

# Combining Multiple Methods in Establishment Questionnaire Testing: The 2017 Census of Agriculture Testing Bento Box

*Jaki S. McCarthy<sup>1</sup>, Kathleen Ott<sup>1</sup>, Heather Ridolfo<sup>1</sup>, Pam McGovern<sup>1</sup>, Robyn Sirkis<sup>1</sup>, and Danna Moore<sup>2</sup>*

There are many methods that can be used to test questionnaires, each with its own strengths and weaknesses. The best approaches to questionnaire testing combine different methods to both broaden and strengthen the results. The US Census of Agriculture (COA) is conducted every five years and collects detailed information on agricultural production, inventories, practices, and operator demographics from agricultural establishments. Preceding each COA, evaluation and testing is done to test new items in the questionnaire and improve data quality for the subsequent COA. This article will describe how a multi-method approach, which we call Bento Box Testing, was applied to establishment questionnaire testing leading up to the 2017 COA. Testing included solicitation of expert opinion, historical data review, cognitive testing, a large scale field test, and qualitative follow-up interviews. The benefits of these testing methods, considerations for establishment survey testing, and how their results in combination provide a stronger evaluation are discussed.

*Key words:* Cognitive interview; field test; expert review; multi-method testing; establishment survey.

## 1. Introduction

Social science researchers have increasingly been employing multiple methods in their research. Multiple methods can be used to triangulate results and gain better understanding of the focus of study (Creswell 2014). There are various ways of testing and evaluating survey questionnaires (Presser et al. 2004; Groves et al. 2009), and it is clear that choosing one method over another will provide different results (Presser and Blair 1994; Forsyth et al. 2004). While any individual questionnaire evaluation method may provide useful information, survey methodologists can also triangulate results from multiple methods (OMB 2016; Yan et al. 2012; Madans et al. 2011; Presser et al. 2004; Schaeffer and Dykema, 2004). Indeed, survey organizations such as Statistics Sweden have developed formal strategies to combine particular methods based on specific characteristics of their surveys (Persson et al. 2015).

The majority of research in survey methods has focused on surveys of households or individuals. Thus, examples of multi-method questionnaire testing for establishment

<sup>1</sup> USDA – National Agricultural Statistics Service, 1400 Independence Ave., SW, Washington, DC 20250, U.S.A. Emails: Jaki.McCarthy@nass.usda.gov, kathy.ott@nass.usda.gov, heather.ridolfo@nass.usda.gov, mcgovern.pam@nass.usda.gov, and robyn.sirkis@nass.usda.gov.

<sup>2</sup> Washington State University, Social and Economic Services, Pullman, WA 99164-4014, U.S.A. Email: moored@wsu.edu

surveys are not as common as for household surveys, but there are some (Phipps et al. 1995; Tuttle et al. 2010; and Willimack 2013). Although many elements of the survey response process are common to both establishment and household surveys, they are not one and the same (Edwards and Cantor 1991). Discussions of testing for establishment survey questionnaires note that factors such as response burden and the technical nature of the data may introduce issues that are not faced in surveys of households (Willimack 2013; Willimack and Nichols 2010) and these may be important in different ways for different methods of testing. Establishment respondents report as representatives of their establishments and appropriate respondents may be more difficult to identify and contact, or multiple respondents may be necessary. Establishment respondents often consult and transform records in order to answer the survey questions when responding. In addition, establishment populations may use specialized terminology or be highly skewed with respect to their size or other characteristics (Snijkers et al. 2013).

In multi-method testing, the results from each method can be complementary, either supporting or expanding on the information obtained in each. In the Census of Agriculture (COA) we call this approach a ‘Bento Box’ approach to questionnaire testing (McCarthy 2016). Using multiple methods of testing in tandem is a particularly useful strategy for establishment surveys, such as the COA, where characteristics unique to businesses may impact the strengths and weaknesses of each method.

Because the COA is the largest data collection conducted by the United States Department of Agriculture’s National Agricultural Statistics Service (NASS) with many types of establishments, the evaluation and testing for the COA is quite wide-ranging and extensive. Smaller data collections may not have the resources available to test as extensively as was done for the COA. However, the strengths and weaknesses of the methods and how they complement each other would be similar. The use of a bento box testing approach in the COA can serve as a case study of how combining methods results in a stronger overall questionnaire evaluation. It also serves to illustrate some of the special considerations for establishment survey testing.

A traditional Japanese bento box is a meal prepared according to five principles with five elements each (see Figure 1). While any ingredients can be included, if the five principles are used, the resulting meal will be nutritious and well-balanced with its complementary elements. We strove to combine information from five different, yet complementary questionnaire testing methods in the development of the 2017 COA data collection. The five methods in our COA testing bento box were:

1. historical data review,
2. expert review,
3. cognitive interviews,
4. field testing, and
5. qualitative follow up interviews.

Each of these methods has its own strengths and weaknesses and complements the others. Paralleling bento meal preparation, using information from all of these methods together provides a well-balanced questionnaire evaluation. Each method as implemented in our testing will be briefly described below. Issues unique to establishment surveys will

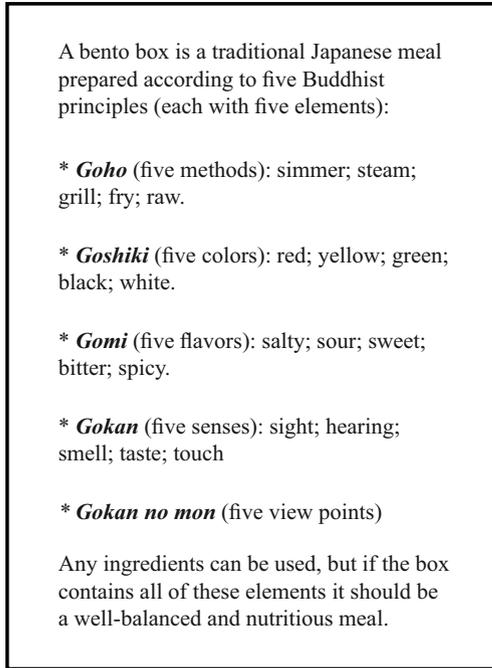


Fig. 1. Bento boxes.

also be discussed. Examples of how the results were used together for the COA will follow in the subsequent section.

## 2. Using Information from Multiple Sources in Testing for the 2017 Census of Agriculture

### 2.1. Brief Background on the COA

The Census of Agriculture is conducted by NASS every five years for the reference years ending in 2 and 7. Data from the 2017 COA will be published in February 2019 and is the only source of uniform, comprehensive, and impartial agricultural data for the United States, and its states, and counties (more detail at <https://www.agcensus.usda.gov/>). It is a complete count of US farms and ranches and the people who operate them. The definition of a farm (also referred to as an operation) for the COA is any place that raised and sold or would normally sell USD 1,000 or more of agricultural products. Operations can run the gamut from a few acres of cropland or a few animals to multimillion dollar corporations. In addition, every kind of agricultural production is included, from commonly grown commodities such as corn, soybeans and cattle to rare commodities such as alpacas, mink, crayfish, okra, kumquats, miscanthus, jojoba, and more. Approximately three million known and potential farms are contacted with a self-administered questionnaire mailed via US Postal mail. Questionnaires are mailed at the end of the reference year (in December 2017 for the 2017 COA) with an early February deadline for reporting. There are multiple

questionnaire mailings, followed by computer-assisted telephone interview (CATI) follow-up and limited in-person interviews. Respondents also have the option to report online via the internet.

