4 An Overlooked Approach in Survey Research: Total Survey Error

René Bautista

The author is grateful to NORC Center for Excellence in Survey Research (CESR) for the grant support provided to complete this chapter. All errors and omissions, of course, remain the sole responsibility of the author.

4.1 Social Research and Survey Methods

Social scientists are constantly challenged to understand and measure human behavior. Researchers are frequently pressed to make inferences about populations of interest based on individual level data, but usually have limited resources at their disposal. The challenge is even bigger when researchers aim to obtain accurate, reliable, cost efficient, and representative results of the population of interest. Fortunately, there are scientific tools that allow social researchers to gather valuable data. Scientific surveys are useful instruments to measure concepts and behaviors based on a sample of cases.

Recent developments in the field of survey methodology are having a great impact upon the way individual data-based research is conducted in the social sciences. Advances in survey methods are changing the way data collection strategies are conceived. Ultimately, they are changing the way we interpret survey data. There are sources of error in surveys, nonetheless, that limit our ability to make inferences about populations of interest. The degree of accuracy in a survey (that is, obtaining survey measures that reflect population parameters or "true values") depends on several sources of error.

R. Bautista (✉)
NORC at the University of Chicago,
55 East Monroe St, 20th floor, Chicago,
IL 60603, USA
e-mail: Bautista-Rene@norc.org

In this chapter, we will describe potential sources of error in a survey, and we will discuss how survey design features can help us minimize the effect of error sources on survey estimates. The information presented in this chapter is based on current theories in the field of survey methodology as well as practical experience. We will discuss a usually overlooked survey methodology paradigm, the Total Survey Error (TSE). The TSE is a useful framework to understand survey errors in a comprehensive manner. Although the TSE can be discussed in mathematical terms, we will not follow a mathematical style. Instead, we introduce the TSE as way of thinking about the tradeoffs usually involved in survey research. The aim of this chapter is to provide a description of the TSE in such a way that the reader finds it useful, especially during the early planning of a survey.

4.2 The Notion of Error in Surveys

The notion of "errors" or "limits" is unavoidable in the context of scientific surveys. Although social researchers are aware of the existence of errors in surveys, the interpretation of errors tends to be incomplete and frequently oversimplified. The notion of error usually is associated with just one of several types of errors: the "sampling error." Sampling error is perhaps the most well-known source of error in surveys, but it is not the only one—as we will see later on. The sampling error occurs due to the fact that we are studying a fraction of the population randomly

L. Gideon (ed.), *Handbook of Survey Methodology for the Social Sciences*,
DOI: 10.1007/978-1-4614-3876-2_4,

selected in lieu of the whole population. This concept is also commonly referred to as "theoretical margin of error," or simply "margin of error." The underlying idea is that of a statistical basis to create "confidence intervals" or boundaries within which the statistic of interest (whether a proportion, a mean, regression coefficient, or other statistic) is expected to lie. Theoretically, as the sample size increases, the margin of error decreases.

The margin of error has been widely used in reporting univariate survey statistics in almost any probability survey. Starting with Neyman's (1934) seminal article, the field of statistics devoted a great deal of effort to the understanding of the sampling process. With today's access to information technologies, it is relatively easy to calculate the margin of error for proportions under the assumption of simple random sampling.

Unfortunately, the routinely reporting of the sampling error in surveys has mistakenly lent credence to the idea that surveys need to be designed based on the "margin of error" alone. Even more, sometimes scientists misguidedly assume that errors other than those related to sampling will occur at random, and will cancel each other out with no impact on the accuracy of the estimate.

The sampling error is a very limited representation of errors in survey design, leading social scientists to overlook other sources of error. Nonetheless, our understanding of errors in surveys is rapidly changing. In recent decades, starting with Groves' (1989) book *Survey Errors and Survey Costs*, methodologists have begun to devote a great deal of effort to investigate sources of "nonsampling errors." Over the years, it has become clear that even if we had resources to collect information from the whole population of interest (that is, if we conduct a census to eliminate the sampling error), a poorly design measurement instrument, inadequate response categories, a faulty opening script, the appearance of the interviewer, the lack of interviewer training, the channel of communication used to gather data, and other important survey design features, may jeopardize the quality of the data. Questionable data collection strategies may result in a significant waste of time and resources—even if the "margin of error" is presumably small.

