Using cognitive interviewing to improve the Women's Empowerment in Agriculture Index survey instruments: Evidence from Bangladesh and Uganda

Hazel Jean L. Malapit*, Kathryn Sproule, and Chiara Kovarik
*Corresponding author: h.malapit@cgiar.org; International Food Policy Research Institute.

Abstract

The purpose of cognitive interviewing is to systematically identify and analyze sources of response error in surveys, and to use that information to improve the quality and accuracy of survey instruments. This paper describes the cognitive interviews undertaken in Bangladesh and Uganda in 2014 as part of the second round of pilots intended to refine the original version of the Women's Empowerment in Agricultural Index (WEAI). The WEAI is a survey-based tool that assesses gendered empowerment in agriculture. Baseline data were collected in 19 countries, but implementers reported some problems, such as confusion among both respondents and enumerators regarding the meaning of abstract concepts in the autonomy sub-module and difficulties recalling the sequence and duration of activities in the time-use sub-module. The results revealed potential problems with the survey questions and informed the revision of the WEAI, called the Abbreviated WEAI (or A-WEAI), which has less potential for response errors.

Keywords
Agriculture; Bangladesh; cognitive interviewing; cognitive testing; survey methodology; Uganda; women's empowerment

Introduction
Cognitive interviewing is a useful methodology that can be used to improve the quality of data collected in survey instruments, particularly tools that are new or under development, such as the Women's Empowerment in Agriculture Index (WEAI). Cognitive interviewing aims to systematically identify and correct problems with survey questions in order to ultimately improve the quality and accuracy of survey instruments (Willis 2005; Beatty & Willis 2007). Even when using well-established and validated questions and survey instruments, it is still worthwhile to conduct cognitive testing when using the tools in a new context or country, to ensure that respondents understand the intent of the questions. As a new tool, the first iteration of the WEAI raised a number of implementation problems and concerns. We used cognitive interviews to address these issues in the piloting of a revised questionnaire, which ultimately led to better ways of measuring women's empowerment. While the focus of this paper is to illustrate how we used this methodology for improving the WEAI, this approach is broadly applicable to any survey instrument.

In this paper, we describe the process of conducting cognitive interviews for the revised WEAI in Uganda and Bangladesh in June and July 2014, and how this process has informed the second version of the WEAI, now called the Abbreviated WEAI (A-WEAI). [2] We begin with a brief introduction to the WEAI and the implementation problems and concerns associated with its roll-out, an overview of the methodology of cognitive interviewing, the different sources of response errors, and how cognitive interviewing can be used to discover and correct potential problems with survey questions. This is followed by a discussion of our sampling strategy and

implementation approach. Next we present our findings, organized by WEAI domain, and the subsequent changes in the survey questions that resulted from these findings. Finally, we summarize the key lessons that we can take away from this process, which may be useful to others who are developing and testing standardized cross-national survey instruments.

Background on the WEAI

Overview

Launched in 2012, the Women's Empowerment in Agriculture Index (WEAI) is the first comprehensive and standardized survey-based measure to directly capture women's empowerment and inclusion in the agricultural sector. The International Food Policy Research Institute (IFPRI), in collaboration with the Oxford Poverty & Human Development Initiative (OPHI) and the United States Agency for International Development (USAID), designed the WEAI as a monitoring and evaluation tool for the US government's Feed the Future (FTF) initiative, operating in 19 geographically and culturally diverse focus countries. The WEAI was piloted in Bangladesh, Guatemala, and Uganda in 2011 and was included as part of the baseline population-based surveys rolled out by FTF in its 19 focus countries. The WEAI is constructed using information collected from the self-identified primary male and female adult decisionmakers, aged 18 and over, in the same household.

As described in Alkire *et al*., (2013), the WEAI is an aggregate index that is reported at the program or country-level, and is a weighted average of two sub-indexes: the five domains of empowerment (5DE), and the gender parity index (GPI), weighted at 90 percent and 10 percent, respectively. The 5DE assesses the extent to which women are empowered across five domains, which include (1) production, (2) resources, (3) income, (4) leadership, and (5) time. An individual's achievements in these five domains are measured by a set of 10 indicators with their corresponding weights (Table 1). A person is defined as empowered if she has adequate achievements in four out of the five domains, or 80 percent of the weighted indicators. The GPI, on the other hand, reflects the gap in achievements – the inequality in empowerment scores – between the primary male and female adults in the household, usually husband and wife but not always (e.g., mother and son, father and daughter-in-law, etc.).

All of these indexes have values ranging from 0 to 1, with higher values reflecting greater empowerment. Note that the WEAI focuses specifically on women's empowerment within agriculture, and does not include measures of other areas like education or socio-economic status. Like poverty indexes that are used to track overall trends in poverty, the WEAI is useful as a headline indicator that summarizes women's empowerment in a given population (program or country). However, the most useful feature of the WEAI is its decomposability, which allows users to disaggregate the 5DE achievements by domain and by indicator to see which specific areas contribute the most to both women's and men's disempowerment. Details about the WEAI's methodology, piloting, and validation are available in Alkire *et al*., (2013). [1]

Table 1: The domains, indicators, and weights in the Women's Empowerment in Agriculture Index

Domain	Indicator	Definition of Indicator	Weight
Production	Input in productive decisions	Sole or joint decisionmaking over food and cash-crop farming, livestock, and fisheries	1/10
	Autonomy in production	Autonomy in agricultural production: the extent to which the respondent's action reflects his or her own values rather than a desire to please others or avoid harm	1/10
Resources	Ownership of assets	Sole or joint ownership of major household assets	1/15
	Purchase, sale, or transfer of assets	Whether respondent participates in decisions to buy, sell, or transfer assets	1/15
	Access to and decisions about credit	Access to and participation in decisionmaking concerning credit	1/15
Income	Control over use of income	Sole or joint control over income and expenditures	1/5
Leadership	Group member	Whether respondent is an active member in at least one economic or social group	1/10
	Speaking in public	Whether the respondent is comfortable speaking in public concerning issues relevant to the respondent or their community	1/10
Time	Workload	Allocation of time to productive and domestic tasks	1/10
	Leisure	Satisfaction with time for leisure activities	1/10

Source: Alkire *et al*., (2013).