The COA collects information on the characteristics of the operation, the commodities produced, production practices, farm income and economics, and farm operator demographics. The COA questionnaire is a complex 24-page questionnaire, and commodity-specific information is collected in separate sections of the questionnaire (see [Appendix 1](#)). For example, there are separate individual sections for detailed information on vegetables, fruit, hay, field crops, cattle, hogs, aquaculture, poultry, sheep, and so on. Other general information such as expenses, income, participation in government programs, and characteristics of the farm operators is collected for all operations. Due to the volume of records in the COA, records are run through an automated data editing process. However, a substantial number of records are reviewed by analysts if inconsistent data cannot be resolved automatically or if it has the potential for a substantial impact on resulting estimates. Inaccurate or missing data on questionnaires can substantially increase the amount of manual analyst data review needed. Increased review is both costly and increases the potential for errors in the results.

As with any large complex data collection, there are always potential improvements or changes possible both to improve data quality and to accommodate the changing agricultural sector and data needs. Thus, preceding each COA, NASS conducts activities to improve the questionnaire and data collection procedures for the next COA. As in previous COAs ([McCarthy and Buysse 2010](#)), NASS chartered a team (formed in mid-2014) to design and carry out multiple tests to evaluate and propose improvements to the forms and data collection. The team initially included 13 core members with various expertise, including survey methodologists, survey operations staff, subject matter experts, and field staff. However, many other NASS staff were also involved as needed for specific tasks or selected tests. For example, additional survey methodologists were involved during the cognitive interview phases of the testing, and additional subject matter experts participated when input was needed for specific topic areas. The intent of these tests was to improve the questionnaire and data collection procedures to improve accuracy of the data collected, improve response rates, and reduce respondent burden. While multi-method testing has been used in previous COA testing, the testing conducted for the 2017 COA was the largest and most extensive to date.

Preceding the 2017 COA, NASS used multiple methods of testing and evaluation to make improvements to the questionnaires. Using them together provided more information about potential problems and solutions to those problems. Each test provides information that can be used in isolation, but multiple techniques used together can be much more powerful. The testing used for the 2017 COA and several examples of the benefits of combining multiple techniques are provided below. Section 3 provides some specific illustrations of the power of research method triangulation.

## *2.2. Historical Data Review of the 2012 COA*

A bento box often begins with a bed of rice. Like rice in the bento box, historical data review can provide some basic quantitative information. To begin the evaluation and

testing activities for the 2017 COA, data from the preceding 2012 COA were reviewed. Edits and imputed data in the COA are the result of automated editing and imputation procedures or analysts' manual reviews and can be considered corrections for measurement errors. Both the edit rates and the absolute number were ranked. This is because some items, such as the number of acres in the operation, apply to all respondents. Others, such as items related to a specific commodity, may be reported by only a few respondents. For example, total land must be reported by all respondents, but the number and characteristics of goats (or ducks, boysenberries, turnips or any other individual commodity) is reported by relatively few respondents. Items with many edits are important, but high edit *rates* are also of interest. Items with either high numbers or rates of edit and imputation were considered for further review.

During each COA, a toll-free telephone helpline is available. During the 2012 COA, for each caller asking for help completing the questionnaire, the section of the questionnaire for which help was requested was logged. These data were also reviewed to identify sections where most help was requested. It can also be compared over time to assess the impact of changes from one COA to the next.

Reviews of the information from the 2012 COA (both edits and calls to the toll-free line) were conducted as soon as the testing team was formed. Because the COA form is so large, these reviews were used to narrow the scope of further evaluation and testing.

### 2.3. Expert Reviews – Varied Expertise Both Internal to NASS and External

Following all major data collections, NASS formally solicits feedback from staff on any problems and potential improvements. For the 2012 COA, these comments were collected after the bulk of census-processing activities in 2014. All comments pertaining to the questionnaires or data collection were reviewed by the COA testing team. Areas that field staff identified as requiring increased editing or that were problematic for interviewer administration were flagged for attention. Internal subject matter experts also provided input on items that required most analyst review or with suspected measurement error. This type of internal expert review is conducted as soon as possible after data collection so that any issues are not forgotten and details about them can still be recalled.

The most common type of expert reviewer in questionnaire evaluation is the survey methodologist or questionnaire design expert (Presser and Blair 1994; Olson 2010). In NASS, internal questionnaire design expert reviews are not conducted as a separate activity, because survey methodologists with questionnaire design expertise typically provide review and advice throughout survey development and testing. However, NASS did work with outside survey design experts to help design and test an alternative 'short' form intended for use with a sample of the COA (Moore et al. 2016). NASS also consulted with an outside expert in web survey design, who provided suggestions for improvements to the new online version of the questionnaire.

NASS also has ongoing communication with other USDA agencies, data users, and various agricultural industry groups. Each of these groups has specific expertise and can provide useful input in questionnaire review. This is an especially useful source of information for establishment surveys which commonly use technical terminology and typically contain specialized concepts. Some groups also provide information about new

or evolving data needs. For example, commodity organizations such as the National Corn Growers Association or the American Farm Bureau can provide insight into whether industry-specific terminology is appropriate or whether the format of information requested is likely to match farm record-keeping systems for their members. NASS contacts many different types of commodity organizations for feedback on their sectors.

In addition to routine feedback, special panels of experts can be commissioned for more detailed or focused review. For example, NASS convened a panel of subject area and survey methodology experts to address concerns that the COA was not adequately representing the role of women and beginning farmers in agriculture (Ridolfo 2016). This type of external review is typically undertaken only for the largest data collections, but survey organizations may convene this type of group when they are looking for major changes to questionnaires. In establishment surveys, these panels can also bring expertise and knowledge of the types of establishments targeted and their reporting processes. This panel, discussed in the example in the next section, met in April 2015 and provided recommendations for changes to the 2017 COA questionnaires. As expected, their initial recommendations for questionnaire changes were modified several times after testing with respondents.

Many statistical agencies may also have standing advisory bodies such as NASS's Advisory Committee on Agricultural Statistics that may provide input relevant to questionnaire content. It was critical for the input from each set of experts to be obtained early in the questionnaire evaluation process. Changes based on expert input may require more subsequent testing, as expert knowledge may or may not reflect respondents' knowledge. Appropriate time must be set aside for this in the testing schedule. A clear benefit to the use of experts is that their specific specialized knowledge can be obtained using relatively few resources. Like wasabi, it should be used judiciously and as an accompaniment to the other ingredients in your bento box, as their results are often idiosyncratic (Olson 2010).

