, Sudman et al. (2000) suggeste the following hybrid response process model for establishment surveys:

1. Encoding of information in company records.
2. Selection and identification of the respondent or respondents.

[1]This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review by the Census Bureau than its official publications. This report is released to inform interested parties and to encourage discussion.

3. Assessment of priorities.
4. Comprehension of the data request.
5. Retrieval of relevant information from existing company records.
6. Judgment of the adequacy of the response.
7. Communication of the response.
8. Release of the data.

While development of this model relied on inductive reasoning based on the findings, we now offer some conceptual bases for the additional steps and for redefining some of the basic steps in the context of establishments.

Encoding Revisited

Edwards and Cantor (1991) suggest a dichotomy – that information to answer survey questions is either in the respondent's memory or in business records. Thus, the respondent must choose between memory and records regarding the source of information used to answer a given survey question – hence, the addition of Step 3 in their model presented earlier.

The source decision varies with the type of question – whether the question requests figures (e.g., dollars, quantities), offers response categories, or uses an attitude scale to codify opinions. Even the Census Bureau asks some questions that do not require figures as answers, instead requesting categorical or Yes/No answers, such as the type of jurisdiction within which the establishment is located (e.g., city, town, township, etc.), kind of business, class of customer, or method of selling.

While it may be true that the respondent has encoded in memory basic figures that also appear in records, such as employment, payroll and revenues, it is even more pertinent that an establishment respondent's "knowledge to answer survey questions" also includes knowledge of the business' various records and information systems, along with company structure. Particularly in large companies, divisions of labor and decentralized data sources distribute knowledge across company units, requiring data may to be assembled to satisfy survey requests (Groves et al., 1997; Tomaskovic-Devey et al., 1994). Thus, knowledge of multiple data sources and locations must also be encoded in the respondent's memory. To the extent that the respondent does not possess this knowledge, or that the respondent's knowledge is flawed or incomplete, there is potential for measurement error.

In summary, there are two types of knowledge encoded in memory important to the response process in business surveys:

- Personal knowledge from which a question may be answered directly.
- Knowledge of records from which the answer to a question can be retrieved.

Thus, regardless of whether a particular question can be answered from business records, what and how information is encoded in the respondent's memory is of utmost importance to measurement error.

Record Formation

Let us now turn attention to record formation. The end product of record formation is recorded data. What data appear in business records? Our previous research on the reporting process in large companies (Sudman et al., 2000), along with previous record-keeping studies (U.S. Bureau of the Census, 1990), found that data recorded in company records are primarily driven by three things:

- Management: Data are kept to manage the business, to ensure the goals of the business – e.g., to make a profit – are met. Items tracked vary depending on data needed to monitor activities for particular units or levels of the company. For example, many companies in our research indicated that more detailed data were tracked by individual business units, levels or locations, while the corporate financial office was responsible for tracking and reporting aggregate or consolidated figures.
- Regulations: Data are kept to meet regulatory requirements, such as those imposed by the Internal Revenue Service (IRS), the Securities Exchange Commission (SEC), the Federal Trace Commission (FTC), workers compensation, or unemployment insurance, among others.
- Standards: Recording of figures are guided by Generally Accepted Accounting Principles (GAAP).

Furthermore, large companies utilize multiple information systems. Most, if not all, are automated, although they may or may not be linked. For example, one system may handle payroll information, while another monitors inventories.

Which data are actually recorded in business records clearly impacts the availability of data to meet a survey request. Besides our large company research, a number of research papers discuss data availability, often discerning whether data requested in surveys exist in company records. Examples include Carlson et al. (1993), Eisenberg and McDonald (1988), Ponikowski and Meily (1989), Sykes (1997), U. S. Bureau of the Census (1990), and Utter and Rosen (1983).