4.3 The TSE Approach

Developments in the field of survey methodology during past decades have incorporated contributions from a variety of fields besides statistics. Disciplines such as psychology, sociology, economy, communication, linguistics, cognitive sciences, computer sciences, and others, are reshaping the notion of error in surveys. Currently, the field of survey methodology is multidisciplinary, and its scientific value is reflected in how survey errors are understood in the field (see for instance, Biemer 2010; Biemer and Lyberg 2003; Groves and Lyberg 2010; Groves 1989, 2004; Leeuw et al. 2008; Lessler and Kalsbeek 1992). In survey methodology, sources of error are studied from different perspectives because errors can be different in nature, but they all have a potential effect on survey results.

The TSE paradigm is a useful integrated theoretical framework that will help us understand survey errors. This paradigm puts emphasis on data quality, and help researchers prepare and evaluate survey designs. Social scientists can organize their survey preparation using as a theoretical reference the TSE. TSE considers important sources of error altogether. Ultimately, the TSE goal is to obtain survey statistics as accurate as possible, given available resources.

The TSE is a function of two kinds of limitations: "sampling" and "nonsampling" errors.

The literature in survey methodology typically breaks down the nonsampling errors into at least three sources of error: coverage error, nonresponse error and measurement error. Each source may contribute to the total error of the survey. Thus, we define TSE as the sum of sampling and nonsampling errors, as shown in Fig. 4.1.

We will discuss later on in this chapter each component of the TSE individually.

4.4 Two Building Blocks: Bias and Variance

Each of the error sources in the TSE potentially contribute to the *bias* and *variance* of a survey statistic. *Bias* is defined as a systematic departure of our survey estimate from the "true" population value—such departure can occur in a particular direction. *Variance* is defined as the uncertainty introduced in our survey estimate to represent the "true" population value. The terms *bias* and *variance* have been adopted into the survey methodology field from the survey statistics field; in particular, they come from a concept known as the Mean Square Error (MSE). The MSE is a function of *bias* and *variance*. In short, MSE is the expected difference between an estimated statistic and the true value due to systematic (*bias*) and random variations (*variance*) over theoretical repeated events. Formally, the MSE $=$ bias^2 $+$ Variance.

In statistics, the MSE has been used as composite measure that helps identify unbiased mathematical procedures to estimate population parameters; these procedures are also known as *estimators*. The resulting MSEs from different procedures, or estimators, are usually compared to each other. The best estimator is said to be "unbiased" when *bias* is non-existent. The underpinnings of the MSE are beyond the scope of this chapter; nonetheless, the idea of an unbiased estimator helps us understand the logic behind the TSE: TSE aims to achieve accurate survey statistics.

In the same MSE fashion, each component of the TSE can be theoretically decomposed into bias and variance; thus the TSE is equal to the sum of sampling bias, sampling variance, coverage bias, coverage variance, nonresponse bias, nonresponse variance, measurement error bias, and measurement error variance. In this classification of survey errors, when a systematic error occurs leading to over- or under-estimate the "true" value, we are in the presence of *bias*; when errors occur in an unrelated way introducing uncertainty about the "true" value that we want to estimate, we are in the presence of *variance*.

4.5 Specification and Processing Errors

There are two other types of errors frequently mentioned in the context of the TSE framework: the *specification error* and the *processing error*. *Specification errors* occur when there is no clear connection between theoretical concepts or constructs, and survey variables. This kind of error happens when key concepts are excluded or misspecified during early stages of the study. In other words, it is the consequence of an ill-defined research protocol or the lack of a research protocol at all.

Specification errors are more frequent than one may think. Less experienced social researchers frequently start crafting a survey questionnaire without having a clear written research protocol. Needless to say, such practice leads to weak conclusions as the resulting data will not reflect specific social constructs. Specification error can be easily minimized if researchers devote more resources (i.e., time and effort) to carefully select their theories, concepts, hypotheses, and survey metrics, before developing a questionnaire. Ideally each survey variable should be related to a concept, and each concept should be related to a hypothesis, which in turn, should be derived from theoretical knowledge.

Processing errors refer to flaws that occur once the survey data have been collected. These errors are often introduced inadvertently. They include errors due to computer programing, coding of open ended questions, data entry of close ended questions, data cleaning, imputation of missing data, weighting, and data reporting or tabulation.

Processing errors can be minimized with a careful revision of protocols that will be implemented after the data collection period. Implementation of quality control and quality assurance procedures throughout the data processing stage is usually more efficient, than a full final revision of the data, once they all have been processed.

Fig. 4.1 Sources of error which can have an impact on the bias and variance of a estimate

Total Survey Error =
Sampling error + Coverage error + Nonresponse Error + Measurement Error

A good practice is to document all coding rules and programing codes used during data processing. A well-documented data processing plan makes easier the identification of errors at this stage.