#### 2.4. Cognitive Interviews

Cognitive interviews are routinely used in survey questionnaire development and testing (Willis 2004, 2015). Like the ubiquitous California sushi roll in the bento box, these are staples in many questionnaire evaluations. For establishment surveys, such as the COA, traditional cognitive interview techniques may have to be modified to probe usual reporting processes and the use of records as well as technical terminology (Stettler et al. 2001). Cognitive interviews for the 2017 COA were scheduled as soon as draft questionnaires were available.

Areas identified in the historical data and expert reviews were specifically targeted (McCarthy 2013). For example, the sections collecting information on land had historically been problematic, and these sections continued to top the list in number of edits and number of requests for help on the toll-free telephone line. Thus, these were targeted in the cognitive testing. In addition, the operator demographics section, with the new content suggested by the expert panel, was also targeted. Iterative rounds of cognitive interviews were scheduled so findings from the early testing could be incorporated into later testing, but were limited by staff availability and delays in finalizing some of the questionnaire content.

Because new content was integrated into the questionnaire, the layout and ordering of some sections was also altered from 2012. Several alternative versions of the questionnaires were also developed to test different orders of the content and alternative formats of some sections. Testing was also needed to make sure that these changes did not have any unintended negative consequences.

NASS conducted an unusually large number of cognitive interviews so that a wide variety of agricultural establishments and the proposed alternatives of the draft questionnaire could be tested. As in many establishment surveys, establishments in different subsectors may answer questions differently, have disparate reporting processes, or interpret terminology differently. Selecting several establishments from different subsectors can increase the size of cognitive interview projects significantly. In our testing, operations raising different types of livestock and crops, as well as those with specific types of activities, such as marketing through production contracts, were targeted.

Because of the lengthy nature of the COA questionnaire (34 sections on 24 pages, see [Appendix 1](#)), cognitive interviews also targeted subsets of sections of the questionnaire. Specific subsets of between 8 and 14 sections were included in each interview in order to keep interviews to manageable lengths. Interviewers recruited participants and interviews were conducted in person at the respondents' locations. The questionnaire was NOT mailed (as in the COA) but provided to the respondent in person. Because most COA questionnaires are self-administered, respondents were provided a paper questionnaire by the interviewer. They were asked to complete it on their own without the interviewer's assistance before the cognitive interview began, as if they had received it in the mail. Then the interviewer followed up with both scripted and emergent probes.

The first round of cognitive interviews for the COA began in August 2015 with the goal of finalizing the questionnaires for field testing in early 2016 (described below). Thirty seven interviews were conducted with respondents with specific characteristics (growing specific crops or livestock, participating in government programs, etc.). Interviews were conducted in 18 states by 26 survey methodologists and trained statisticians. Findings from this initial round resulted in changes to the questionnaires that were included in subsequent testing. An additional set of nine interviews was also conducted during this round, with a specific focus on the changes in the personal characteristics section.

Another round of 71 cognitive interviews with respondents was conducted in early 2016. Thirty eight trained interviewers conducted interviews in 32 states. This round tested sections of the questionnaire and types of operations that were not tested in the initial round, a short-form version, and alternative formats that were not included in the initial round. For example, in the initial round of cognitive testing, no respondents reported fruit or berries, so none of the questions in the sections collecting data on these crops were tested. These types of respondents were specifically recruited in the second round. Again, as in the initial rounds of cognitive interviews, the interviews provided qualitative information about how respondents were interpreting questions and formulating their answers.

A final round of 20 interviews was conducted by 14 trained interviewers between March and May 2016. This round of cognitive interviews focused on the variations in how commodities were preprinted in tables or in listings in the commodity sections.

Ideally, multiple rounds of cognitive testing would be conducted before using near-final questionnaires in the field test. Unfortunately, only results from the initial round of

cognitive testing were available in time to incorporate their findings into the field test. However, results from the second and final rounds of testing were evaluated in parallel with the field test and ultimately incorporated into the 2017 COA forms.

### 2.5. *Field Test – 2015 Content Test*

The largest part of our COA testing bento box was the 2015 COA Content Test, a large field test, using the questionnaires resulting from the earlier historical data review, expert reviews and initial cognitive testing. Like the main dishes in a bento box, field tests come in different varieties, to suit the tastes of the survey organization. Unlike cognitive interviews with small idiosyncratic samples, field tests can provide generalizable results with quantitative evidence resulting from more realistic survey field conditions (Dillman and Redline 2004; Fowler 2004).

The Content Test had several objectives: to verify that respondents would complete the questionnaires and provide data as expected in each section, to test new questions under production conditions, to test (with respect to both data quality and response rates) alternative versions of the questionnaire that differed in format and layout using embedded split samples, to evaluate the performance of a proposed short form, to test the utility of a new experimental pre-census mail contact, to test a redesigned web version of the questionnaire, and provide data to be used in the initial COA imputation donor pool (Ott et al. 2016; McCarthy 2017). The test emulated the operational census procedures and schedule to the extent possible.

Unlike cognitive interviews, which include interviewers and motivated respondents, the field test provided data which were collected with procedures similar to the COA and should more closely resemble production COA data. In addition, the large sample size allowed us to examine errors that might arise in rare circumstances or for only a small percentage of respondents. However, unlike the cognitive interviews, the field test yielded little information about why errors occurred.

A stratified random sample of approximately 30,000 operations was selected from NASS's list frame. The sample was not selected to produce population statistics, but to include a set of respondents to allow testing of all sections of the questionnaire. As is common for establishment survey frames, NASS has descriptive information about most of the units on the sample frame. Thus, the sample was selected to include all regions of the country, and a range of sizes and types of operations. Specific minimum numbers were also sampled for types of operations in order to ensure that all sections of the form would have some respondents reporting those data.

Because this is an establishment survey, several concessions were made in selecting the sample. As is typical in establishment surveys, some agricultural operations may be included in many independent survey samples. For example, large agricultural operations may be selected with high probability (or certainty) in all surveys targeting the commodity they produce. Any operations who were included in other NASS surveys in the same data collection time frame were not eligible for the sample, in order to reduce their response burden. This included all known organic operations, since NASS was also conducting a census of those operations at that time. In addition, some large establishments have special handling arrangements for their data collection due to their importance for specific

estimates in other surveys or the need for their ongoing cooperation. These establishments were removed from the sample.

The field test schedule mimicked the COA timeline so data for the calendar year could be collected and seasonal effects on response would be similar. An initial mailing of a questionnaire and cover letter was sent to 29,740 operations in January 2016, collecting information for the 2015 calendar year. A reminder postcard followed two weeks later and a second questionnaire mailing two weeks after that. An online questionnaire was also available for response. This was also followed by CATI telephone follow-ups for nonrespondents. Conducting testing using realistic survey procedures may also be particularly important for establishment surveys. While cognitive interviews yield useful information, we had several cognitive interview respondents tell interviewers that they did not bother to consult records during the cognitive interview but would have used them to report in the COA itself.