When new items are proposed for surveys at the Census Bureau, a small sample of businesses (or their trade associations) are contacted to learn whether these items exist in records or if there are sufficient data from which requested items can be estimated. This is research

into the results of record formation, rather than the retrieval step.

Furthermore, if data collectors could impact record formation – that is, if we could influence businesses to record the data that we will ultimately be collecting for statistical purposes – then retrieval of the data would be eased. There are a few rare instances of this – e.g., influencing payroll software developers to incorporate data requested by the Covered Employment and Wages (ES-202) program conducted by the Bureau of Labor Statistics (Searson, 2001).

Record Formation, Retrieval and Data Availability

However, record formation is only one dimension of data availability. Others include the respondent's access to data sources and the respondent's ability to retrieve data from these sources. Thus, availability may be conceived to be a function of record formation, the respondent's access to records, and retrieval.

Retrieval incorporates three components:

- the cognitive act of retrieving from memory knowledge of data sources, company records, information systems, and company structure;
- access to appropriate records; and
- the physical act of retrieving data from records and/or information systems, which includes
  - extracting information from computer and paper files,
  - consulting multiple sources, both people and records, because of distributed knowledge, and
  - compiling information.

As a result, the function defining data availability suggests particular inferences and conclusions. For example, if data exist in records and the respondent knows this, but the respondent does not have access, then the data are not available. If data exist in records and the respondent does not know this, then access is irrelevant, and the data are not available. Thus, data availability requires both the respondent's knowledge of and access to records. Moreover, access is a necessary but not a sufficient condition for retrieval, and thus data availability.

Because knowledge and access vary with the respondent, data availability varies with the respondent. This adds another dimension of the retrieval step that is subject to respondent variation, resulting in potential measurement error.

As we noted earlier, our reconsideration of the literature suggests that much of this previous research has addressed the retrieval step by focusing on data availability in terms of the results of record formation. To a lesser degree, if at all, previous research has considered the attributes of the respondent's a) source knowledge, b) records access, and c) retrieval activities.

In our original research on large companies, we too interpreted retrieval primarily in terms of availability. Nevertheless, we did identify the following information retrieval strategies. Nearly all reporters kept documentation related to government reporting, along with supporting "work papers" or accounting schedules documenting calculations of items and/or indicating data sources. For repeated periodic surveys, all respondents used this documentation of previous period reporting as a guide for completing subsequent forms. This strategy has the advantage that changes from one period to another reflect real changes in the business and not changes in question interpretation. However, a disadvantage is that any previous errors are perpetuated.

In addition, we discovered two common completion strategies related to the role of the respondent. Sometimes the respondent coordinated survey response, by distributing the report form(s) to local data providers, that is, to staff at other levels or units in the technical core of the company. In some cases, these local data providers completed and mailed the form directly back to the Census Bureau, with little intervention from the coordinator; in other cases completed reports were returned to the coordinator for corporate-level review. In an alternative strategy, the respondent compiled the necessary data for the request, gathering data from multiple sources and/or requesting data from local data providers. Copies of the form and instructions were sent to the local providers, or data needs were interpreted by the compiler and communicated via telephone or e-mail to staff with access to the data. The compiler then completed the form(s) with the forwarded data. In either case, involving local data providers results in a second, and perhaps embedded, round of cognitive response processes, with consequences for measurement error.

Respondent Selection / Identification

Clearly the identity of the respondent has implications for the efficacy of the four core cognitive steps and, subsequently, for measurement error. Different respondents for the same company may –

- have differing knowledge of available records encoded in memory;
- comprehend questions differently;
- have varying degrees of access to different records as well as varying abilities to retrieve data from them;
- make different judgments regarding the adequacy of the information retrieved; and
- communicate the response in different manners.

Edwards and Cantor (1991) suggest measurement error due to the respondent may be minimized by

selecting the person most knowledgeable of the requested data to be the respondent. Thus, the desired respondent would be the person closest to the record formation process, thus having knowledge of both contents of the records as well as understanding of the concepts being measured by recorded data. Tomaskovic-Devey et al. (1994) call this the "technical core." Indeed, selecting such a person for the respondent should reduce variation in the cognitive steps.