Specification and processing errors can drastically affect our survey estimates. In this chapter we choose not to include them as part of our theoretical equation in Fig. 4.1, because they do not have an obvious *bias* and *variance* component, as the other components have. Despite the fact that we are excluding the specification and processing errors from our TSE equation, the reader needs to be aware of their existence and to take preventive measures while planning a survey, to minimize their effects.

4.6 Sampling Error

Cost constraints commonly dictate the way empirical research is conducted in the social sciences. Since it is nearly impossible to collect data from all of the elements in the population of interest, scientific sampling has proven to be a valuable tool. Consequently, social researchers are in constant need of utilizing probability sampling to establish conclusions about a certain population.

Sampling error occurs because we do not collect information from all individuals in our population due to the fact that we analyze a sample of cases (instead of conducting a census of the whole population). Sampling error can be further divided into two components: *sampling bias* and *sampling variance*. Suppose we were to repeat several times the sampling process by drawing the same number of cases (not the same individuals) in exactly the same manner, out of the same population. Chances are that we would obtain a slightly different result every time. This is called the *sampling variance*. The basis to derive the variance is theoretical because we usually do not have resources to conduct many surveys under the same survey conditions.

A key aspect of scientific sampling is that all members in the population of interest—or target population—should have a known, non-zero, probability of being selected. When this condition is not satisfied, our sample is likely to be biased. *Sampling bias* occurs if the sampling process systematically gives no probability of selection to some individuals of the population, or gives a disproportionately small or large probability of selection, to some subjects. A way of minimizing the sampling bias is to give all subjects in the population an equal chance of selection.

If sampling bias exists due to different probabilities of selection, we could compensate for such differential probabilities of selection by means of *weights*—an adjustment that gives individual a disproportionate weight relative to their original weight. However, such adjustment may increase the uncertainty of our estimates, because subjects that were drawn with different probabilities will now have a disproportionate influence in the sampling variance.

The conceptualization and estimation of sampling variance heavily rely on theoretical statistical formulations because we do not actually draw and collect information from multiple surveys at the same time, under the same survey conditions. The computation of the sampling bias and sampling variance depends on assumptions about what would have happened if the sample were drawn numerous times—hence the name of "theoretical margin of error."

The sampling variance is a concept we use to establish the boundaries within which our estimate is expected to lie. Under this logic, the greater the number of cases in the sample, the smaller the variability. Table 4.1 illustrates the effect of the sample size on the "margin of error." These computations assume that cases were selected using simple random sampling. Also, they assume that the proportion of an estimated survey statistic is equal to 0.5,

Table 4.1 Margin of error for proportions under simple random sampling, assuming a 95 % confidence level

Number of cases	Margin of error (%)
100	± 9.8
500	± 4.4
1,000	± 3.1
5,000	± 1.4
10,000	± 1.0

and assume that the true value lies within the proposed range, 95 % of the time.

If we select 100 cases, we would have an estimate with a boundary of ±9.8 % points; but, if we set our sample size to 500 cases, we would reduce our uncertainty by almost 5 and a half percentage points (= 9.8–4.4 %). Increasing sample cases does not decrease the margin of error in a linear fashion, nonetheless.

If we compare a survey design having 1,000 cases with a survey design having 10,000 cases, we would see a reduction in the margin of error of nearly 2 % points (3.1–1.0 %). Increasing our sample size from 1,000 to 10,000 would improve our precision, but also would increase the costs of data collection efforts.

Survey designs with large number of cases represent also a great challenge in terms of survey quality. Having more cases mean hiring and training more survey personnel, doing more attempts to persuade people to participate in a survey, doing more supervision of data collection and data processing, more data management, and other burdensome logistic aspects. In some instances, it would be preferable to have a manageable, cost effective, closely monitored sample of cases, than a very large sample that can potentially increase other sources of error.

There are occasions of course, where a larger sample of cases may be needed. For instance, if the researcher wants to conduct analysis of subgroups in the population with relatively high statistical precision, then more sample cases are necessary. The point to emphasize, nonetheless, is that there are tradeoffs between feasible well-monitored samples and statistical power.

Besides increasing the sample size, *sampling variance* can be also minimized by using a procedure known as *stratification*. For instance, before drawing a sample we could divide our population into mutually exclusive groups (known as strata), based on available information. Once people are assigned to a particular stratum, we can proceed to select cases randomly from all of the strata. A population can be previously stratified into age groups, occupation, gender, literacy level, or other meaningful criteria. Stratification usually represents a type of assurance that we are randomly drawing cases from all possible groups in the population, therefore, decreasing the uncertainty of survey results.