Implementation problems with the original WEAI

The implementation of the WEAI survey module, as a brand-new tool in a brand-new program, raised many concerns. For example, would the respondents interpret concepts relating to decision making, autonomy, and empowerment in similar enough ways to allow for cross-national comparisons? Would they report truthfully about the different types of production activities they participated in, what types of decisions were made, and who made them? Could they accurately recall all the different activities they engaged in the previous day, and report on both the sequence and duration of those activities? Although the version of the WEAI that was implemented in the population-based surveys included only those survey questions that worked well in the pilot districts in Bangladesh, Guatemala, and Uganda, applying these questions in new contexts was expected to result in other unforeseen implementation issues. Indeed, many of these concerns were borne out in the FTF baseline surveys in 2012 and 2013.

A learning event held in November 2013 among users of the WEAI revealed that autonomy in production, speaking in public, and time use were among the most problematic modules:

Autonomy in production. In the original WEAI, the autonomy indicator is collected using a set of questions on motivations for decisions undertaken regarding four activities: (a) getting inputs for consumption and sale at the market, (b) the type of crops to grow, (c) taking crops to the market (or not), and (d) livestock raising:

- *My actions in [ASPECT] are partly because I will get in trouble with someone if I act differently. Would you say that this is always true, somewhat true, not very true or never true?*
- *Regarding [ASPECT] I do what I do so others don't think poorly of me. Would you say that this is always true, somewhat true, not very true or never true?*
- *Regarding [ASPECT] I do what I do because I personally think it is the right thing to do. Would you say that this is always true, somewhat true, not very true or never true?*

Implementing partners reported that both respondents and enumerators found these questions regarding motivations for decisions difficult to understand because people have a hard time disentangling why they do what they do. Even when a respondent gives an answer, it is not easy for the enumerator to figure out in which response code it maps to. For example, in some languages it is nearly impossible to distinguish between "somewhat true" and "not very true". Another concern is that people may not actually feel like they are making decisions in these aspects, especially if they feel they are following cultural norms, or when other factors constrain the decision (e.g., only certain types of plants or animals grow in this climate). To make these questions more concrete, we converted each motivation into a short story or vignette where a hypothetical person acts according to the specific motivation. Instead of indicating the extent to which a particular motive is true, respondents are asked to indicate whether they are like the person described in the story, and whether they are exactly alike or just somewhat alike.

***Speaking up in public*.** In the original WEAI, respondents are asked:

- *Do you feel comfortable speaking up in public to help decide on infrastructure (like small wells, roads, water supplies) to be built in your community? [Responses: No, not at all; Yes, but with difficulty; Yes, comfortably; Not applicable]*
- *Do you feel comfortable speaking up in public to ensure proper payment of wages for public works or other similar programs? [Responses: No, not at all; Yes, but with difficulty; Yes, comfortably; Not applicable]*
- *Do you feel comfortable speaking up in public to protest the misbehavior of authorities or elected officials? [Responses: No, not at all; Yes, but with difficulty; Yes, comfortably; Not applicable]*

Implementing partners found that respondents may not be willing to identify as someone who would speak up in public because of fear of political risk. In many countries, asking about payment of wages and the misbehavior of authorities or elected officials were considered inappropriate. These questions ignore the wide variety of interpretations of "speaking in public" – is it considered public enough to speak with a small group, or does it have to be in an official community meeting? We are also not able to identify the specific challenges women face in voicing their opinions. For example, women may be comfortable speaking up about some topics but not others. To avoid the politically-sensitive subjects, we replaced the three questions with

just one question phrased more generally: *"Do you feel comfortable speaking up in public about any issue that is important to you, your family, or your community?"*

Time use. The time-use sub-module is collected by asking the respondent to narrate all the activities they did in the last 24 hours. The enumerator records each episode into the appropriate activity classification (out of the 26 categories) and rounds off the duration of the activity to the nearest 15-minute interval. When simultaneous activities occur, up to two activities can be recorded for a given time interval, the primary and secondary activities. The primary activity is the activity which the respondent considers as their main focus during that particular time interval. Although the short recall period of 24 hours minimized difficulties with recall, the module was still extremely time-consuming to administer and resulted in poor-quality data due to respondent fatigue. It was also difficult to administer the module using tablets, which required switching between paper and tablet surveys. We tested two versions of the time module in the second pilot, one was the same 24-hour recall diary with primary activities only, and the second was a stylized 7-day recall time module that collects only the paid and unpaid work categories that are used in constructing the workload indicator.

Using cognitive interviews to test revised questions

To address these issues, IFPRI, OPHI, and USAID developed a second version of the questionnaire that included either new or revised survey questions for the problematic sections and was more streamlined and easier to administer. One specific suggestion from implementers of the WEAI was to use cognitive interviewing to ensure that the revised WEAI succeeded in resolving the aforementioned problems in the original WEAI. The FTF Haiti team members used this approach as part of their baseline pretests, although they did not revise the WEAI survey based on their findings because they wished to maintain comparability with the rest of the FTF portfolio (Johnson and Diego-Rosell 2015).

Following the recommendation by Johnson and Diego-Rosell (2015), as part of the pretesting phase we administered draft survey questions from the revised WEAI while collecting additional information about the survey responses (Beatty and Willis 2007). This information was then used to evaluate the quality of the response or to help determine whether the question was generating the information we expected. To the extent possible, we used these findings to further refine the survey questions that were administered at the pilot sites.

We chose to conduct our pilots in the FTF zones in Bangladesh and Uganda because these are countries where the original WEAI pilots were administered and where IFPRI has had long-standing relationships with local partners who have prior experience collecting the WEAI data. These countries are also geographically and culturally different from one another, which is useful when testing a tool that will be applied in many different contexts. Although the entire revised instrument was tested, we prioritized the sections on time use, autonomy in decision making, and speaking up in public, as well as specific questions related to asset ownership and production decisions that were highlighted as problematic by FTF implementing partners. More details on the development and piloting of the A-WEAI is available in Malapit *et al*., (2017).

Cognitive interviewing: A brief overview

Cognitive interviewing (or cognitive testing) is an evidence-based, qualitative method for assessing whether survey questions are being understood by respondents in the way they were originally intended (Willis 2005). It is particularly useful when designing new survey questions, especially when these questions will be used cross-culturally or cross-linguistically (Willis & Miller 2011). Cognitive interviewing requires administering draft survey questions (in our case from the revised WEAI) while collecting additional information about the survey responses, which is then used to evaluate the quality of the responses or to help determine whether the questions are generating the expected information (Beatty and Willis 2007). In essence, it helps test the validity of survey questions by identifying the potential sources of response error and allowing researchers to devise solutions that will improve the quality and accuracy of survey instruments (Willis 2005).

