Improving Data Quality in Face-to-Face Survey Research

Carolyn Logan, *Michigan State University*
Pablo Parás, *Data OPM*
Michael Robbins, *Princeton University*
Elizabeth J. Zechmeister, *Vanderbilt University*

ABSTRACT Data quality in survey research remains a paramount concern for those studying mass political behavior. Because surveys are conducted in increasingly diverse contexts around the world, ensuring that best practices are followed becomes ever more important to the field of political science. Bringing together insights from surveys conducted in more than 80 countries worldwide, this article highlights common challenges faced in survey research and outlines steps that researchers can take to improve the quality of survey data. Importantly, the article demonstrates that with the investment of the necessary time and resources, it is possible to carry out high-quality survey research even in challenging environments in which survey research is not well established.

Ensuring data quality in survey research remains a paramount—if often under-discussed—issue for research into mass political behavior. During the past two decades, this has been especially true as surveys became a more common method of research in countries around the world. Best practices, often developed in the United States, have been imported to other contexts despite differences in survey modes, techniques, and challenges. Data quality encompasses a wide range of issues, including implementing sample plans correctly and ensuring that field teams have the capacity to carry out the design. Meanwhile, in recent years, the issue of data fabrication—that is, the intentional departure from specified protocols by a member of the data-collection team—also has become a more prominent concern (Bredl, Winker, and Kötschau 2012; Kuriakose and Robbins 2016; Spagat 2016).

Afrobarometer, AmericasBarometer, and Arab Barometer have all confronted the challenges of ensuring data quality and preventing fabrication in the work they have done in more than 80 countries worldwide. Evidence from these projects makes clear that it is possible to conduct high-quality, nationally representative, and reliable public-opinion surveys even in challenging conditions. However, this process is time and resource intensive, and investments in data quality must be made throughout the life of the project. Achieving high-quality data is not only a question of monitoring and policing by principal investigators (PIs) but also of partner commitment, capacity, and correctly structured incentives.

Following is a summary of best practices developed by these organizations during each stage of the research process. When these methods are followed, the likelihood of obtaining valid and reliable data and reducing fabrication is high. Although our own focus is on projects in settings that are relatively less developed with relatively more challenges, many of these practices are relevant to researchers working in other contexts as well, whether comparative survey projects, single-country surveys, or surveys of a specific population within a country.

STAGE 1: PRE-FIELDWORK

Investments in data integrity must start before fieldwork begins, when designing the survey instrument, identifying suitable partners, selecting and training fieldworkers, and planning fieldwork logistics. A well-designed questionnaire not only ensures conceptual equivalence but also helps interviews to flow quickly and smoothly, reducing the likelihood of unauthorized shortcuts. Partner organizations that demonstrate real commitment and buy-in to the methodology and that have adequate resources and the necessary capacity to do the work well will be on the frontlines in producing high-quality data. Effectively recruited fieldworkers who are adequately trained, resourced, fairly compensated, and motivated are more likely to commit to collecting good quality data and foregoing fabrication. Finally, adequately planning, resourcing, and supervising

Carolyn Logan *is associate professor of political science at Michigan State University and deputy director of the Afrobarometer. She can be reached at clogan@msu.edu.*
Pablo Parás *is president and founder of Data OPM in Mexico City. He can be reached at pp@dataopm.net.*
Michael Robbins *is director of the Arab Barometer based at Princeton University. He can be reached at mdr7@princeton.edu.*
Elizabeth J. Zechmeister *is Cornelius Vanderbilt Professor of political science at Vanderbilt University and director of the Latin America Public Opinion Project (LAPOP). She can be reached at liz.zechmeister@vanderbilt.edu.*
This is an updated version of the original article. For details please see the notice at https://doi.org/10.1017/S1049096519001689

© American Political Science Association, 2019
doi:10.1017/S1049096519001161
https://doi.org/10.1017/S1049096519001161 Published online by Cambridge University Press

fieldwork logistics reduces incentives for both partner organizations and individual interviewers to cut corners on samples or to fabricate data.