Sometimes grouping occurs naturally in the population of interest. People may be grouped into clusters defined as classrooms, neighborhoods, offices, election precincts, and so forth. Clusters represent a way of reducing costs because there are marginal costs associated with collecting more information from the same cluster. The researcher may decide to take a sample of clusters in order to collect information and calculate inferential statistics; for example, but the researcher needs also to be aware of the fact that cluster sampling may increase the variance associated with survey estimates.

Cluster cases are usually related to one another. People in the same clusters are exposed to the same social environment, information, social influences, context, and so on; consequently, they tend to be very similar. Thus, even if data were collected from all of the individuals in the selected clusters—creating the illusion of an increased sample size—the research will not gain more knowledge from such a sample because those cases are correlated.

In practice, cost-effective scientific sampling strategies combine stratification and clustering. They may even use different selection probabilities for people from some groups. When the

Fig. 4.2 Coverage of population by sampling frame

(a). Undercoverage **(b).** Overcoverage

selection of sample cases is made using such strategies, the sampling approach is called *complex random sampling*. When a survey is designed using a complex sampling approach, the margin of error (i.e., the variance) is not calculated in the same way as in simple random sampling. Clusters, strata, and differential probabilities of selection need to be considered in the computation of the variance.

There are tradeoffs in the selection of strategy for sampling that may have an impact on the variance of our survey estimates. While some elements may minimize the sampling error, others can potentially increase it. The researcher needs to carefully consider such tradeoffs to have a cost efficient and statistical efficient sampling design.

4.7 Coverage Error

In any scientific survey, we need a listing of elements in the population to be able to draw a sample. This list of elements in the population is known as *sampling frame*. A sampling frame is usually comprised of households, individuals, telephone records, mailing addresses, administrative records, institutions, and others. Existing lists usually serve as sampling frameworks, and represent great advantages to survey projects because information is readily available. In cases where existing sampling frames do not cover the population of interest, researchers may need to create their own sampling frame in order to cover the entire population under study.

Coverage error occurs when individuals in the population of interest are missing from the sampling frame use to draw a representative sample. Coverage error can be determined even before drawing the sample. This is possible because we can anticipate that some elements of the population are not part of the sampling frame.

Coverage error can take the form of undercoverage or overcoverage (Fig. 4.2). When we systematically exclude population elements from our sampling frame, we are dealing with undercoverage. When elements other than those described as our target population are part of the sampling frame, we are in the presence of overcoverage.

Coverage error can be broken down into *bias* and *variance*. *Coverage bias* occurs when elements in the population are systematically excluded from the sampling frame. For example, if we conduct a survey of students in a particular college, we may ask the administrative office for a list of the students' email addresses, with the purpose of using it as our sampling frame. Students who do not have an email address registered in the administrative office, will not be part of the sampling frame. Accordingly, the bias of survey statistics obtained from our study is likely to increase due to coverage error.

An easy solution to coverage bias is to recognize the limitation of the sampling frame and redefine the population of interest in light of the characteristics of the sampling frame. The researcher can describe upfront the elements in the population from which we are able to make valid statistical inferences. In our example, the target population would be college students with a valid email address registered in the administrative office.

An alternative solution for minimizing bias due to coverage error is to create a sampling frame in such a way that it includes all individuals in the population. Clearly, this option can be very costly because it requires an enumeration of all members in the target population. An additional solution to minimize bias due to coverage is to use supplemental sampling frames.

The use of two or more sampling frames requires posterior statistical adjustments to obtain unbiased results. Survey estimates from two or more sampling frames would be combined by means of weighted averages. Nevertheless, since individuals will be disproportionally contributing to this averaged number, they are likely to increase the uncertainty around our resulting survey estimates. Put differently, in theoretical replicates of the selection process using the same sampling frames, and the same survey design, we will observe a great deal of variability on our estimates, leading to an increase in *variance* due to coverage error.

A careful selection of an up-to-date sampling frame that covers as many sampling units as possible, along with a thoughtful definition of the target population may help minimize coverage errors. If necessary, a supplementary sampling frame for those undercovered individuals may help as well. For instance, if the intention of the researcher were to draw a sample of people from a minority group for which separate sampling frames exist, then the researcher may need to design a multiple sampling frame survey. Importantly, prior to drawing the sample, the researcher needs to exclude ineligible individuals from the sampling frame to avoid misleading conclusions.