The interview itself can comprise one or more of the following main techniques: asking the respondent to "think aloud" while formulating a response to a survey question, placing probing questions within the survey instrument (which may be asked immediately following the relevant question, retrospectively following the interview, or a combination of the two), and having the interviewer observe a respondent's verbal and nonverbal cues (Willis 2005). To minimize the burden on the respondents, which tend to be highest in the "think aloud" technique, the WEAI cognitive interviews consisted primarily of scripted probes administered by survey enumerators with basic training in cognitive interviewing techniques, supplemented with interviewer observations of verbal and nonverbal cues.

Table 2 Stages of cognition

Cognitive stages	Cognitive stage definition	Problems	Causes
1. Comprehension	Respondent interprets the question	Respondent does not understand	Unknown terms, ambiguous concepts, long and overly complex questions
2. Retrieval	Respondent searches memory for relevant information	Respondent does not remember / does not know	Recall difficulty, questions assume respondent has information
3. Judgment	Respondent evaluates and/or estimates response	Respondent does not want to tell / cannot tell	Biased or sensitive questions, estimation difficulty
4. Response	Respondent provides information in the format requested	Respondent cannot respond in the format requested	Incomplete response options, multiple responses necessary

Source: Willis (2005).

Four stages of cognition are required for an accurate response to a survey question: comprehension, retrieval, judgment, and response (Tourangeau 1984; Willis 2005). Problems can

occur during any of these stages (Table 2), leading to response error. Cognitive interviews can thus reveal at which stage in the process issues are likely to arise, and multiple iterations of testing can be used to create improved versions of survey questions. Since we already had specific feedback on which sections of the WEAI had performed poorly in the baselines and why, we used this information as a starting point in developing our cognitive instrument. Our intent was to use the findings from our cognitive interviews to help improve the validity of question responses and also decrease the time required to administer the survey, given that enumerators would have clearer and more precise questions to ask respondents.

Methodology

Sampling

The second WEAI pilot surveys were administered by Data Analysis and Technical Assistance in Bangladesh and by Associates Research Trust in Uganda, in the same FTF zone districts surveyed in the original WEAI pilots. The pilot sample was split between the two versions of the WEAI, the original (WEAI 1.1) and the revised version (WEAI 2.0). [3] Because of budget constraints, we prioritized cognitive interviews for the revised version of the WEAI, which meant that the original questions were no longer cognitively tested, with the exception of those regarding autonomy. The autonomy sub-module had the most substantial revisions, which were guided by the feedback we had received on the FTF baselines. Thus, we decided to test both the original and the new autonomy questions to give us a sense of whether the revised questions indeed helped resolve the cognitive issues raised by users.

The cognitive interviews were conducted as part of the pretesting phase prior to fieldwork. While there is no consensus regarding optimal sample size for cognitive interviews, common practice ranges from 5 to 15 interviews per language group (Beatty and Willis 2007). The goal is not to obtain sample sizes large enough to supply precision in statistical estimates, but rather to maximize variance among the respondent group by interviewing a variety of individuals who will be useful in informing decisions about if and how to modify questions (Willis 2005). Thus, we aimed for a minimum of 15 interviews per language group—four languages in Uganda and one in Bangladesh—and tried to ensure that the respondents were representative of our sample by interviewing men in dual-adult households, women in dual-adult households, and women in female-only households. [4] We also ensured that a variety of ages were represented and interviewed both younger and older men and women.

Table 3 shows the sample breakdown by household type. Because the Uganda FTF zone has four major language groups, we conducted interviews with a total of 60 respondents per round (120 total). Testing was conducted in the Jinja, Kiryandongo, and Mukono districts. Only one language is used in Bangladesh, so fewer interviews were required there. However, instead of interviewing only one respondent per household type, the Bangladesh team conducted cognitive interviews for both the primary male and primary female decision makers in the selected households, as is the standard protocol for the WEAI.

Table 3 Sample by household type

	Bangladesh	Uganda		
	Round 1	Round 1	Round 2	Total
Revised WEAI module				
Men, dual-adult households	22	20	12	32
Women, dual-adult households	23	16	12	28
Women, female-only households	7	16	8	24
Original autonomy questions				
Men, dual-adult households	8	0	8	8
Women, dual-adult households	8	6	12	20
Women, female-only households	2	2	8	8
Total	70	60	60	120

Source: Authors' calculations.

The Bangladesh field team viewed the cognitive testing exercise as a way for enumerators to learn more about the WEAI and practice conducting the WEAI interviews. This is why they chose to adhere to the WEAI protocol of interviewing two respondents per household, and instead of selecting a few enumerators to conduct the cognitive interviews, all the enumerators were asked to conduct the cognitive interviews. This resulted in a much larger sample size for Bangladesh than the minimum requirement of 15 interviews per language group, totaling 70 respondents. Testing was conducted in the Manikgonj District, roughly a two-hour drive from Dhaka. Table 4 compares respondent characteristics by country and cognitive testing round.

Table 4 Respondent characteristics

Variable	Bangladesh (%)	Uganda (%)	
	R1 (n=70)	R1 (n=60)	R2 (n=60)
Female	55.7	65	64.4
Age (years)	47.2	40	46.2
Age less than 35 years	17.1	45	41.7
Married	92.9	48.3	54.2
Highest education is less than primary, or no schooling	74.3	21.7	10.2
Female-only household	13.7	26.7	25.4
Religion of primary decision maker	100.0 Muslim	86.5 Christian	90.3 Christian

Source: Authors' calculations.

We originally attempted two rounds of cognitive interviews in both countries, in order to observe whether any problems remained after adjustments were made to the survey questions following the first round. Unfortunately, due to the short spacing between the two rounds in Bangladesh (about one week, to ensure that both pretests were completed before the upcoming Ramadan), there was not enough time to make adjustments to the instrument. Although the Bangladesh field team implemented two rounds of cognitive interviews, they tested the same questionnaire in each round. Thus, for the purpose of this paper we treat both Bangladesh rounds as one and compare their pooled results to Uganda's first-round results.

In Uganda, the rounds were spaced about one month apart, allowing for a thorough analysis of the first round of data, which provided insights into how the questionnaire or enumerator manual could be improved for better understanding by respondents and interviewers in the second round. While field teams in both countries found the cognitive testing exercise useful, we recommend following the protocol Uganda used, allowing for a sufficient period of analysis and revision between the two rounds.