The Survey Instrument

Designing a survey instrument, especially for use in multiple countries and/or linguistic subgroups, requires careful attention to issues of measurement and translation. Survey items may have different cultural meanings or implications across or within countries. Instrument design should include a team of topical, cultural, and linguistic experts as well as survey specialists to address these challenges. Convening a team that includes those knowledgeable about the country(ies) or key target population(s) ensures that concepts and constructs are designed to be meaningful, relevant, and equivalent across contexts.

Partner Selection and Support

Most multinational survey research projects work with in-country partners to lead data collection; therefore, partner selection is critical in all aspects of data quality. There are key points to consider when selecting a partner.

Conducting high-quality, nationally representative surveys—especially when relying on face-to-face interviews in logistically challenging environments—is demanding work. It is ideal to find a partner that has a strong foundation of relevant experience in the topics, the type of research (e.g., market, academic, or electoral/political), and the methods (e.g., computer-assisted personal interviewing [CAPI] and experiments) of the planned survey, as well as other aspects of the survey-management process (e.g., data management and analysis). Inexperienced partners may have unrealistic expectations about what the work will entail, as well as insufficient budgets and inadequate planning, which creates a situation ripe for problems.

Evidence from these projects makes clear that it is possible to conduct high-quality, nationally representative, and reliable public-opinion surveys even in challenging conditions.

Translation of the instrument is another critical and time-consuming step. Translation should seek conceptual equivalence rather than a word-for-word approach. Back translations—that is, the process of translating a survey instrument back into the original language—identify certain problems but are likely to miss many others that can limit conceptual equivalence. A more useful approach is a multistep process including a team that first can perform an in-depth assessment and then the initial translation; undertake an adjudication of any issues that arise; pretest the draft translation; and again adjudicate any problems that arise to yield a final translation for fieldwork.

The design of the survey instrument also has a direct bearing on the quality of data collection and the propensity to fabricate data. Sensitive questions that interviewers are reluctant to ask may be omitted by field teams. Lengthy and/or complex questionnaires that test the patience of respondents may lead interviewers to cut short questions or interviews and potentially fabricate data to avoid respondent fatigue. Elaborate skip patterns that are difficult to implement or skips that significantly change the length of an interview (e.g., if a particular response results in skipping a set of follow-on questions) also may lead interviewers to "pre-select" responses that facilitate implementation of the interview. Question wording and translation quality also are important: better questions and translations are easier for interviewers to read and for respondents to answer. That is, interviews will flow more smoothly, thereby reducing the incentive to manufacture data to avoid an uncomfortable process.

To summarize, it is essential to work closely with partners who have expertise in the relevant context, culture, and language(s) and with interviewers—especially during pretests—to obtain feedback on the survey instrument and to identify challenges to implementing it correctly and completely. It may not be possible to eliminate all relevant challenges, but reducing them as much as possible facilitates data-collection processes, thereby enhancing data quality. These conversations also provide insight about which parts of the questionnaire implementation require the closest monitoring to ensure correct implementation.

Whereas experience is important, it cannot substitute for a partner's fundamental commitment to the quality and integrity of the survey. Although electronic data-capture methods make it increasingly feasible to monitor partners throughout the survey process, it will never be possible for PIs and/or sponsors[1] to observe and validate every aspect of data quality. It thus is essential to have partners who are committed to meeting high standards not only because it is in the contract or because the PIs and sponsors might be monitoring them, but also because they are committed to the process and the integrity of the data being collected.

Transparency, referred to by Data-OPM—the AmericasBarometer partner in Mexico—as the "*hasta la cocina*" principle, is another key indicator, especially with new partners. Committed and reliable partners should be willing to grant unrestricted access to all aspects of the research process, even inviting their collaborators "into the kitchen." Project PIs and/or sponsors should be able to observe any part of the data-collection process, from enumerator selection and training to pilot interviews and fieldwork, as well as data-validation processes. Ideally, no part of the project should be off limits.