The field test also allowed us to randomly assign treatment groups within the sample to directly compare the performance of a number of questionnaire versions. This included alternative formats for some questions and a 'short form'. A subset of large operations was also selected to evaluate whether a presurvey mailing to collect contact information for the operation increased response rates or was useful for early identification of COA nonrespondents (McCarthy 2017). While the procedures were similar to those planned for the 2017 COA, there were a few notable differences. For example, the 2017 COA carries mandatory reporting authority, but the field test did not. In addition, there was no final in-person nonresponse follow up. Finally, no population estimates were generated from the data, so no processing, such as weighting or summarization, was included in the field test.

Overall, 18,864 responses were gathered for evaluation. The field test results included comparisons of response rates for alternative questionnaire versions and a review of errors (such as sub-items not adding to totals, other inconsistent answers, item nonresponse and outliers or unusual answers) both overall and across versions (Ott et al. 2016). Results were used to determine the final 2017 COA questionnaires and to identify any items with unacceptable error rates.

A second smaller field test was also conducted the following year to collect data for 2016. The objective of this test was only to evaluate the online questionnaire, since the paper questionnaire had already been finalized and provided for printing companies to bid on the print contract. Changes were still possible for the online instrument, so additional usability testing and a smaller field test were conducted. For this test, survey invitations with instructions for accessing the online form were mailed and responses were collected solely through the online questionnaire. The test resulted in additional recommendations for changes to the online form.

## 2.6. Qualitative Follow up Interviews

Similar to response analysis studies (Phipps et al. 1995), respondents can be asked follow-up questions about their existing survey responses. In many establishment surveys where much is known about sample units, it can be easier to identify records that report unexpected or inconsistent data for follow up. Unlike response analysis studies, which typically recontact randomly selected respondents, we identified responses where further

information from respondents could be helpful in examining problems. Because these interviews were quite short and respondents were not easily accessible for in-person interviews (i.e., farmers in rural locations), these interviews were all conducted by telephone. Twenty operations reporting unexpected data in the field test were recontacted by telephone and asked to provide more detail about their answers. For example, all known organic operations were excluded from the field test, as an organic survey was also being conducted at the same time and additional burden of the COA field test was deemed too high. However, some field test respondents unexpectedly reported organic production. In order to verify that these respondents were not reporting in error due to the questions on the COA questionnaire, several were recontacted and asked to provide more information about their answers. Their responses indicated no major problems with the questions, but were mainly operations that were not USDA *certified* organic producers and thus may be less likely to be known to NASS as organic producers. This information about the validity of their answers could not be obtained from the field test data alone. Several other responses with suspect field test data were also recontacted to collect additional information.

### *2.7. Planning the Testing Bento Box*

In order to plan a bento box of questionnaire evaluation, both the strengths and weaknesses of the methods and how to schedule them together need to be considered. In addition, each method should be planned with respect to whether information from one method will be used as input or in parallel to another. For our bento box, the field test had to be conducted at the end of a calendar year. The test could not be fielded before new content was finalized and questionnaires were drafted (2015) and results had to be available before the deadline for questionnaires to be finalized for the 2017 COA questionnaire print contract (late 2016). Thus, the field test had to be conducted at the end of 2015, continuing through early 2016. Given this, the other methods of testing were scheduled to work around the field test schedule.

As illustrated in the descriptions above, each of the methods we used in testing provided different information and had its own strengths and weaknesses. These are summarized in [Table 1](#) below.

## **3. Selected Results of Bento Box Testing in the COA**

While each of the methods described above can provide useful information for questionnaire evaluation, when combined, their results can be much more powerful. Multiple methods can help compensate for some methods' weaknesses or amplify their strengths. Several examples will serve to illustrate this.

### *3.1. Example 1: Combining Expert Review, Cognitive Testing, and Field Testing*

Based on data from the 2012 COA, 30 percent of all operators were women, and only 14 percent of principal operators were women. Following the publication of the 2012 COA, NASS received feedback from the agricultural and government sectors with concern that

Table 1. Summary of the questionnaire testing bento box methods.

Method	When to use	Strengths	Potential weaknesses
Historical data review	Early; As soon as previous collection's data is available	<ul style="list-style-type: none"> <li>• Can be used to focus smaller scale testing</li> <li>• Provides quantitative measures</li> <li>• Reflects operational procedures/conditions</li> <li>• Can fully represent population</li> <li>• Can allow examination of small subpopulations</li> <li>• Inexpensive</li> </ul>	<ul style="list-style-type: none"> <li>• Little information on reasons for problems</li> <li>• Cannot examine changed or new content</li> </ul>
Expert review	Early; To allow iterative testing and later respondent testing	<ul style="list-style-type: none"> <li>• Can leverage specialized knowledge</li> <li>• Can broaden perspective</li> <li>• Inexpensive</li> </ul>	<ul style="list-style-type: none"> <li>• Idiosyncratic results</li> <li>• May only identify 'major' problems</li> <li>• Limited sphere of consideration</li> <li>• May not provide enough detail for questionnaire revisions</li> <li>• May not consider respondent perspective</li> <li>• May not consider operational constraints</li> </ul>
Cognitive interviews	Once questionnaire drafts are available	<ul style="list-style-type: none"> <li>• Provides rich insight into respondent processes</li> <li>• Can suggest solutions for improvements</li> <li>• Inexpensive compared to field testing and post-field editing (although may be less so for establishment surveys)</li> </ul>	<ul style="list-style-type: none"> <li>• Requires specially trained staff</li> <li>• Does not provide measure of size or extent of problems</li> <li>• Does not replicate survey conditions</li> <li>• Samples may not be representative and may not include important subgroups</li> <li>• Hard to find rare items in small samples</li> </ul>

Table 1. Continued.

Method	When to use	Strengths	Potential weaknesses
Field test	Late; Questionnaires and procedures should be near final	<ul style="list-style-type: none"> <li>• Emulates survey conditions</li> <li>• Provides quantitative measures</li> <li>• Can include wide variety of respondents</li> <li>• Split samples can directly compare alternative questionnaires</li> <li>• Can examine rare events</li> </ul>	<ul style="list-style-type: none"> <li>• Errors producing 'reasonable' data will not be detected</li> <li>• Does not provide reasons for errors</li> <li>• Expensive</li> </ul>
Qualitative follow-up interviews	Late; as a follow up to field tests or other data collection	<ul style="list-style-type: none"> <li>• Can target identified problems</li> <li>• Examines data collected in operational conditions</li> <li>• Provides insight into respondents' cognitive processes</li> <li>• Inexpensive</li> </ul>	<ul style="list-style-type: none"> <li>• Requires skilled interviewers</li> <li>• Includes small samples</li> <li>• Does not provide measure of size of problems</li> <li>• May occur too late in process for questionnaire changes</li> </ul>

the role of women and new/beginning farmers was not being adequately measured in the COA and survey programs.