Instrument design

Following the approach used by Johnson and Diego-Rosell (2015) in their cognitive testing of the WEAI in Haiti, for each question in the revised WEAI module and original autonomy sub-module we identified the stages in the cognitive process that were likely to break down and used scripted probing questions for each. Scripted probes are recommended for minimizing both enumerator error and respondent burden (Willis 2005), because this type of probing does not require expert WEAI knowledge and can be carried out by regular professional interviewers (Johnson & Diego-Rosell 2015). We relied primarily on five key probing questions, which enabled us to identify all four stages of cognitive breakdown (the cognitive interview questionnaire is available at the WEAI Resource Center [1]).

- **Comprehension:**
  - Recall period: What recall period did you include in your answer?
  - Abstract terms/concepts: Can you describe for me in your own words what the term [TERM] means?
- **Retrieval**: Many people find it difficult to recall [X]; how well do you remember [X]?
- **Judgment**: Do you think other people would find this question difficult? If so, why? Do you think other people would be reluctant or afraid to answer this question? If so, why?
- **Response**: Did you find this question easy or difficult? If difficult, why?

As in Johnson and Diego-Rosell (2015), the probing questions were asked after each sub-module of the WEAI questionnaire. This information was supplemented by enumerator observations of the verbal and nonverbal behavioral responses to each question, as well as any difficulties the respondent had with specific questions.

Analytical approach

We used cognitive interviewing to answer the following research questions:

1. **Revised WEAI instrument:** Which survey questions did the respondents find difficult to answer and why? Are there any observed differences in the cognitive interview results between different household types?
   - Regarding questions that referred to a recall period, do respondents understand the time frame in the same way that it was originally intended? Can respondents recall what occurred during the time frame referenced?
   - Regarding questions that use abstract or unfamiliar terminology, can respondents understand these terms in the way they were originally intended? Can respondents paraphrase these terms in their own words?
2. **Autonomy sub-module:** Are there observed differences in the difficulty reported by respondents answering the original autonomy questions compared with the revised autonomy questions using vignettes?
3. **Round 2, revised WEAI instrument:** Have the cognitive problems identified in the first round improved in the second round of cognitive testing in Uganda?

Because our goal was to identify all potential problems with the questionnaire, we explored every survey question where any respondent cited difficulty in answering, incorrectly paraphrased terms, or cited an incorrect recall period. We analyzed the cognitive interview responses as well as direct feedback from the field teams to assess the main causes of the issues and develop solutions through modifying the wording of questions or providing more specific enumerator instructions such as offering culturally appropriate examples or defining terms explicitly. A comparison between the original and revised autonomy sub-modules, supplemented with feedback from the field teams was used to assess which version of the autonomy questions was less likely to pose a cognitive risk. To the extent that significant differences exist between responses from different household types, we highlight those differences in the results. Findings from both the Bangladesh and Uganda cognitive interviews informed the revisions to the WEAI module, which was then subjected to a second round of cognitive testing in Uganda. Comparing the two rounds of cognitive interviews in Uganda allowed us to look at the extent to which the adjustments in the questionnaire resulted in improved responses to the cognitive probes.

Results

In this section, we focus on the WEAI survey questions that presented the greatest difficulties for respondents, organized by domain. We discuss our findings on each topic in turn and highlight cases in which we observe significant differences between household types.

Production domain: Recall period for production activities

In the Uganda cognitive interviews, the most glaring recall comprehension issue came from the question that asks respondents about their participation in production activities in the past year (Table 5, Panel A).

Table 5 Cognitive interview results: Production recall

A. Draft form of survey question

R1 WEAI 2.0 [G2.01]: "Did you yourself participate in [ACTIVITY] in the past 12 months (that is, during the last [one/two] cropping seasons)?

B. Probes

- **(Response)** Earlier I asked you if you had participated in several activities in the past 12 months, that is, in the last (one/two) cropping seasons. How was this question for you? Was this question easy or difficult? [Q4]
- **(Comprehension)** What period of time did you include in your response? [Q6]
- **(Retrieval)** Many people find it difficult to recall activities done a long time ago. How well do you remember which agricultural activities you have participated in in the last 12 months? [Q7]

C. Results	**Bangladesh**	**Uganda**	
	R1 (n=52)	R1 (n=47)	R2 (n=32)
Reported difficulty with question [Q4]	3.8%	6.4%	3.2%
Incorrect or unknown period of time included in response [Q6]	NR	59.6%	30.8%
Reported difficulty remembering events of past year [Q7]	11.9%	12.8%	0.0%

D. Suggested revision

R2 WEAI 2.0 [G2.01]: "Did you yourself participate in [ACTIVITY] in the past 12 months (that is, during the last [one/two] cropping seasons), *from [PRESENT MONTH] last year to [PRESENT MONTH] this year?"*

Source: Authors' calculations.

The question was originally formulated to be as general as possible, so that users who are implementing the survey in different countries can adapt the question to the agricultural cycle in their respective survey area to cover a full year. In the baselines, however, this was a source of confusion (Johnson and Diego-Rosell 2015). In the cognitive interviews, we used response, comprehension and retrieval probes (Table 5, Panel B) to assess whether the recall time frame specified in the question matched the same time frame that respondents gave in their answers. Although only 3.8 percent of respondents in Bangladesh and 6.4 percent of respondents in Uganda reported having difficulty with this question, a much higher proportion of respondents reported having difficulty remembering events of the past year: 11.9 percent in Bangladesh and 12.8 percent in Uganda (Table 5, Panel C). When asked how much time respondents included in their response, answers ranged anywhere from 3 to 12 months. In Uganda, 59.6 percent of first-round respondents either could not say what recall period they used or referred to a time frame other than 12 months.

To address this problem, we modified the question to, "Did you yourself participate in [ACTIVITY] in the past 12 months (that is, during the last [one/two] cropping seasons), **from [PRESENT MONTH] last year to [PRESENT MONTH] this year**?" In the second round of cognitive interviews in Uganda, this change reduced the recall error by half, with now only 30.8

percent of respondents either stating periods of less than one year or unable to respond. This problem appears to have been mitigated with a more clearly defined time frame.

Production domain: Testing alternative autonomy sub-modules

The original WEAI questions on autonomy ask about respondents' motivations for their actions. The aim of the questions is to assess the extent to which an individual's actions are motivated by his or her own values, by coercion, or by fear of others' disapproval (Alkire *et al*., 2013) (Table 6, Panel A).