However, due to distributed knowledge, especially in large companies, a single respondent may not be the most knowledgeable of all the requested data items for surveys that consist of multiple data items. Multiple "most knowledgeable respondents" may be needed. Who then should be selected as the respondent? Is there one respondent with sufficient knowledge to answer multiple survey questions? Because of organizational hierarchies (Tomaskovic-Devey et al., 1994), it is more likely that such a person may be knowledgeable of the existence of these data, but not have intimate knowledge of the figures themselves, nor have the ability to directly retrieve these figures. As a result, the single most knowledgeable respondent may be redefined as the person who has broader knowledge of the existence of a variety of types of requested data.

In addition, this person must have the ability to gather the requested data from multiple sources in the technical core. This ability is often associated with a certain level of authority within the company. Thus the person with authority must also have institutional knowledge, although staff with intimate knowledge of the records and the data – those at the technical core – may not or need not have authority.

Indeed, authority is a second attribute of the desired respondent identified by Edwards and Cantor (1991). They note that the desired respondent must not only have knowledge of the requested data, but the authority to release it. Edwards and Cantor (1991) note, however, that authority and knowledge may not reside in the same person.

Findings from our large company research suggest that authority is manifested in three decisions:

- the survey participation decision (see Willimack et al., 2001, for a discussion of this);
- delegation of the response task, that is, selection of the respondent;
- assessment of the priority of the response task relative to the respondent's other duties.

The latter two decisions have implications for the response model and the potential for measurement error. First, respondent selection is under the control of the business, and not the survey organization. Our research on the response process in large companies found that survey response was often delegated or assigned to someone other than an authority figure. In addition, limited empirical research has shown that, even when survey organizations make the effort to identify staff believed to be most knowledgeable about the requested data – that is, the desired respondent according to Edwards and Cantor (1991) – it is not uncommon for the actual respondent to be someone else (Ramirez, 1996; Davie, 1999).

Second, the person with authority needs to have, at least, a certain level of knowledge of company records in order to select an appropriate respondent that will minimize measurement error. That "certain level of knowledge" may vary with the type of survey and the breadth of data requested. Third, priorities set by the authority impact the respondent's motivation, or attentiveness to the response task, subsequently affecting data quality.

Assessment of Priorities

As a result, we believe that assessment of priorities is an explicit step in the response model. Since it influences the respondent's motivation, it too impacts the potential for measurement error. The greater the respondent's motivation to do a good job responding to the survey, the more attentive the respondent will be to the four core cognitive steps, reducing measurement error. The higher the priority for the response task, the greater will be the motivation.

Our research with large companies found the following priorities assigned to various tasks and assignments among financial reporting staffs:

1. Requests from management, as well as those from investors for public companies.
2. Preparation of Annual Reports, SEC and IRS filings.
3. Other periodic financial statements (quarterly or monthly).
4. Other government or non-government regulatory requirements (e.g., Hart-Scott-Rodino filings, Department of Energy or Department of Transportation regulatory filings).
5. Other government data requests (such as those from the Census Bureau and other statistical agencies).
6. Non-government data requests.

Priority is given to activities required to keep the business open and growing. Respondents noted that government reporting is not a revenue-producing activity – it bears a cost without an associated revenue or penalty. Thus, while reporting on Census Bureau surveys is taken seriously, it is not a company's highest priority task.

In addition, since respondents complete a survey within the context of their jobs (Nichols et al., 1999),

motivation is related to job performance and evaluation criteria. Finally, motivation is also associated with pride in one's work and professional standards.