NASS convened an independent panel consisting of 13 experts from academia, government, and agricultural sectors to evaluate how NASS measures the contribution of women and new/beginning farmers (Ridolfo 2016). Stakeholders and the general public had opportunities to provide public comments or to present their points of view in person. In April 2015, the panel met in person and reviewed the operator characteristic information currently published by NASS. It provided guidance on how to improve the reporting of women and new/beginning operators in the 2017 COA. The panel met two subsequent times following the iterative rounds of cognitive testing to provide further guidance on the modification of the 2017 COA questionnaire.

The panel identified two issues that could impact questions to collect data about women and new/beginning farmers. First, the evolution of farm structure into complex entities, both large and small, has created problems for assessing the roles of individuals in the operation. Responsibilities are often divided amongst several individuals whether the operation is a small family farm or a large corporation. The COA must be able to capture the full range of individual responsibilities within these small and large operations. A second, even more challenging issue raised by the panel was the cultural norms in the farming community that reinforce the perception that farm operators are men. Cognitive testing at NASS has found that the oldest, male family member is viewed as the principal operator even if this individual has limited or no day-to-day involvement in farming, while the roles of women often go unreported in the COA and other surveys.

The panel provided specific recommendations for changes to the COA questionnaire that would address the key issues detailed above. First, the panel recommended defining operators in terms of function rather than titles. The panel argued that titles such as ‘operator’ and ‘principal operator’ do not have universal definitions and are open to interpretation. Given the deference to the oldest, male family members when reporting operators, it was recommended that NASS define operators based on function (i.e., involvement in decisions pertaining to the operation) and allow for joint involvement. Furthermore, they recommended that the COA ask about involvement in a variety of decisions, not just day-to-day decisions. The panel also recommended that respondents be allowed to report up to four operators on the COA. Finally, the panel recommended transitioning away from the concept of a single principal operator. Although they recommended the retention of this measure for bridging purposes, they suggested that respondents be allowed to report more than one principal operator.

Questions were drafted based on the expert panel recommendations, but it was unknown whether respondents would interpret and answer these questions as intended. Once drafted, they could be administered to respondents in cognitive interviews to test this. For the initial cognitive interviews, all labels that the panel identified as problematic (i.e., operator, principal operator) were removed and ‘day-to-day’ decisions was removed from the first question (see [Figure 2](#)).

During cognitive testing, it was found that when answering Question 1 (“In 2015, how many individuals were involved in the decisions for this operation?”), respondents reported individuals with a broad range of involvement in decisions. While reporting was

SECTION 6 PERSONAL CHARACTERISTICS							
1. In 2015, how many individuals were involved in the decisions for this operation (include family members and hired managers)? Enter the total number of individuals and the number of women. Exclude hired workers unless they were a hired manager or family member. . . . 1575	<table border="1"> <thead> <tr> <th colspan="2">Number of Individuals Making Decisions</th> </tr> <tr> <th>Total</th> <th>Women</th> </tr> </thead> <tbody> <tr> <td></td> <td>1574</td> </tr> </tbody> </table>	Number of Individuals Making Decisions		Total	Women		1574
Number of Individuals Making Decisions							
Total	Women						
	1574						

Fig. 2. Revised questionnaire used in cognitive testing.

not consistent across respondents, information was being collected on a broader range of people than just those making day-to-day decisions.

Although the scope of individuals reported was broadened, undercounts of women were still present. During the cognitive interviews, some male respondents discussed how women were involved in the decisions regarding the farm but they did not report them on the questionnaire. When asked why, they gave reasons in alignment with cultural norms that subordinate women and their role on the farm, such as “we (the men) have the final say,” or “she is only involved in certain decisions.” In one case, a woman was reluctant to report herself despite her involvement in decisions for the farm, stating “the Bible says he’s in charge.” These findings indicated that these revisions to the questionnaire do not completely circumvent norms that preclude the reporting of women’s contributions to agriculture.

An unexpected finding from the cognitive testing was that women were offended by the formatting of Question 1. They felt that asking about women (but not men) separately highlighted the fact that women are a minority in agriculture. They felt the question should ask for the number of men and women. Other respondents assumed that this question was asking for the number of men and women and reported the number of men in the “total” answer space.

In the third round of testing, “women” was removed from Question 1 and a second question, which asked for the total number of men and women who were involved in decisions for the operation was added (see Figure 3).

In the final round of cognitive testing, this change worked well. Responses to Question 2 summed to responses given to Question 1 the majority of the time. Respondents with no female operators sometimes left the “women” response space blank. One limitation to this question was that some respondents felt Questions 1 and 2 were redundant. However, in all, this question was an improvement to what was used in 2012, because respondents were thinking of more than just individuals who are involved in the day-to-day decisions and were including a wider variety of individuals who are involved in various decisions for the operation.

The panel also recommended that NASS ask about individuals’ involvement in a variety of decisions. A series of new questions suggested by the expert panel was added

SECTION 6 PERSONAL CHARACTERISTICS					
1. In 2015, how many individuals were involved in the decisions for this operation (include family members and hired managers)? Exclude hired workers unless they were a hired manager or family member. . . 1575	<table border="1"> <thead> <tr> <th>Total</th> </tr> </thead> <tbody> <tr> <td></td> </tr> </tbody> </table>	Total			
Total					
2. Of the total number of individuals making decisions for this operation, how many are men and how many are women? . . . 1571	<table border="1"> <thead> <tr> <th>Men</th> <th>Women</th> </tr> </thead> <tbody> <tr> <td></td> <td>1574</td> </tr> </tbody> </table>	Men	Women		1574
Men	Women				
	1574				

Fig. 3. Final round cognitive test version.

to the questionnaire to capture individuals' varying levels of decision-making involvement. Respondents were asked to report how much involvement each person had in specific decisions. Cognitive testing indicated that the length of the section was burdensome, respondents had difficulty with the response options (All, Some, None, or Not Applicable) and respondents felt that some of the types of decisions were redundant. These cognitive interview findings prompted changes to the expert panel's suggested questions.

Following the cognitive interviews, modified questions (with modified response options) were included in the field test. Review of the field test responses showed that data reported in this section appeared reasonable and respondents appeared to complete the new questions correctly.

One of the experimental comparisons in the field test was between forms with the operator demographic questions near the front or back of the questionnaire. Considered sensitive, NASS was concerned that if these questions appeared early in the questionnaire, response rates or item nonresponse would be worsened. The field test showed that overall response was comparable for both versions of the form, but that item response rates were higher when this section appeared early (McCarthy 2017).

In addition, the field test data in this section was reviewed and compared to the information reported by the field test respondents in the 2012 COA. Results showed that the field test respondents reported more operators than they had in 2012, both men and women. While not accounting for changes in operations between 2012 and 2015, the data was reviewed to see if more operators were reported than with the previous questions. The field test respondents reported an average of 1.27 operators per operation in 2012. These same respondents reported an average of 1.52 operators in the 2015 field test with the modified questions. The mean number of women operators reported increased from .40 to .54.