For example, "Regarding [ASPECT], I do what I do so others don't think poorly of me." Many WEAI users struggled to implement this section because the phrases were thought to be abstract and difficult to understand, which may also have contributed to inappropriate translations to local languages in the field. In Haiti, for example, Johnson and Diego-Rosell (2015) found that the term "think poorly" in the example above was particularly confusing for respondents because it had many different connotations in Haitian Creole, ranging from the intended meaning of "think poorly" to more forceful connotations such as "wishing bad things to other people" and "destroying someone." To address this concern, we developed an alternative autonomy sub-module that uses short stories, or vignettes, to demonstrate each type of motivation (Table 6, Panel A). Instead of asking the respondents how true different statements on motivations are for them, we ask the respondents whether they are similar to the person described in the short hypothetical story. This approach is an attempt to make the concepts more concrete and easy to understand, and at the same time avoids the use of problematic phrases such as "get into trouble with someone" and "think poorly."

Following Johnson and Diego-Rosell (2015), in both versions we used response probes that ask not only whether the respondent thought a question was difficult, but also judgement probes on whether he or she thought other people would find the question difficult (Table 6, Panel B). Inconsistencies between these answers can reveal problem areas, as a respondent may not admit to finding a question difficult to answer but would note that others might find it difficult. Consistent with what Johnson and Diego-Rosell (2015) found in Haiti, in both versions we found large discrepancies in the percentages of respondents who found the question difficult themselves versus how difficult they thought others would find it. The original autonomy questions did slightly better in Bangladesh, with 16.7 percent of respondents reporting difficulties themselves and 38.9 percent suggesting potential difficulties for others, compared with 17.3 percent and 48.8 percent, respectively, for the vignettes (Table 6, Panel C). However, the vignettes appeared to do better in Uganda, compared with the original questions. About half the Uganda first-round respondents reported difficulties with the original questions, compared with only 17 percent of respondents reporting difficulties with the vignettes (Table 6, Panel C).

Three-quarters of the Uganda respondents also thought others might find the original questions difficult, whereas 53.2 percent of respondents thought that others would find the vignettes difficult (Table 6, Panel C). In Uganda, respondents cited comprehension issues, lack of experience regarding the situation described, and different approaches to decision making as the primary sources of difficulty. In Bangladesh, women in dual-adult households seemed to find these questions slightly more difficult than men in dual-adult households or women in single-

adult households. This may be due to their lack of experience in making decisions about selling crops/products, because these tend to be activities or decisions more commonly handled by men.

Table 6 Cognitive interview results: Autonomy

A. Draft form of survey questions

Original WEAI 1.1 [Module G5(B)]: This set of questions is very important. I am going to give you some reasons why you act as you do in the aspects of household life I just mentioned. You might have several reasons for doing what you do and there is no right or wrong answer. Please tell me how true it would be to say:
... My actions in [ASPECT] are partly because I will get in trouble with someone if I act differently. Would you say that this is always true, somewhat true, not very true, or never true?
Vignettes WEAI 2.0 [Module G4]: Now I am going to read you some stories about different farmers and their situations regarding different agricultural activities. This question format is different from the rest so take your time in answering. For each I will then ask you how much you are like or not like each of these people. We would like to know if you are completely different from them, similar to them, or somewhere in between. There are no right or wrong answers to these questions.
... [PERSON'S NAME] can't grow other types of crops here for consumption and sale in market. Beans, sweet potato, and maize are the only crops that grow here.
... Are you like this person? Are you completely the same/different, or somewhat the same/different?

B. Probes

- **(Response)** Is this question easy or difficult? [WEAI 1.1, Q5/9/13; WEAI 2.0, Q50/54/58/62]
- **(Judgement)** Do you think some people may find this question difficult to answer? [WEAI 1.1, Q7/11/15; WEAI 2.0, Q52/56/60/64]

C. Results	**Bangladesh**	**Uganda**	
Original WEAI 1.1	R1 (n=18)	R1 (n=8)	R2 (n=28)
Reported difficulty with question	16.7%	50.0%	28.6%
Reported others would have difficulty with question	38.9%	75.0%	71.4%
Vignettes WEAI 2.0	R1 (n=52)	R1 (n=47)	R2 (n=31)
Reported difficulty with question	17.3%	17.0%	21.9%
Reported others would have difficulty with question	48.8%	53.2%	65.5%

D. Suggested revisions

Round 2, WEAI 2.0: No change
A-WEAI: Dropped because it took the longest time to administer among the sub-modules (23% of interview time)

Source: Authors' calculations.
Note: Panel A shows only excerpts of the survey questions for illustrative purposes. The complete survey instruments and datasets are available from the WEAI Resource Center: www.ifpri.org/topic/weai-resource-center.

Although the respondents reported more difficulties with the original questions, the Uganda field team still found the original questions much easier to implement than the vignettes, which were more complicated to translate to local languages.

Overall, these mixed results made it difficult to draw strong conclusions. Respondents found the vignettes to be more difficult than the original questions in Bangladesh, and in Uganda they were harder to implement. However, the Uganda respondents appeared to find the vignettes easier to understand, given the lower incidence of reported difficulties. The Bangladesh field team also used a storytelling enumeration style in general, so the original questions were already implemented in a vignette-like fashion. Since the means of implementation of the two versions were more similar in Bangladesh, the main difference with the vignettes was the increased length, which might explain the higher difficulty reported in the cognitive interviews.

Taking all this information into account, including feedback we received from the field teams in the original FTF baselines, we opted to use the vignettes in the pilot surveys to allow for further analysis. Therefore, there were no changes in the survey questions between the two rounds, and both versions of the autonomy sub-module were tested in the second round of cognitive interviews in Uganda.

Surprisingly, the reported difficulty with the vignettes increased during the second round, while the reported difficulty with the original questions improved (Table 6, Panel C). However, this may be due to differences in the sample. Focusing only on the second-round results, we observe that the vignettes have a smaller proportion of respondents reporting difficulties, which suggests that comprehension may be better for the vignettes compared with the original questions. This supports the view that a storytelling approach is a promising mechanism for asking these types of abstract questions, although storytelling takes more time to implement than conventional survey questions.

Nevertheless, the proportion of respondents citing difficulty in this section is still very large, leaving much room for further refinement. [5] Even with the improvements achieved by using vignettes, the autonomy section is still one of the most challenging parts of the WEAI to implement. Following the cognitive testing exercise, we used the vignettes in the WEAI 2.0 pilot surveys, which worked well in the field and captured more nuanced responses compared with the original autonomy questions. In the end, we recommended that the autonomy vignettes be dropped from the A-WEAI, in part because of the length of time it took to administer (around 23 percent of survey time).