4.8 Nonresponse Error

Ideally, in a study based on scientific sampling, all selected individuals should answer all of the questions included in the data collection instrument. In practice, this rarely happens.

Nonresponse error occurs when part of the information sought is not collected because sample individuals choose not to respond; as a result, survey statistics may or may not be representative of population parameters. If sample individuals choose not to answer just some of the questions, we will experience *item nonresponse*. If sample individuals refuse to answer the entire questionnaire will be in the presence of *unit nonresponse*.

$$\text{Non response bias} = (\bar{Y}_r - \bar{Y}_m)\left({}^{M}/_{N}\right)$$

Fig. 4.3 Computation of nonresponse bias

As it was the case with sampling and coverage error described in the two previous sections, nonresponse error can also be decomposed into *nonresponse bias* and *nonresponse variance*. *Nonresponse bias* occurs when responses obtained from those who accepted to participate in the survey are systematically different from those who chose not to participate. Such bias can increase depending on the proportion of people who chose not to participate. The proportion of people who participate in the survey is typically represented as a *response rate*.

The response rate in a survey is incorrectly considered an indicator of survey bias. Response rate alone is just one part of the story in the estimation of nonresponse bias. A basic equation can help us understand the role of (non)response rates (for details on response rate calculation, see AAPOR 2009). To obtain the nonresponse bias, one could calculate the difference between the respondent and nonrespondent means multiplied by the nonresponse rate (Fig. 4.3):

In Fig. 4.3, the expression $\bar{Y}_{\mathrm{r}} - \bar{Y}_{\mathrm{m}}$ represents the difference in the means of respondents ($\bar{Y}_{\mathrm{r}}$) and nonrespondents ($\bar{Y}_{\mathrm{m}}$), and M/N represents the ratio of nonrespondents to all sample people (M = number of respondents; N = total of respondents and nonrespondents). Although the mean of nonrespondents ($\bar{Y}_{\mathrm{m}}$) cannot always be determined, the equation shows the theoretical logic of the nonresponse bias.

Using response rates alone as a way of judging the quality of a survey can be misleading. A high response rate (i.e., a small nonresponse rate) can still introduce a sizable bias in an estimate if the difference between those successfully sampled and those who refuse, is considerable. Likewise, a small response rate (i.e., a large nonresponse rate) does not necessarily mean there is bias if the difference between respondents and nonrespondents is negligible.

Statistical adjustments could be implemented to compensate for nonresponse bias. In the case of *unit nonresponse*, weighting techniques can help reduce the difference between respondents and nonrespondents once the data have been collected. In the case of *item nonresponse*, single or multiple imputation techniques (i.e., entering model-based simulated values for data which do not exist) can help represent data that otherwise would not be available for analysis. Statistical approaches that compensate nonresponse bias can introduce uncertainty to our results. That is, we modify the bias to the expense of the variance.

If we theoretically repeat exactly the same survey over and over again, using the same survey design and the same weighting techniques, the range of possible survey outcomes is likely to be wider than the one we would have had, if those values had not been missing. The fact that we would give respondents disproportionally different weights (in the case of unit nonresponse) or simulating data using imputation techniques (in the case of item nonresponse) is likely to increases the *variance*.

There are survey design features that can help us improve our cooperation rates, whether at the item or unit level. For example, if the researcher anticipates higher levels of item nonresponse because of the sensitive nature of the questions, the researcher may consider a self-administered mode of data collection (whether paper- or computer-based instruments) instead of an interviewer-administered method.

If solutions to improve response rates include the use of information technologies, the researcher needs to be aware of the tradeoffs involved. At this time in history, younger and educated people tend to be more likely to have positive attitudes toward technology than older and less educated people, making those who are open to technologies more likely to participate in the survey using information technologies.

Offering incentives, whether monetary or nonmonetary, may also help motivate people to participate in a survey. The kind of incentives offered to sample persons depends on the kind of target population. Sometimes offering a copy of the final report is a good incentive. Other times, people are not interested in the topic of the survey and they place value on time, thus monetary incentives become more relevant in the study.

Overall, researchers should strive to reduce the burden imposed on respondents. This means that the researcher can potentially use multi-mode data collection approaches. Some of the information can be collected using a particular mode among some subgroups, and other subgroups may answer using a different mode of data collection. For instance, older and less educated people may be measured using a mail survey, and younger and more educated people can be surveyed using web-based modes.