Survey PIs and sponsors also have a critical role in building effective partnerships. They must be willing to pay for and incentivize quality. Local partners must have sufficient resources to conduct the research according to specifications, especially when some of these practices may differ from the partner's normal practices. Simply stated, the project leader should not automatically accept the lowest bid. Sponsors *must* be willing to provide adequate resources. This includes avoiding unnecessarily contentious bargaining processes that can lead to under-budgeting, which can generate perverse incentives for local partners to relax standards or avoid complying with the required survey specifications.

PIs and sponsors also should become acquainted with their local partners to identify their skills and capabilities and to provide the necessary technical support and supervision. This is especially critical for less-experienced partners that may not be fully

https://doi.org/10.1017/S1049096519001161 Published online by Cambridge University Press

prepared for the demands and expense of conducting a nationally representative survey to a high standard. PIs and sponsors also should consider steps that foster partner commitment, such as investing in the organization and incentivizing high-quality work. Investments take many forms, both financial and nonfinancial, and may include provision of hardware or software, technology transfer, and capacity building and human capital.

Interviewer Recruitment, Training, and Support

The goals are similar for the fieldwork team, which is the most critical link in ensuring collection of reliable data. Maximizing commitment to the survey protocols and minimizing incentives to fabricate requires the development of a pool of interviewers who have a clear understanding of the relevant survey methods and their role in implementing them; understand the importance of data quality; are both safe and well supervised in the field; and are fairly compensated.

Finally, PIs and sponsors should ensure that the local partner pays fair wages and makes adequate provision for interviewers while they are in the field. Participating in training sessions or observing fieldwork can help the sponsor to evaluate whether fieldworkers are satisfied with their working conditions. High dropout rates indicate that the demands of the work and the compensation being offered are not well balanced.

STAGE 2: FIELDWORK

The foci of real-time quality control should be twofold: (1) correct errors and poor performance by members of the field team, and (2) reduce fraud. Until recently, most face-to-face surveys relied on local teams to oversee both concerns. Small groups of interviewers were overseen by a field supervisor, who typically sat in on a certain percentage of interviews, performed callbacks in the field, and reviewed completed questionnaires to ensure that interviews were valid. Subsequently, the local partner's central office might perform more callbacks as additional validation.

It is essential to work closely with partners who have expertise in the relevant context, culture, and language(s) and with interviewers—especially during pretests—to obtain feedback on the survey instrument and to identify challenges to implementing it correctly and completely.

Achieving these goals begins with recruiting qualified interviewers. Fieldworkers who understand the context of the survey and the ultimate value of the work that they are doing are more likely to commit to the data-collection process. Comprehensive training in the specific protocols of the survey and thorough familiarization with the questionnaire—including pilot interviews—also are critical. Adequate and standardized training is essential even when working with experienced interviewers who have conducted many surveys. This ensures that they follow the particular survey protocols rather than rely on "usual practices" that may not reflect the same methodological standards. It is ideal to have a representative from the sponsor organization participate in the training not only to monitor and observe but also to impart an understanding of the larger context and body of work to which the survey will contribute. It also is advisable to ask local partners to train more interviewers than required so that weak performers can be removed from the team. Ideally, survey sponsors should participate in the selection or approval of enumerators.