Resource domain: Sensitivity to credit questions

The resource domain of the WEAI involves relatively few implementation issues, having undergone extensive testing by IFPRI in its data collection efforts on intrahousehold allocation in recent years. Although credit was not flagged in our cognitive interviews as problematic, we found interesting gender differences in the degree of sensitivity to responding to these questions. Questions related to personal finances and financial decisions may be viewed as personal and sensitive, particularly if this information is not readily shared among household members, creating reluctance to respond truthfully (Table 7, Panel A).

Table 7 Cognitive interview results: Credit

A. Draft form of survey questions

Access to credit [Module G3(B)]: Next I'd like to ask about your household's experience with borrowing money or other items in the past 12 months.
…Would you or anyone in your household be able to take a loan or borrow cash/in-kind from [SOURCE] if you wanted to?
... Has anyone in your household taken any loans or borrowed cash/in-kind from [SOURCE] in the past 12 months?
... Who made the decision to borrow from [SOURCE] most of the time?
... Who makes the decision about what to do with the money/item borrowed from [SOURCE] most of the time?

B. Probe

- **(Judgement)** Do you think other respondents being asked these questions would be embarrassed or reluctant to share information on household borrowing activities? [Q46]

C. Results	**Bangladesh**		**Uganda**	
	Men (n=21)	Women (n=29)	Men (n=27)	Women (n=51)
Think other respondents being asked these questions would be embarrassed or reluctant to share information on household borrowing activities [Q46]	38.1%	10.3%	66.7%	60.8%

Source: Authors' calculations.

Thus, for the cognitive interview in the access-to-credit sub-module, we asked respondents the judgement probe, "Do you think other respondents being asked these questions would be embarrassed or reluctant to share information on household borrowing activities?" We find that, overall, respondents in Uganda are more likely to find these questions sensitive, compared to respondents in Bangladesh. In addition, we find gender differences: in Uganda, 66.7 percent of men and 60.8 percent of women report that other respondents will be reluctant to answer, while in Bangladesh, 38 percent of men and only 10.3 percent of women report sensitivity to this issue.

Some of the reasons cited by Ugandan men include "fear to be identified with debts," "not want[ing] to share their information with others," and embarrassment "for those who borrowed loans and used it in a wrong way." Ugandan women also tended to cite general discomfort with speaking about money issues as well as embarrassment "because they failed to pay the loan" or "fear of their spouse," and some noted that "spouses don't always share with their household members after borrowing money." Although we did not make any changes to the questions in this sub-module as a result of the cognitive interviews, these findings confirm the importance of conducting the WEAI interviews separately and privately for the male and female respondents.

Leadership domain: Use of ambiguous terms

Based on feedback from the WEAI baselines, we knew that the speaking-in-public sub-module was politically sensitive because of the examples that were originally given, such as building community infrastructure, ensuring proper payment of wages for public works programs, and

protesting misbehavior of authorities or elected officials. In an attempt to move away from these politically sensitive issues, we tested a more general version of this question: "Do you feel comfortable speaking up in public about *any issue* that is important to you, your family, or your community?"

Table 8 Cognitive interview results: Speaking in public

A. Draft form of survey question

R1 WEAI 2.0 [G6.01]: "Do you feel comfortable speaking up in pubic about *any issue* that is important to you, your family, or your community?"

B. Probes

- **(Response)** Earlier I asked you if you felt comfortable speaking in public about issues that are important to you, your family, or your community. How was this question? Was this question easy or difficult? [Q82]
- **(Comprehension)** Can you tell me in your own words what it means when I say "issues that are important to you, your family, or your community"? [Q84]

C. Results	**Bangladesh**	**Uganda**	
	R1 (n=52)	R1 (n=47)	R2 (n=32)
Reported difficulty with question [Q82]	9.8%	8.5%	3.1%
Gave correct definition and/or examples of issues [Q84]	NR	27.7%	62.5%

D. Suggested revision

R2 WEAI 2.0 [G6.01]: "Do you feel comfortable speaking up in public about *anything* that is important to you, your family, or your community?"

A-WEAI: Dropped because concerns about the social or cultural acceptability of asking these questions in specific contexts remained, even with the new wording.

In the cognitive interviews, we asked the response probe on whether respondents found this question to be difficult, and, to assess whether the phrasing of "any issue" comes across as less political, we also asked respondents a comprehension probe to describe in their own words what they thought was meant by "issues that are important to you, your family, or your community" (Table 8, Panel B).

In the first round of testing, we found that less than 10 percent of respondents found this question difficult to answer (Table 8, Panel C). However, we learned that in Uganda, the word "issue" translates to "problem" or "challenge" in local languages and thus has a more negative rather than neutral connotation. Not surprisingly, only 27.7 percent of respondents cited a definition of "issue" that included both positive and negative topics. Therefore, in order for people to include issues with both positive and negative connotations, *any issue* was changed to read *anything* (Table 8, Panel D). With this small change in wording, 62.5 percent of respondents gave a neutral definition of "issue" in the subsequent round of testing (Table 8, Panel C). Despite the improvement in the wording of this question, we ultimately recommended dropping this from A-WEAI because the concerns about the social or cultural acceptability of asking these questions in specific contexts remained.

Time domain: Testing alternative recall periods

In the original WEAI, time use was collected using a 24-hour recall time diary, collecting up to two activities per time interval. This sub-module was often cited as problematic because it took substantial time to administer and may not be indicative of a usual day. We tested two alternatives in response to this concern. First, we used the same 24-hour recall sub-module but recorded only the primary activity for each time interval. In our analysis of the original pilot data, dropping secondary activities did not significantly change the WEAI results. Second, we tested an alternative seven-day recall sub-module. Instead of a complete time diary, this is a stylized time-use module that collects information only on the activities that were relevant for calculating the WEAI workload indicator, which includes all types of paid and unpaid work, including domestic work and childcare. The purpose of testing the time-use section is to determine a best practice for quickly capturing an accurate depiction of a respondent's time allocation. Table 9 shows the original form of the survey questions (Panel A) and the selected probes that we found most useful (Panel B).

During the first round, 23.5 percent of respondents in Bangladesh and 26.1 percent of respondents in Uganda reported that they had difficulty remembering their activities during the *past week*, as compared to 6 percent and 14.9 percent of the same respondents in Bangladesh and Uganda, respectively, when asked the retrieval probe about how well they remembered their activities over the *past 24 hours* (Table 9, Panel C). As one might expect, these results suggest that respondents find it much easier to recall events that occurred in the past 24 hours than events that occurred in the past week. Although this can be taken as evidence against using the seven-day recall approach, we still opted to use this sub-module so as to be able to compare the differences in reported work time between the two types of recall periods. Consistent with the cognitive testing findings, we did find that the range of working time reported for the seven-day recall sub-module was much higher than with the 24-hour recall, which suggests higher measurement error. Ultimately, for the A-WEAI, we opted to keep the 24-hour recall sub-module rather than switching to seven-day recall in order to minimize the potential retrieval error.