Regardless of the mode of data collection, the researcher needs to be aware that respondents need some kind of reward (whether a social or economic reward), and that such reward needs to be delivered. There are tradeoffs, of course, in the use of incentives. Some respondents may rush through the survey so they can obtain the promised reward while others may be offended if money is offered. A careful selection of rewards (monetary or in-kind) as well as a carefully designed data collection protocol, can improve response rates. Toepoel (Chap. 13) provides a thorough discussion on the effect of incentives in reducing response rates and decreasing nonresponse bias, and the potential consequences of such incentives to quality of responses.

The way respondents are first contacted to participate in the survey has a great impact on their decision. To the extent that respondents perceive the survey solicitation as a legitimate request, survey cooperation is likely to happen. Also, it is important that respondents perceive that his cooperation and the corresponding exchange of information is beneficial for both, the respondent and the researcher. Likewise, if the respondent perceives that the sponsor of the survey is a prestigious institution or an authoritative figure, survey cooperation is likely to improve.

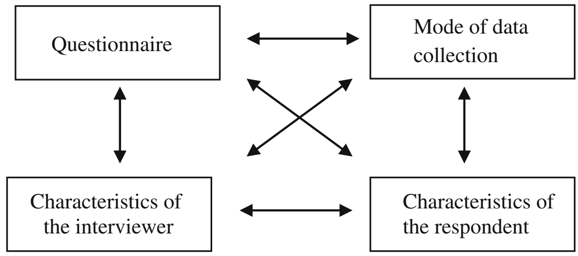

Fig. 4.4 Interacting sources of measurement error

Further, the researcher needs to carefully design a first-contact protocol as well as a follow-up protocol for refusals. These two protocols usually involve preparation of letters, friendly reminders, preparation of standardized scripts, and intense interviewer training on persuasion of sample individuals.

A follow-up protocol for nonrespondents is just as important as the first-contact protocol. The researcher needs to have these protocols prepared before starting the data collection process. Respondents may reply they are busy at the moment of contact, that they do not feel completely comfortable answering questions, that they have answered many surveys already, among many other answers. Depending on the type of refusals (whether "soft" or "core"), more experienced and better trained interviewers should try to persuade them again. Since refusal conversion is usually costly and time consuming, the researcher may consider collecting data from a subsample of nonrespondents to have at least an approximation of the bias due to nonresponse. For further discussion on why people agree or refuse to participate in surveys, see Albaum and Smith (Chap. 11), and Glaser (Chap. 12).

4.9 Measurement Error

In survey research we aim to represent social constructs with survey measures as accurate as possible. Unfortunately, several aspects related to the survey design impose restrictions to gauge the "true" value of a social concept. The "true" value can be an observable variable—which can be directly measured, such as weight, size, age, income, and others—and can be an unobservable or latent measure—and cannot be directly measured such as intelligence, happiness, quality of life, and satisfaction among others.

Measurement error occurs when there are differences between the estimated value and the "true" value due to survey design elements. Explicitly, measurement errors come from inaccuracies in responses due to four sources (Fig. 4.4): the questionnaire, the mode of data collection, characteristics of the interviewer, and characteristics of the respondent. All these sources interact and can be the cause of measurement error at the same time.

When measurement error occurs in a systematically fashion, the concept under study is misrepresented in a particular direction—whether positively or negatively. This is known as *measurement bias*. If the error occurs in a random fashion (not in a particular direction) over theoretical replicates of the same survey design, then we would likely say that we have an increase in *variance* due measurement error.

Examples of survey design aspects that can lead to measurement error include poor question wording, unclear question instructions, erroneous skip patterns, lengthy questions, inadequate response options, the topic of the questionnaire, timing, sponsorship, confusing visual designs, data collection methods, interviewer characteristics, faulty interviewer training, interviewer actions (whether indicated by the training or unforeseen behaviors), interviewer expectations, respondent reactions (whether to the topic or to the interviewer appearance), social pressure in the interviewer-respondent interaction, and respondents' memory erosion among many others.

Responses are provided based on the interpretation of interactive survey design elements. For instance, the measurement process in interviewer-administered modes requires conversational actions between interviewer and respondents to be able to complete the questionnaire. Thus, the researcher needs to decide if the interviewing protocol will utilize a fully scripted, standardized approach (i.e., having a standardized questionnaire), or if the interviewing protocol will take a flexible approach (which means the interviewer will have latitude to rephrase questions as needed).

One of the reasons for choosing standardized interviewing techniques is because interviewers represent a source of error. Different interviewers may interpret questions differently, deviating from the intended meaning of each question; therefore, they should not be allowed to change words or terms. If interviewers introduce unanticipated variability (whether systematic or random) the accuracy of the survey data would decrease.