Minimizing the risks that interviewers face during fieldwork—whether from crime and insecurity, government interference or even arrest, or other factors that may threaten their well-being—also reduces fieldworkers' incentives to deviate from survey protocols. Steps that can be taken in high-risk areas include the following:

- provide fieldworkers with letters of introduction explaining their work, including contact information for supervisors and managers
- require fieldworkers to introduce themselves to local authorities before beginning their work in a primary sampling unit
- substitute sample locations that put the personal security of interviewers at high risk
- plan work hours and locations to avoid more dangerous periods of the day
- have interviewers work in teams with knowledge of a project security plan

The use of CAPI approaches provides a greater ability to monitor and correct fieldwork while the survey is in progress. Such real-time quality-control efforts allow errors, misunderstandings, and poor performance by the fieldwork team to be corrected in the early stages of data collection rather than persisting throughout the project. For example, interviewers who read questions too quickly or skip instructions for show cards can be coached to improve their performance (or removed if coaching is ineffective). In addition, such real-time quality control identifies interviews that need to be canceled because of unacceptably low data quality while teams are still in the field and, therefore, at lower cost to the project. Finally, the array of paradata that can be captured via CAPI approaches allows research teams to identify cases of fraud that might escape detection in the case of post-hoc quality-control approaches alone.

Real-time monitoring takes various forms depending on the tools available to the researcher. First, GPS coordinates can be used to evaluate whether an interviewer is conducting fieldwork within the selected sampling area. Second, it is possible to use automated timers within the survey to check the length of the interview and determine whether it was conducted too quickly to be of reasonable quality. Third, CAPI may be used to capture images of interviewers at the start of the interview to validate that they are the ones who attended training and not surrogates. Fourth, it may be possible to record portions of the interview to validate whether sensitive questions indeed were read out or to permit the data-quality team to assess and provide feedback on how well the questions were read. Similarly, audio recordings can be used to validate the respondent's gender by listening for the interviewer's gender-specific references to the respondent. Fifth, it is possible to more easily track attempted contacts made by the interviewer and combine them with GPS data to ensure that appropriate household skip patterns were implemented. Sixth, if data are provided throughout fieldwork, data-analytic approaches

https://doi.org/10.1017/S1049096519001161 Published online by Cambridge University Press

can be implemented at earlier stages of fieldwork than with pencil and paper interviewing (PAPI) approaches.

However, although CAPI offers enormous possibilities for improving data quality, there is one caveat: the need to balance monitoring with the responsibility to maintain respondent confidentiality. Capturing photographic, audio, and/or GPS-coordinate records of an interview has the potential to reveal respondent identities. Protocols for collecting, using, and preserving these data must be carefully developed, considering both the country context and the risks involved, as well as respondents' own level of understanding of these processes. It is essential to maintain the respondents' *perception* of confidentiality as well as the reality.

Committed and reliable partners should be willing to grant unrestricted access to all aspects of the research process, even inviting their collaborators "into the kitchen."

Several projects are developing and standardizing these approaches with specialized software and programs, including the Latin American Public Opinion Project (LAPOP) FALCON© (i.e., Fieldwork Algorithm for LAPOP Control over Survey Operations and Norms). Using data collected for each interview from software and hardware, FALCON© analyzes GPS coordinates, time, audio, and images in the course of the interview in close to real time.[2]

The use of a program such as FALCON© provides additional benefits to the fieldwork and quality-control process. First, because many supervisors' "policing" responsibilities become automatized or are audited in real time from a central location, they can switch to being "research agents" who work more collaboratively with interviewers to improve data quality. Supervisors can focus on key research aspects such as respondent/sample selection, revisit processes, substitution procedures, improving and correcting enumerators' performance, and more efficient communication with the central office.

Second, to implement the tools of complex CAPI systems, the research company must develop auditors. These new staff people work from the central location in real time and become the key point of contact with fieldwork supervisors and interviewers. They rely on a control console that gives them key performance information and alerts them to minor or major problems that can be corrected promptly. This becomes an extra layer in the data-collection process entailing additional cost, which should be viewed as an investment in quality control. In short, CAPI works best to improve data quality if PIs and sponsors can devote the human resources required to track, analyze, and utilize the rich body of data that the system provides.

STAGE 3: ASSESSING DATA QUALITY AFTER FIELDWORK ENDS

Following the completion of fieldwork, several methods can be used to detect potential issues with data quality.