Releasing the Data

Finally we add the release of the data as an explicit step in the response process. In our research on large companies, we found that, while the assigned respondent may be responsible for reporting individual data items on the survey questionnaire, it was not uncommon for authority figures to re-enter the response process prior to releasing the data to statistical agencies. These authority figures reviewed and verified survey forms for completeness and consistency, essentially performing their own internal "edit" step. They also reconciled reported data with other aggregated company figures to ensure that a consistent picture of the company was presented to the outside world, of which statistical agencies are a part.

In addition, authorities considered the confidentiality and security of the data release relative to the sensitivity of the data being requested. In some instances, release of the data was delayed until figures could be reviewed by upper management, or were provided to stockholders.

The Complete Model

The discussion in the previous sections suggests some minor revisions to the original Sudman et al. (2000) response model. These appear in ***bold italics*** in the complete model of the response process for establishment surveys proposed below:

1. ***Encoding in memory / record formation.***
2. Selection and identification of the respondent or respondents.
3. Assessment of priorities.
4. Comprehension of the data request.
5. Retrieval of relevant information ***from memory and/or*** existing company records.
6. Judgment of the adequacy of the response.
7. Communication of the response.
8. Release of the data.

The core cognitive steps – comprehension, retrieval, judgment, and communication – remain in tact, although we suggest some modification to the dimensions of the retrieval step. Our primary contention is that the three steps added to the model preceding the core cognitive steps – encoding/record formation, respondent selection/identification, and assessment of priorities – set a context for the cognitive process, and themselves potentially contribute to measurement error. The final step added to the model – releasing the data – also has consequences for measurement error.

Discussion

Our model was built inductively based on results of exploratory research on the response process in large companies, and primarily referring to survey requests for numerical data. Does the model describe the response process in small and medium-sized businesses? Is the model appropriate for non-numerical information requests? What modifications or caveats are needed so that it applies more generally?

We suggest a greater distinction between record formation and information retrieval than currently found in the literature, while data availability encompasses both. We suggest that cognitive research into the retrieval step more explicitly consider 1) the respondent's knowledge of records sources and how that knowledge is retrieved from memory; 2) the respondent's access to relevant records; and 3) the strategies and activities involved in physically retrieving data from records.

We have argued that the attributes of data availability should be evaluated relative to a particular use – i.e., survey response. However, it could be argued that availability be defined in terms of existence only. Does the fact that data exist in records mean that they are available? I.e., is existence a sufficient condition for availability? How should the consequences for measurement error be assessed under each scenario?

It is important to recognize that an organization such as a business or an establishment cannot respond for itself. It relies on a person within the organization to provide information on its behalf, that is, to act as a proxy for the organization. The literature on proxy response calls such a person an "informant." Regardless whether we refer to this person as an informant, a respondent, or a reporter, the key point is that this person is a proxy for the establishment, subject to issues of knowledge formation and memory retrieval raised by Eisenhower et al. (1991). In particular, how do respondents for businesses obtain and retain knowledge about records, information systems and data sources? How does the context of the job affect the encoding step and retrieval from memory? Does memory related to one's job differ from memories associated with oneself? How does this, vary across different possible respondents/informants who may be involved in responding to a survey? What does this suggest about selecting a respondent so that measurement error is reduced?

We have discussed the implications that respondent selection has for the four core cognitive steps embedded within our proposed response process model. We have contended that respondent selection is under the control of the business. Thus, it is possible for survey response to be delegated to different respondents/informants within

the firm. Who are these different candidates for being the respondent and what are the criteria by which they are selected? How do these criteria vary by survey? What alternative cognitive processes may be used by the different candidates and how is the quality of the data affected? Additionally, multiple respondents/informants may be required to complete a single survey request, causing successive cognitive processes as survey questions and instructions are communicated from one informant to another. What is the affect of this successive processing on measurement error?

We have suggested that priorities and other aspects of motivation affect a business respondent's cognitive processes. What can survey designers do to relieve the potential measurement error related to reduced attention to the response task? Lastly, the final step in our response process model, releasing the data, may also have consequences for measurement error.