If we choose a flexible interviewing approach is because we assume that respondents are a source of measurement error. Respondents may have different interpretations of the terms; therefore, the interviewer should intervene to provide clarification and assistance to gather accurate data.

In the end, these two competing approaches aim to improve accuracy of the data; they just identify the source of measurement error differently. Standardized interviewing aims to minimize measurement error due to the interviewer. Flexible interviewing aims to minimize measurement error due to the respondent. They both have advantages and disadvantages. The researcher needs to decide the most suitable approach depending on the anticipated error source in the study: interviewers or respondents, and depending on the challenges the researcher faces.

To date, the majority of surveys utilize standardized interviewing techniques. Under this approach, interviewers are trained to have a neutral position toward respondents' answers, read questions exactly as worded, follow established skip patterns, probe inadequate answers in a nondirective manner, record answers as expressed, and maintain an interpersonally, nonjudgmental relationship with the respondent.

Despite these guiding principles in standardized interviewing, the interviewer faces situations where respondents do not provide information needed due to a poorly developed question. Interviewers are naturally tempted to change or reword the question in order to get an answer—even when they are trained to leave the interpretation up to the respondent. This emphasizes the importance of resources allocated to questionnaire design and to interviewer training.

Seemingly irrelevant aspects of questionnaires may introduce measurement error to our survey. The order in which questions appear in a questionnaire may have an impact on the way questions are interpreted; namely, preceding questions affect the way in which subsequent questions are answered. Respondents usually provide answers in the context of an interview, not considering each item separately.

Another relevant feature in questionnaire design is the order of response options. When respondents do not have strong positions or definitive answers in some questions, the order of response categories may be a matter of highly importance. In this context, the mode of data collection may have an influence on the answer. For example, in auditory data collection modes (such as in a telephone surveys), the last response option that the respondent hears is more likely to be selected. In visual data collection modes (such as in a paper and pencil survey, or a web survey), the first visualized response category is more likely to be endorsed, relative to the rest of response options.

A solution to avoid measurement error due to the order of response categories is to rotate response options throughout the administration of the survey; however, that might not be feasible in all modes of data collection. In the context of mail surveys, for instance, it may require creating different versions of the same questionnaire. This can complicate the data entry process once the forms have returned from the field. Clearly, it will be easier to implement

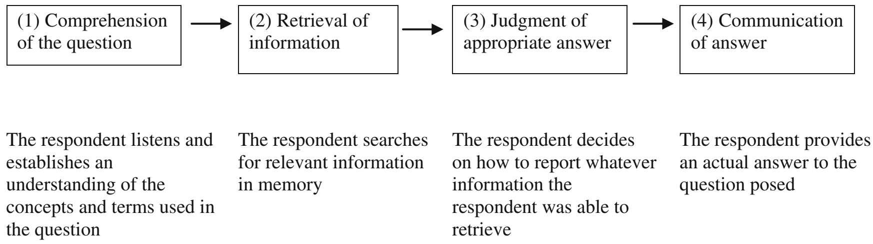

Fig. 4.5 Basic theoretical model of the survey response process

rotation of response categories in computer-assisted surveys, but that would require a significant investment in technology resources.

The selection of data collection method (telephone, Internet-based, in person, paper and pencil, computer-assisted, and others) is usually tied to other survey design decisions such as sampling frame, sampling strategy, expected response rates, and costs. An Internet-based probability survey could be feasible if we have an adequate sampling frame; but it may not be feasible due to uneven access to information technologies among members of the target population.

In general, measurement error is hard to measure because we usually do not know the true value, or because it requires advanced statistical techniques, or simply because we use qualitative methods to study and understand it. However, measurement error is one of the greatest sources of error in survey research. Among the qualitative techniques that can help us understand measurement error (in particular errors related to the respondents) are *cognitive interviews*.

Cognitive interviews are semi-structured interviews that provide insights on the cognitive process that takes place in respondents' mind while answering a survey question. This interviewing technique makes extensive use of a cognitive model which describes the steps respondents take to come up with an answer. The basic foundation of such cognitive model is presented in Fig. 4.5.

While the specifics of cognitive interviewing are beyond the scope of this chapter, the aim of a cognitive interview is to determine what the respondent experiences at each stage of the process. For example, if respondents experience difficulty in understanding some of the terms used in a particular question, or if they constantly ask for repetition of a particular question, that would suggest a problem of comprehension and the researcher would have to make improvements to the question.