While our expert panel was able to suggest possible ways to increase the reporting of women operators, cognitive interviews were needed to modify their suggested questions so respondents could answer them appropriately. The interviews also confirmed the expert panel input that strong cultural norms might still limit reporting of women. The subsequent field test showed empirically that broadening the scope of "operators" prompted the reporting of additional people, both men and women. It also provided evidence for the optimal placement of this section. Using these three elements of the bento box together gave us confidence that the new and modified questions will perform well for 2017.

### *3.2.. Example 2: Combining Expert Review, Cognitive Testing, Field Testing, and Follow up Interviews in Development and Testing of a Short Form*

Since the COA uses a lengthy 24-page questionnaire, NASS was also interested in developing a short form to reduce respondent burden, reduce costs or decrease nonresponse. Many of the sections of the full 24-page questionnaire do not apply to individual respondents if they do not produce specific types of commodities or have specific characteristics. Based on previous data collected or other auxiliary information, NASS can identify specific operations that should or should not receive a short form. For example, an operation that has historically only grown corn can receive a questionnaire

that does not ask detailed questions about fruit, berries, Christmas trees, aquaculture, hogs, and so on. In addition to reducing respondent burden, reduced mail and processing costs for a short form would save money. The possibility that a short form could increase response rates was also of interest.

In order to develop and evaluate a short form, NASS worked with survey experts at the Washington State University (WSU). WSU survey methodologists developed two versions of a short form and detailed specifications for how to select operations eligible to receive it (Moore et al. 2016). The 24-page questionnaire was reduced to either 16 or 12 pages by combining the commodity listings across different sections and reducing the amount of space available to provide answers. In addition, they made visual design changes to the questionnaire.

Both of these test questionnaires were tested by WSU in a field test, which included mailed questionnaires similar to COA data collection procedures. The WSU field test showed that comparable response rates were obtained for each of the two shortened questionnaires. That is, reducing the number of pages from 16 to 12 did not increase response rates. With respect to data quality, comparisons of the two questionnaires showed that the shortest form prompted slightly lower data quality as measured by missing data, reduced reporting of some items, and missing codes. This is most likely due to the fact that less information is preprinted on the form and respondents must write in more information on this version of the form. Based in part on this research, NASS decided not to pursue the twelve-page short form.

Building on the development and testing from the WSU experts, NASS continued development of a short form. A short form was designed for the field test by removing sections of the questionnaire for which selected respondents would not report (vegetable, fruit, or aquaculture production, organic agriculture, hired labor, etc.). In order to verify that respondents did not have these items, simple questions about their presence/absence were included (Figures 4 and 5). These questionnaires were then tested in 40 cognitive interviews. The cognitive interviews showed that respondents appeared to be able to report accurately and also answered the presence/absence questions as expected (typically reporting that they did not have the items).

The short form was then included in the 2015 Content Test. The objectives of this field test related to the short form were to evaluate the data collected with this questionnaire, measure the response rate relative to the longer questionnaire, and evaluate whether NASS was able to accurately identify respondents who should receive the short form. In order to receive a short form in the COA, operations that were known to have the items removed from the short form were excluded. In other words, operations that were known to have fruit, vegetables, berries, nursery, Christmas trees, maple, sheep, goats, bees, hogs, aquaculture, specialty livestock, labor or renewable energy systems were NOT eligible to receive the short form. Any short form respondents in the COA who indicated they did have these items would have to be recontacted to collect the detailed information about them missing from the short form.

The field test data confirmed that the short form did *not* significantly increase response with comparable split samples' response rates for mail returns. Two subgroups of short form eligible operations were sent versions of a short form and another two subgroups were sent the long form. Response rates for the groups are shown in Table 2 below.

**SECTION 11 FRUITS AND NUTS**

1. Were any fruit or nut trees, including grapevines, on this operation in 2015? Report berries in SECTION 12.

**INCLUDE**  
• crops grown under contract

**EXCLUDE**  
• abandoned acres  
• home garden, personal or home use crops

1047 1  Yes - Complete this section      3  No - Go to SECTION 12

Total Acres		Acres Irrigated	
Acres	Tenths	Acres	Tenths

2. Acres in bearing and nonbearing fruit orchards, vineyards, and nut trees ..... 0121

3. Report gross value of fruits and nuts sold from this operation in 2015. Include the value of your landlord's share, marketing charges, taxes, hauling, etc. Exclude dollars for items produced under production contracts ..... 1329  Mark "X" if None

Gross Value of Sales (Dollars)
\$ .....00

4. Fill in the columns below for all fruits and nuts on this operation in 2015. For those commodities not printed in the table, enter the crop name and code from the list below or the commodity listing and codes in the instruction booklet.  
• Include acres even if not harvested because of low prices, damage from hail, frost, etc.

Enter Fruit or Nut Name	Enter Code	Total Acres		Bearing Age Acres		Nonbearing Age Acres	
		Acres	Tenths	Acres	Tenths	Acres	Tenths
Apples	0123						
Grapes	0366						
Peaches	0228						
Pears	0246						
Pecans, improved	0336						
Pecans, native and seeding	0342						
Other fruits and nuts, specify below:							

If more space is needed, use a separate sheet of paper.

NON-CITRUS FRUIT	CODE	CITRUS FRUIT	CODE	NUTS	CODE
Apricots .....	0129	Grapefruit .....	0267	Almonds .....	0321
Cherries, sweet .....	0345	Lemons .....	0279	Chestnuts .....	0324
Cherries, tart .....	0587	Oranges, Valencia .....	0150	English walnuts .....	0357
Figs .....	0165	Oranges, other than Valencia .....	0174	Hazelnuts (Filberts) .....	0327
Kiwifruit .....	0189	Tangelos .....	0303	Other nuts, specify above .....	0363
Nectarines .....	0201	Tangerines .....	0309		
Persimmons .....	0237	Other citrus fruit, specify above .....	0315		
Pomegranates .....	0255				
Other non-citrus fruits, specify above .....	0261				

Fig. 4. Example of long form section removed from the short form.

The field test also provided vital information about the accuracy of the selection of the short form eligible members. In the field test, the percentage of respondents who reported the presence of the items in the sections removed from the questionnaire could be evaluated. The two items reported at the highest rate were the use of hired labor at almost five percent and vegetable production at six percent. Given the rarity of these positive reports, it was not surprising that we did not find this problem in the cognitive interviews.