We also found that over half of respondents in both Bangladesh and Uganda said that their activities change from day to day, which could influence the respondent's ability to remember (retrieval) or estimate the correct response (judgement). In Uganda, 16.7 percent of these respondents cite seasonality specifically as a reason for the change in their daily activity. Although we recognize that seasonality is an important aspect of time allocation in agriculture, this is something that ought to be addressed in the timing and frequency of surveys rather than in the rewording of survey questions.

Despite the findings from the first round, we still opted to use both versions of the time module in the pilot surveys, and therefore there were no changes in the survey questions between the two rounds. Nevertheless, during the second round in Uganda, we observed substantial improvement in the cognitive interviews (Table 9, Panel C). Only 6.3 percent reported difficulty remembering how they had spent their time over the past week, and none reported difficulty with the 24-hour recall. Thus, differences in reported difficulty may be attributable either to better enumerator administration of the questionnaire, as field teams gain more experience in conducting the survey, and/or to a different respondent pool.

Table 9 Cognitive interview results: Time use

A. Draft form of survey questions

7-day recall [G5.01]: Now I'd like to ask you some questions regarding how you've spent your time over the last week and whether or not this was typical.
… In the last 7 days, how much time in hours did you spend on [ACTIVITY]?
… Did you spend a usual amount of time on [ACTIVITY] in the last 7 days?
... Since the last week was not usual, within the last 6 months how much time do you usually spend on [ACTIVITY] per week?

24-hour recall [G5.02]: Now I'd like to ask you about how you spent your time during the past 24 hours. We'll begin from yesterday morning and continue through to this morning. This will be a detailed accounting. I'm interested in everything you do (i.e., resting, eating, personal care, work inside and outside the home, caring for children, cooking, shopping, socializing, etc.), even if it doesn't take you much time.

B. Probes

- **(Retrieval)** Many people find it difficult to recall how many hours they spent on certain activities in the last week. How well do you remember which specific activities you spent time on during the past week? [Q69]
- **(Retrieval)** Many people find it difficult to recall every activity done in a day. How well do you remember which specific activities you were doing at every time in the last 24 hours? [Q70]
- **(Retrieval / Judgement)** Is your daily schedule always the same, or do your activities change from day to day? [Q72]

C. Results	**Bangladesh**	**Uganda**	
	R1 (n=51)	R1 (n=47)	R2 (n=32)
Reported difficulty remembering events/activities done during the past week [Q69]	23.5%	26.1%	6.3%
Reported difficulty remembering events/activities done during the past 24 hours [Q70]	6.0%	14.9%	0%
Reported schedule varies from day to day [Q72]	56.9%	63.8%	56.3%

D. Suggested revisions

Round 2, WEAI 2.0: No change

A-WEAI: To minimize measurement error due to retrieval difficulties, we use the 24-hour recall sub-module, collecting data on only primary (not secondary) activities to reduce complexity.

Source: Authors' calculations.

Conclusion

In this paper, we described how we used cognitive interviewing to guide the design of the second pilot surveys for the WEAI, which ultimately led to the final recommendations for the Abbreviated WEAI. Overall, we are convinced that cognitive interviewing added value to our fieldwork and directly contributed to reducing the potential for response error in our survey instruments. We recommend that all researchers implementing surveys adopt this method as part

of their standard pretesting protocol, particularly for studies of highly complex and sensitive topics such as women's empowerment and gender issues.

Although the cognitive testing exercise was crucial to the redesign of the WEAI, we also learned that cognitive testing should not be the only consideration in making questionnaire design decisions (Willis 2005). For instance, the cognitive testing results provided inconsistent information that made it difficult to draw any strong conclusions about which version of the module on autonomy in decision making to use. Speaking directly with the field supervisors was most beneficial, as it allowed us to learn where they themselves had had difficulty with the questions and thus may have influenced the results we obtained. Thus, qualitative insights are especially useful in informing the design of the survey instrument and should be done in an iterative process in conjunction with cognitive testing.

Lessons learned and best practices

While we referred to published literature and consulted experts prior to conducting the cognitive interviews, our own experience taught us a number of lessons and confirmed some, but not all, of the best practice suggestions from the literature.

1. **Be explicit about the cognitive testing process.** Given that many local data firms commonly conduct what is referred to as pretesting prior to fieldwork, it is important to be clear and specific about the process for cognitive pretesting and how this differs from general pretesting.

2. **Be aware of multicountry challenges.** We encountered little guidance regarding the challenges of conducting cognitive testing in multiple countries. Our survey instrument was originally designed to assess women's empowerment in a methodologically consistent manner in 19 countries (and is now used more broadly), and thus we wanted to ensure that changes to questions in one country did not jeopardize respondent comprehension in another country. Although we wanted to conduct pretesting in at least one site in all the regions where FTF countries are located, budgetary considerations permitted covering only one site in Asia and one in East Africa. A related issue is reconciling feedback from a cognitive testing round between multiple countries. Fortunately, the Bangladesh team was always in agreement with the changes proposed by the Uganda team. However, disagreements between country teams would have made it difficult to modify the questionnaire. A final, though almost overlooked practical challenge, is coordinating the timing of pretesting between countries. We staggered our work in the two countries in the following manner to allow time for communication of results between the country teams: Bangladesh R1, Uganda R1, Bangladesh R2, Uganda R2. At least one week of time between rounds is necessary when collaborating across countries, especially in different time zones. Note that we report both rounds of cognitive interviews in Bangladesh as Round 1 because there was no change in the survey instrument between rounds.

3. **Be strategic in the amount of information collected.** Our cognitive pretesting ended up using a combination of qualitative and quantitative methods, and we collected a great deal of information that needed to be analyzed in a short period. We even voice recorded our interviews as recommended by experts (Willis 2005; Beatty and Willis 2007). In retrospect, recording the interviews was not necessary and placed an additional burden on the enumerators; we had neither the time nor the budget to pay for translation or analyze the recordings. Additionally, while recording enumerator observations was also a suggested best practice, our enumerators were generally better trained in collecting quantitative information and it was difficult for them to discern nonverbal observations that indicate confusion; the comments they wrote in this regard did not prove useful.