If respondents are not able to remember events or specific information (e.g., visits to libraries, purchases of electronic devices, visits to the dentist, crimes, and other information of interest), for example, over a period of twelve months, but is able to remember events in the past six months, that would suggest that a problem exists at the second stage of the cognitive process. In that particular case, a possible solution is to have a shorter period of reference in the question.

If respondents are able to understand the question, retrieve relevant information, but do not feel quite sure about which response option best reflects his or her thoughts, or are confused on the meaning of each response option, it would suggest that problems are occurring at the third stage of the survey response process. Therefore, the researcher would have to modify response options or response scales to make sure he or she is gathering data from respondents accurately.

When respondents feel their answers would not conform to social norms, and they experience social pressure to communicate their answers in a different way than their true answers, it may be indicative of problems related to the fourth stage in the cognitive process model. For example, in a question about attitudes toward abortion, a respondent may not feel comfortable communicating his or her answer to the interviewer, and may choose to modify their answer on such question.

In that case, the researcher would have to modify the question being posed and how the answers are being collected. Conceivably, self-administered methods could be more convenient for sensitive questions. Perhaps the researcher can take advantage of portable devices (i.e., mp3 players, handheld devices, laptops, and others) to administer the survey and hand the respondent a paper and pencil instrument so only the respondent knows what he or she is answering. This may help minimize the effect of measurement error due to the social pressure that exists in the interviewer-interviewee interaction. Nonetheless, incorporating technology may increase the costs of the project.

In the survey methodology field, there is a growing body of literature devoted to measurement error, and about how social researchers can potentially minimize its effects on survey estimates. Given the fact that sources of measurement error interact altogether, social scientists are encouraged to think comprehensively about these aspects when designing a survey. Solutions involving technology are rapidly changing and are becoming more affordable. When deciding on alternative methods as a way of minimizing measurement error, a revision of the latest literature on data collection methods is highly recommended.

4.10 Measuring the Total Survey Error

Social researchers constantly face a challenge on how to decide about a survey design. The first reaction, very often, is to put most of the funds available into increasing the size of the sample or to improve response rates; however, such a decision might not be necessarily the best choice all the time. There are instances where it is clearly preferable to allocate more resources on interviewer training, questionnaire design, selection of a better sampling framework, or any other survey design feature, than spending resources on getting more respondents. The TSE approach can help social scientists better inform their decisions.

A common question is how to calculate a single measure of TSE to able to compare survey designs. In that sense, the survey methodology field is still evolving. Several mathematical methods could be used to estimate the TSE for survey statistic under a particular survey design. For instance, multilevel models techniques can be used to estimate the influence of interviewer on respondents.

Latent class analysis methods are used to infer unobservable true population parameters, and to compare them to survey estimates. Structural equation models can help represent the correlation between several error sources. Also, methods dealing with inferential statistics are in continuous improvement for better variance estimation processes to account for complex sampling design features.

To date, there is no single straightforward method to compute a measure of TSE. Although the computation of an actual TSE measure may not be readily available in commercial software packages, TSE is still a very useful theoretical framework to help us understand comprehensively aspects that can jeopardize the accuracy of our results. Usually the best decisions in survey research are made when the researcher considers simultaneously the different components involved in the design.

As it should be obvious by now, there are tradeoffs between survey design elements. All decisions have an implication for accuracy, and all decisions are related to costs. There is no one-size-fits-all survey design that can be applied to all research projects. The researcher is encouraged to use the TSE paradigm to guide their discussions and decisions on survey design.

References

AAPOR. (2009). Standard definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys 2009. Available from http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions1&Template=/CM/ContentDisplay.cfm&ContentID=1814

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly, 74*(5), 817.

Biemer, P. P., & Lyberg, L. (2003). *Introduction to survey quality*. Hoboken: Wiley.

de Leeuw, E. D., Hox, J. J., Dillman, D. A., & European Association of Methodology. (2008). *International handbook of survey methodology, EAM book series*. New York: Lawrence Erlbaum Associates.

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly, 74*(5), 849.

Groves, R. M. (1989). *Survey errors and survey costs, Wiley series in probability and mathematical statistics. Probability and mathematical statistics*. New York: Wiley.

Groves, R. M. (2004). *Survey methodology, Wiley series in survey methodology*. Hoboken: Wiley.

Lessler, J. T., & William D. K. (1992). *Nonsampling error in surveys*, *Wiley series in probability and mathematical statistics. Probability and mathematical statistics*. New York: Wiley.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society, 97*(4), 558–625.