1. Were any of the following crops grown or harvested on this operation in 2015?			
a. Vegetables, potatoes, sweet corn, or melons? . . . . .	1101	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
b. Fruit or nut trees, including grapevines? . . . . .	1047	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
c. Strawberries or other berries? . . . . .	1041	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
d. Nursery, floriculture, or greenhouse crops, including ornamental plants, flowers, mushrooms, aquatic plants, sod, food crops under protection, vegetable seeds, flower seeds, or other propagative materials? . . . . .	1032	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
e. Christmas trees or short rotation woody crops? . . . . .	1208	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
f. Maple syrup? . . . . .	1209	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
2. Did this operation own or custom feed for others any sheep, lambs, goats, or kids in 2015, regardless of location? . . . . .			
	1102	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
3. Did this operation own honey bees in 2015, regardless of location? . . . . .			
	1104	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
4. Did you or anyone else have any of the following livestock species on this operation in 2015?			
a. Hogs and pigs? . . . . .	1211	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
b. Aquaculture? . . . . .	1157	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No
c. Any other livestock or livestock products not already mentioned? . . . . .	1239	1 <input type="checkbox"/> Yes	3 <input type="checkbox"/> No

Fig. 5. Presence/absence questions on the short form.

While these percentages are low, the specifications for the population eligible for the short form would have identified almost a million records. If 5–6% of these records must be recontacted for each item, tens of thousands of recontacts would need to be made. This was deemed an unacceptably high number.

The field test also included a subsample of records eligible for the short form that were sent the long form. Data reported in sections of the long form that did not appear on the short form gave an indication of the detailed data that would have been missed. While much of the data was for small amounts, there were several large reports, such as 10,000 hogs, 240 acres of fruit, and so on.

This information was a critical finding from the field test that was not possible to learn from small-scale cognitive interviews. The field test information showed that problems still existed, but did not provide the explanatory information for them. To provide more information on field test respondents reporting these unexpected items, a small sample of these short form respondents who reported presence of unexpected commodities were recontacted in follow up interviews.

Survey methodologists from the testing team conducted these semi-structured interviews. Respondents were asked to verify their answers and then provide more detail on the unexpected commodities. Four respondents who reported vegetables were contacted and stated that they had home gardens or vegetables grown only for their own

Table 2. Response rates for questionnaires sent to short form eligible respondents.

Form	Response rate (%)
Short Form ( $n = 4,911$ )	61.05
Long Form ( $n = 4,887$ )	60.47

$$X^2(1, N = 9,798) = .35, p = .56.$$

home consumption. Home gardens should not be reported on the COA, but the short form did not include the instruction to exclude these types of gardens (these instructions do appear on the longer questionnaire). Even with a small, specially selected group of recontacts, it was clear that the missing instructions were prompting errors. Many of the short form eligible respondents who reported small amounts of vegetables when mailed a long form were likely reporting home gardens. Based on the field test combined with the follow-up interviews, the short form instructions were modified. The decision was also made to change the presence/absence questions to questions collecting the number or acres of the commodities. Based on the field test results, the farm labor section was also put back onto the short form used in the 2017 COA.

Survey design experts provided initial input in the development of a short form. Cognitive interviews then provided initial insight into how respondents answered the short form. But the field test provided quantitative measures of how the short form performed with a much larger and diverse set of respondents. In the cases where the short form produced unexpected reports of commodities, the follow-up interviews provided clear information on why we got these results and how we might make improvements to the form. Because these interviews could be targeted, a few interviews provided crucial insights. Together, these elements of the bento box were far superior to each one alone.

#### 4. Conclusion and Lessons Learned

As in a well-balanced meal, in a large and complex data collection, combining diverse ingredients in questionnaire testing can provide better results than any one type in isolation. A bento box approach yields complementary information and greater confidence in questionnaire revisions. Each element of the bento box – the rice, the sushi, the salad, the grilled meat – brings certain elements to the meal and lacks others – so using them together can help overcome any methods' individual weaknesses. Some methods will provide information across a broader swath of respondents but with little detail, filling you up, but offering few nutrients. Others will provide smaller portions with much more vibrant flavors – less breadth but a deeper look with richer information.

Each testing method has its own strengths, and its weaknesses can be supplemented with the strengths of the others. The examples discussed above illustrate how information from multiple kinds of questionnaire evaluations can be combined. Still other methods can be added to the testing and evaluation bento box ([Office of Management and Budget 2016](#)), such as focus groups, interaction behavior coding, validation studies, latent class analysis, usability testing, review of paradata ([Couper 2008](#)), and others. Other survey organizations are increasingly using multiple testing methods to triangulate results ([Persson et al. 2015](#) and [Willimack 2013](#)). NASS has found the methods described in this case study to be a particularly effective set of ingredients to use in combination.

In addition, some of the considerations discussed may prove particularly useful in testing establishment survey questionnaires. The specific differences between household and establishment populations, such as the importance of large units or the variety in specific characteristics, may be important. Expert input may be more important for establishments if technical terminology is involved. We also found that the use of records

was likely different across different testing methods. There are also special considerations for respondent selection in establishment survey testing. We were able to select our samples using extensive information known about the establishment size, location, previous commodities raised, and so forth. Key subgroups of respondents were specifically targeted. However, we specifically excluded establishments that had been selected in other surveys or that may have had existing special data collection handling.

The multiple methods used for the 2017 COA, implemented sequentially and in parallel, proved to be complementary, balancing out the shortcomings of each testing method. In our testing, each evaluation method provided evidence of potential problems and areas for improvement. However, using evidence from multiple sources provided us with more confidence that we had correctly identified their source and scope, or a deeper understanding of our data quality.

While we consider our testing efforts a success, we did learn some lessons that may help others who are considering multi-method questionnaire testing. As for all large testing projects, planning is critical to success. Evaluation of existing data can (and should) be started as soon as data is available to allow maximum time for questionnaire design and testing. Larger scale elements like the field test had the strictest scheduling constraints, so these can be scheduled first, and most of the other evaluation methods can be scheduled around that. While we planned to use results from early testing as inputs for subsequent rounds of testing, we did not always allot enough time to thoroughly evaluate results before other methods had to be fielded. In addition, we should have identified, from our early testing, more areas where the late qualitative follow-ups would be useful. In addition, with multiple methods, more resources are required for implementation and evaluation.

In such a large and complex data collection as the COA, the resources spent to test and retest our questionnaires will be paid back in better data quality from our respondents and less staff time needed in data editing, processing, and analysis. Using multiple methods gave us different kinds of information and a fuller picture of how our questionnaires were working. While the COA is a large data collection, the methods described can be combined for surveys of any size. Taken together, the multiple questionnaire evaluation elements of our bento box provided a much more satisfying meal than any of the evaluations consumed alone.