Recontact

A common method is to recontact a portion of the survey respondents. As part of this process, respondents usually are asked factual questions about the interview, including when and where it was conducted, content of the survey instrument, duration of the interview, and a description of the interviewer. It also is common to re-ask a small portion of the interview questions to validate that recorded responses—including both demographic and attitudinal variables—are consistent with what was entered by the interviewer.[3]

Recontact is especially important to validate PAPI surveys that were not recorded or observed by a team member other than the interviewer. The use of CAPI for real-time monitoring has decreased the importance of recontact as the primary means for detecting data-quality issues. However, recontact can be used as an additional safeguard or verification in cases in which questions remain about the quality or validity of a specific interview.

Paradata

To ensure data quality, it is essential to record extensive paradata—that is, variables relating to how the survey was collected such as interviewer identity, time, date, and geolocation. Implementing surveys using CAPI facilitates the collection of paradata because most survey programs allow for automatic collection. It also is easier to collect data throughout the interview—for example, by placing timers at multiple points in the survey.

Several quality checks can be accomplished using the paradata. Some of the simplest and most important relate to time stamps. Extremely short interview times likely mean that the full instrument was not implemented properly, whereas extremely long interview times may be the result of a break during the interview or another abnormality. Extremely short duration (e.g., five minutes or less) between face-to-face interviews also signals that the sampling plan is not being followed correctly. The time of day also is instructive because valid interviews are unlikely to take place overnight or at other unusual hours.

Additionally, there is a limit to the number of interviews that an interviewer can conduct each day. A sudden surge of interviews for one or more interviewers—especially close to the fieldwork completion deadline—may suggest a problem. Likewise, if interviewers report conducting interviews in distant locations on the same day, it may indicate a departure from the sampling plan.

Data-Analysis Methods

Several checks within the dataset can be used to identify potential data-quality issues. First, it is important to conduct basic logic checks within the survey, some of which relate to sheer impossibilities. For example, a 25-year-old respondent could not have lived in the same city for 35 years. If an interview is fabricated without an actual respondent present, the fabricator may make such errors. Other checks should include rare response combinations across variables. For example, a respondent who identifies as an atheist is unlikely to also attend church regularly and pray daily. These combinations suggest either fabrication or a respondent's lack of comprehension or attentiveness, which are data-quality issues.

Undifferentiated response patterns also may indicate fabrication or satisficing that affects data quality. In the simplest form, this outcome includes straight-lining, wherein the same response is selected for each variable in a section—or even throughout the

https://doi.org/10.1017/S1049096519001161 Published online by Cambridge University Press

entire survey. More sophisticated fabrication methods include an interviewer asking only a single question and then filling out the remaining items in a battery by inferring the respondent's likely answers. Principal-component analysis (PCA) determines if an interviewer has a lower variance for a set of questions, which could indicate fabrication by this means (Blasius and Thiessen 2015).

Because of the potential for logic checks or correlational analyses across the dataset, convincingly fabricating an interview is a difficult undertaking. For this reason, a common strategy is to generate fraudulent observations by duplicating valid ones. This form of fabrication can be done by an individual interviewer or another member of the data-collection team. However, given that most statistical programs include means to check for exact duplicates, fabricators may alter a small number of variables to evade these checks. Recent advances, including the use of the PercentMatch program, are effective ways to check for near-duplicate observations to prevent this type of fraud.

Additional checks include looking for short paths throughout the survey, particularly if there are extensive sections with questions that are asked only if a respondent provides a given response to a filter question. Similarly, high percentages of "don't know" or refusals by the respondent, including break-offs, may indicate a problem. A high percentage of missing telephone numbers, meaning that callbacks are less likely to be performed, is another potential red flag.

CONCLUSION

Evidence from across these three survey projects demonstrates that with the necessary investment of time and resources, it is possible to carry out high-quality face-to-face survey research even in challenging environments in which survey research is not well established. In the Arab Barometer's fourth wave, for example, less than 1% of interviews were flagged for signs of potential fabrication.