4. **Prioritize problematic sections.** One of the major reasons our team undertook cognitive testing was to scrutinize questionnaire modules that were identified as problematic during the baseline. We also wanted to test new and/or modified questions and assess how well respondents understood these as compared to the original questions. However, researchers with experience conducting cognitive testing suggested that we focus our questions on the sections deemed most problematic, while asking fewer questions in sections that had not experienced known problems to date. This worked well, as it allowed us to more fully understand where comprehension was breaking down in problematic questions and sections while still being able to flag problems in sections where we were unaware of cognition issues.

5. **Understand where to save and where to splurge.** Perhaps one of the most costly of the best practice suggestions is to use two enumerators per cognitive pretesting interview. One enumerator conducts the interview, while the other records nonverbal observations and manages the voice recorder. However, our budget allowed only one enumerator per interview, and given that the voice recordings and nonverbal observations did not prove useful, we recommend using only one enumerator per interview. That said, researchers should not skimp on training the enumerators. While all the enumerators we worked with had prior quantitative (and some qualitative) interviewing experience, none had formal cognitive interviewing training or experience. Thus, it was important to take the time to ensure that they understood the process and goals of the cognitive interviews. In Uganda this meant spending time in the field to train the enumerators on both the questionnaire and cognitive testing as well as administering the first round of cognitive testing; by the second round, the team had built capacity and was able to do the fieldwork independently. In hindsight, we can see that our enumerators needed more training in the methods of cognitive testing. While they did well in following the scripted interview questions, they struggled to observe more nuanced signs of respondent difficulty such as long silences, contradictions, hesitation or reluctance in answering a question, and other nonverbal cues.

Endnotes

[1] For more information on the WEAI, please see www.ifpri.org/topic/weai-resource-center.

[2] Throughout this paper we refer to several versions of the WEAI. The "WEAI" refers to the original version released in 2012. The "WEAI 1.1" refers to a slightly modified version that was created following feedback from the baseline surveys and remains comparable to the WEAI. In the WEAI 1.1, response codes were streamlined and enumeration instruction was revised for faster and more efficient administration. The "revised WEAI" refers to the version that was piloted in 2014 and that eventually resulted in the final Abbreviated WEAI (A-WEAI) version (Malapit *et al*., 2017). For more details on each of these versions, please see the versions table in the WEAI Resource Center: www.ifpri.org/sites/default/files/Basic%20Page/weai_versions_table.pdf.

[3] The Bangladesh pilot dataset is available at www.ifpri.org/publication/womens-empowerment-agriculture-index-weai-pilot-ii-bangladesh. The Uganda pilot dataset is available at www.ifpri.org/publication/womens-empowerment-agriculture-weai-pilot-ii-uganda.

[4] The WEAI purposely avoids the concepts of male-headed and female-headed households, which are fraught with difficulties and assumptions about "headship" (Buvinić & Gupta 1997; Budlender 2003; Diana Deere *et al*., 2012). Rather, households are classified in terms of whether there are both male and female adults (dual-adult households), only female adults (female-only households), or only male adults (male-only households) (Alkire *et al*., 2012; Alkire *et al.* 2013). The latter are very rarely found in the WEAI study areas, and are excluded from our sample because of our focus on women's empowerment.

[5] The autonomy module is undergoing further testing and development as part of the Project WEAI (Pro-WEAI), which is developing a WEAI for project use under Phase 2 of the Gender, Agriculture, and Assets Project (GAAP2). See http://gaap.ifpri.info/ for more information.

Acknowledgments

This work was undertaken as part of the CGIAR Research Program on Policies, Institutions, and Markets (PIM). This research was supported by the United States Agency for International Development (USAID) under Feed the Future, the US government's global hunger and food security initiative, through its partnership with the International Food Policy Research Institute on the Women's Empowerment in Agriculture Index, USAID Grant Number EEM-G-00-04-00013-00. We thank the survey enumerators and other staff of Data Analysis and Technical Assistance Limited in Bangladesh, and Associates Research Trust in Uganda, for conducting the cognitive interviews and providing assistance with data cleaning. We would also like to thank Kiersten Johnson and Pablo Diego-Rosell for their invaluable advice regarding the cognitive testing process, Sabina Alkire for assistance in refining the autonomy vignettes, and Jessica Heckert and other seminar participants for providing helpful comments and suggestions. We are especially grateful to Ruth Meinzen-Dick, Agnes Quisumbing, Caren Grown, Emily Hogue, and Farzana Ramzan for guiding this effort and for providing invaluable feedback. The contents are

the responsibility of the authors and do not necessarily reflect the views of CGIAR, PIM, USAID or the US government.

References

Alkire, S. et al., 2013. The Women's Empowerment in Agriculture Index. *World Development*, 52, pp.71–91.

Alkire, S. et al., 2012. *The Women's Empowerment in Agriculture Index*, Washington, D.C.

Beatty, P.C. & Willis, G.B., 2007. Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, 71(2), pp.287–311.

Budlender, D., 2003. The debate about household headship. *Social Dynamics*, 29(2), pp.48–72.

Buvinić, M. & Gupta, G.R., 1997. Female-Headed Households and Female-Maintained Families: Are They Worth Targeting to Reduce Poverty in Developing Countries? *Economic Development and Cultural Change*, 45(2), pp.259–280.

Diana Deere, C., Alvarado, G.E. & Twyman, J., 2012. Gender Inequality in Asset Ownership in Latin America: Female Owners vs Household Heads. *Development and Change*, 43(2), pp.505–530.

Johnson, K.B. & Diego-Rosell, P., 2015. Assessing the cognitive validity of the Women's Empowerment in Agriculture Index instrument in the Haiti Multi-Sectoral Baseline Survey. *Survey Practice*, 8(3).

Malapit, Hazel J.; Pinkstaff, Crossley; Sproule, Kathryn; Kovarik, Chiara; Quisumbing, Agnes R.; and Meinzen-Dick, Ruth Suseela. 2017. The Abbreviated Women's Empowerment in Agriculture Index (A-WEAI). IFPRI Discussion Paper 1647. Washington, D.C. http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/131231

Tourangeau, R., 1984. Cognitive sciences and survey methods. In T. B. Jabine, M. L. Straf, & R. Tourangeau, eds. *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Washington, DC: National Academies Press, pp. 73–100.

Willis, G.B., 2005. *Cognitive interviewing : a tool for improving questionnaire design*, Thousand Oaks, CA: Sage Publications.

Willis, G.B. & Miller, K., 2011. Cross-Cultural Cognitive Interviewing Seeking Comparability and Enhancing Understanding. *Field Methods*, 23(4), pp.331–341.