*Appendix 1 Sections Included in the COA Questionnaire.*

Face Page
Acreage in 2015
Land
Land Use Practices
Government Programs
Type of Organization
Personal Characteristics
Field Crops
Hay and Forage Crops
Cultivated Christmas Trees, Short Rotation Woody Crops, and Maple Syrup
Nursery, Greenhouse, Floriculture, Sod, Mushrooms, Vegetable Seeds, and Propagative Materials
Vegetables, Potatoes, and Melons
Fruits and Nuts
Berries
Cattle and Calves
Hogs and Pigs
Equine
Sheep and Goats
Aquaculture
Poultry
Colonies of Honey Bees
Other Livestock and Livestock Products
Organic Agriculture
Production Contracts and Custom Feeding
Practices
Marketing Practices
Fertilizers, Chemicals, and Soil Conductions Applied
Energy
Market Value of Land, Buildings, Machinery, and Equipment
Machinery and Equipment
Income from Farm-Related Sources
Farm Labor
Production expenses
Agricultural Activity within the Borders of American Indian Reservations, Pueblos, and Service Areas
Conclusion

## 5. References

- Couper, M. 2008. *Designing Effective Web Surveys*. Cambridge: Cambridge University Press.
- Creswell, J.W. 2014. *A Concise Introduction to Mixed Methods Research*. Thousand Oaks, CA: Sage Publications, Inc.
- Dillman, D. and C. Redline. 2004. "Testing Paper Self-Administered Questionnaires: Cognitive Interview and Field Test Comparisons." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer, 299–317. New York: Wiley and Sons, Inc.
- Edwards, W.S. and D. Cantor. 1991. "Toward a Response Model in Establishment Surveys." In *Measurement Errors in Surveys*, edited by P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman, 211–233. New York: Wiley and Sons, Inc.
- Forsyth, B., J. Rothgeb, and G. Willis. 2004. "Does Pretesting Make a Difference? An Experimental Test." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer, 525–546. New York: Wiley and Sons, Inc.
- Fowler, F. 2004. "The Case for More Split-Sample Experiment in Developing Survey Instruments." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer, 173–188. New York: Wiley and Sons, Inc.
- Groves, R.M., F.J. Fowler Jr, M. Couper, J. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology*. New York: Wiley & Sons, Inc.
- Madans, J., K. Miller, A. Maitland, and G. Willis. 2011. *Question Evaluation Methods: Contributing to the Science of Data Quality*. New York: Wiley & Sons, Inc.
- McCarthy, J. 2013. "Getting Your Money's Worth: Targeting Resources to Make Cognitive Interviews Most Effective." In *Proceedings of the Section on Survey Research Methods: American Statistical Association*. Alexandria, VA: American Statistical Association. Available at: <http://ww2.amstat.org/sections/srms/Proceedings/> (accessed November 2017).
- McCarthy, J. 2016. "Planning Your Multi-method Questionnaire Testing Bento Box: Examples from the 2017 Census of Agriculture Testing." Paper presented at the 2016 International Conference on Questionnaire Design, Development, Evaluation, and Testing (QDET2). Miami, Florida. Available at: <https://ww2.amstat.org/meetings/qdet2/OnlineProgram/Program.cfm> (accessed November 2017).
- McCarthy, J. 2017. "Multi-use Field Testing: Examples from the 2017 Census of Agriculture Dry Run." In *Proceedings of the Section on Survey Research Methods: American Statistical Association*. Alexandria, VA: American Statistical Association. Available at: <http://ww2.amstat.org/sections/srms/Proceedings/> (accessed November 2017).
- McCarthy, J. and D. Buysse. 2010. "Bento Box Questionnaire Testing: Multi-Method Questionnaire Testing for the 2012 Census of Agriculture." In *Proceedings of the Section on Survey Research Methods: American Statistical Association*. Alexandria,

- VA: American Statistical Association. Available at: <http://ww2.amstat.org/sections/srms/Proceedings/> (accessed November 2017).
- Moore, D., K. Ott, and A. Gertseva. 2016. "Developing and Evaluating a Short Form: Results and Recommendations from Tests of a Form Designed to Reduce Questionnaire Length." In *Proceedings of the Fifth International Conference of Establishment Surveys*, June 20–23, 2016. Geneva, Switzerland: American Statistical Association.
- Office of Management and Budget. 2016. *Statistical Policy Working Paper 47: Evaluating Survey Questions: An Inventory of Methods*. Washington, DC.
- Olson, K. 2010. "An Examination of Questionnaire Evaluation by Expert Reviewers." *Field Methods* 22(4): 295–318.
- Ott, K., P. McGovern, and R. Sirkis. 2016. "Using Analysis of Field Test Results to Evaluate Questionnaire Performance." In *Proceedings of the Fifth International Conference of Establishment Surveys*, June 20–23, 2016. Geneva, Switzerland: American Statistical Association.
- Persson, A., A. Bjornram, E. Elvers, and E. Erikson. 2015. "A Strategy to Test Questionnaires at a National Statistical Office." *Statistical Journal of the IAOS* 31: 297–304.
- Phipps, P. S. Butani, and Y. Chun. 1995. "Research on Establishment Survey Questionnaire Design." *Journal of Business and Economic Statistics* 7: 337–346.
- Presser, S., and J. Blair. 1994. "Survey Pretesting: Do Different Methods Produce Different Results?" *Sociological Methodology* 24: 73–104.
- Presser, S., J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer. 2004. *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley and Sons, Inc.
- Ridolfo, H., V. Harris, J. McCarthy, D. Miller, N. Sedransk, and L. Young. 2016. "Developing and Testing New Survey Questions: The Example of New Questions on the Role of Women and New/Beginning Farm Operators." In *Proceedings of the Fifth International Conference of Establishment Surveys*, June 20–23, 2016, Geneva, Switzerland: American Statistical Association.
- Schaeffer, N.C. and J. Dykema. 2004. "A Multiple-Method Approach to Improving the Clarity of Closely Related Concepts." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer, 475–502. New York: Wiley and Sons, Inc.
- Snijkers, G., G. Haraldsen, J. Jones, and D. Willimack. 2013. *Designing and Conducting Business Surveys*. New York: Wiley & Sons, Inc.
- Stettler, K., D. Willimack, and A. Anderson. 2001. "Adapting Cognitive Interview Methodologies to Compensate for Unique Characteristics of Establishments." In *Proceedings of the Section on Survey Research Methods*: American Statistical Association. Alexandria, VA: American Statistical Association. Available at: <http://ww2.amstat.org/sections/srms/Proceedings/> (accessed November 2017).
- Tuttle, A., R. Morrison, and D. Willimack. 2010. "From Start to Pilot: A Multi-method Approach to the Comprehensive Redesign of an Economic Survey Questionnaire." *Journal of Official Statistics* 26(1): 87–103.
- Willimack, D. 2013. "Methods for the Development, Testing, and Evaluation of Data Collection Instruments." In *Designing and Conducting Business Surveys*, edited by

- Snijkers, G., G. Haraldsen, J. Jones, and D. Willimack, 253–301. New York: Wiley and Sons, Inc.
- Willimack, D. and E. Nichols. 2010. “A Hybrid Response Process Model for Business Surveys.” *Journal of Official Statistics* 26(1): 3–24.
- Willis, G.B. 2004. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications.
- Willis, G.B. 2015. *Analysis of the Cognitive Interview in Questionnaire Design*. Oxford: Oxford University Press.
- Yan, T., F. Kreuter, and R. Tourangeau. 2012. “Evaluating Survey Questions: A Comparison of Methods.” *Journal of Official Statistics* 2: 503–529.

Received October 2016

Revised November 2017

Accepted January 2018