Similarly, LAPOP evidence suggests that fabrication is unlikely to be the primary data-quality issue facing researchers. In an analysis of surveys conducted by LAPOP in the 2017–2018 round, the modal cause for cancelation of an interview was "reading error"—that is, questions were read too quickly or erroneously or interviewer instructions were skipped. On average, only about 1% of surveys showed signs consistent with fabrication.

In contexts in which survey research remains relatively rare and conditions are challenging, data quality is unlikely to reach the same levels as in more established survey environments. However, the collective experience of our projects demonstrates that it is possible to validly measure public opinion in a variety of difficult contexts. In addition to the measures cited previously, cooperation among researchers throughout the field incentivizes local partners to maintain rigorous practices. For example, increasing communication about experiences with local partners can make interactions a multi-round game, whereby poor performance on one survey is likely to cost a local partner future business.

Yet, incentivizing local partners is not the only challenge in collecting high-quality data. The same issue extends to individual researchers and survey projects alike. As described in this article, this process can be costly and time consuming depending on the specific context, and project leaders also may be lured into deviating from best practices. We conclude by recommending an additional enforcement mechanism—one that is actually relatively simple and inexpensive to implement but which is key to ensuring that data collected for social science research is high quality: for publication, researchers should be required to document each step of the survey process, be fully transparent about the methods used to collect the data, and ultimately make their data public. All three of our projects took these steps, making all data available free of charge.

Complete transparency—hasta la cocina—*at all levels remains the most powerful incentive to promote high-quality data.*

This incentivized our project teams to work diligently to ensure that data quality improves with each round. Projects that do not make their data public—or do so only for a substantial fee—do not invite the same scrutiny from other researchers, making such investments in data-quality procedures less likely. Complete transparency—*hasta la cocina*—at all levels remains the most powerful incentive to promote high-quality data. ■

NOTES

1. We distinguish here between the academic project lead investigator(s) and organizations that contribute funding to the project.
2. For more information, including an assessment of the impact of these approaches on data quality, see Cohen and Larrea 2018; and Montalvo, Seligson, and Zechmeister 2018.
3. The same cautions noted previously about maintaining respondent confidentiality—both perceived and actual—apply here as well.

REFERENCES

Blasius, Jörg, and Victor Thiessen. 2015. "Should We Trust Survey Data? Assessing Response Simplification and Data Fabrication." *Social Science Research* 52:479–93.

Bredl, Sebastian, Peter Winker, and Kerstin Kötschau. 2012. "A Statistical Approach to Detect Interviewer Falsification of Survey Data." *Survey Methodology* 38 (1): 1–10.

Cohen, Mollie J., and Sebastian Larrea. 2018. "Assessing and Improving Interview Quality in the 2016/2017 AmericasBarometer." Methodological Note #002. Nashville, TN: LAPOP. Available at www.vanderbilt.edu/lapop/insights/IMN002en.pdf.

Kuriakose, Noble, and Michael Robbins. 2016. "Don't Get Duped: Fraud through Duplication in Public Opinion Surveys." *Statistical Journal of the IAOS* 32:283–91.

Montalvo, J. Daniel, Mitchell A. Seligson, and Elizabeth J. Zechmeister. 2018. "Improving Adherence to Area Probability Sample Designs: Using LAPOP's Remote Interview Geo-Locating of Households in Real-Time (RIGHT) System." Methodological Note #004. Nashville, TN: LAPOP. Available at www.vanderbilt.edu/lapop/insights/IMN004en.pdf.

Spagat, Michael. 2016. "Censorship and Self-Censorship in Research on the Iraq War." Paper presented at the 57th Annual Convention of the International Studies Association. Atlanta, March 16.

https://doi.org/10.1017/S1049096519001161 Published online by Cambridge University Press