While the traditional cognitive response model remains pertinent for evaluating potential measurement error in establishment surveys, it does not address many other issues common to the response process for businesses. To address these issues, we offer the response process model presented in this paper for consideration, while recognizing many remaining questions that warrant further research.

References

Carlson, L. T., J. L Preston, and D. K. French (1993) "Using Focus Groups to Identify User Needs and Data Availability," *Proceedings of the International Conference on Establishment Surveys*, American Statistical Association, pp. 300-308.

Davie, W. (1999) Internal office e-mail to Elizabeth Nichols, dated 2/23/99. U.S. Bureau of the Census.

Edwards, W. S., and D. Cantor (1991) "Toward a Response Model in Establishment Surveys," in P. P. Biemer, et al., eds., *Measurement Error in Surveys*, New York: John Wiley & Sons, pp. 211-233.

Eisenberg, W. and H. McDonald (1988) "Evaluating Workplace Injury and Illness Records: Testing a Procedure," *Monthly Labor Review*, April, pp. 58-60.

Eisenhower, D., N. A. Mathiowetz, and D. Morganstein. (1991) "Recall Error: Sources and Bias Reduction Techniques." In Paul Biemer et al., eds., Measurement Errors in Surveys. New York: John Wiley& Sons, pp. 127-144.

Groves, R. M., Cantor, D., Couper, M., Levin, K., McGonagle, K., and Singer, E. (1997) "Research Investigations in Gaining Participation from Sample Firms in the Current Employment Statistics Program," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 289-294.

Nichols, E., D. K. Willimack, and S. Sudman (1999) "Who Are the Reporters: A Study of Government Data Providers in Large, Multi-Unit Companies," paper presented at the Joint Statistical Meetings of the American Statistical Association, Baltimore, Maryland, Washington, D.C.: U.S Census. Bureau.

Ponikowski, C. H. and S. A. Meily (1989) "Controlling Response Error in Establishment Surveys," *Proceedings of the Survey Research Section*, American Statistical Association, pp. 258-263.

Ramirez, C. (1996) Respondent Selection in Mail Surveys of Establishments: Personalization and Organizational Roles. *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 974-979.

Searson, M. (2001) "Strategies to Implement Electronic Collection of Multiple Worksite Report Data," paper presented at Symposium 2001: Achieving Data Quality in a Statistical Agency: a Methodological Perspective, Ottawa: Statistics Canada.

Sudman, S., D. K. Willimack, E. Nichols and T. L. Mesenbourg (2000) "Exploratory Research at the U.S. Census Bureau on the Survey Response Process in Large Companies," *Proceedings of the Second International Conference on Establishment Surveys*, American Statistical Association, pp. 327-337.

Sykes, W. (1997) *Quality Assurance Study on ONS Employment Questions: Qualitative Research*, London: Social Survey Division, Office of National Statistics.

Tomaskovic-Devey, D., Leiter, J. and Thompson, S. (1994) "Organizational Survey Nonresponse" *Administrative Science Quarterly*, 39, 439-457.

Tourangeau, R. (1984) "Cognitive Sciences and Survey Methods." in T. Janine, et al., eds., *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines.* Washington, D.C.: National Academy of Science.

U. S. Bureau of the Census (1990) *1989 Record Keeping Practices Survey*, Economic Census Staff, Washington, DC: U.S. Bureau of the Census.

Utter, C. M. and Rosen, R. J. (1983) "Employer Records and Recordkeeping Practices of BLS-790 Sample Respondents. *Proceedings of the Business and Economic Statistics Section,* American Statistical Association, pp. 254-259.

Willimack, D. K., E. Nichols, and S. Sudman (2001) "Understanding Unit and Item Nonresponse in Business Surveys," in R. M. Groves, et al. (eds.) *Survey Nonresponse*, New York: John Wiley & Sons.