A Practical Guide to Measuring Women's and Girls' Empowerment in Impact Evaluations

APPENDIX 2: EXAMPLES OF NON-SURVEY INSTRUMENTS AND TIPS ON USING THEM

J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

WHAT ARE SOME EXAMPLES OF NON-SURVEY INSTRUMENTS AND WHEN ARE THEY USEFUL?

Non-survey instruments can be useful for capturing difficult-to-measure outcomes and indicators related to women's empowerment. They can help measure things that people may not report honestly in a survey (e.g., discrimination, gender bias, or prejudice), are highly subject to recall error (e.g., how many times women spoke in yesterday's community meeting), or have aspects that people may not even be consciously aware of (e.g., subconscious or implicit gender bias). They can also be useful in helping us understand agency in decision-making by allowing us to observe decisions between people. However, non-survey instruments tend to be more expensive to implement than a set of analogous survey questions because they require more time to collect data and more in-depth training for enumerators or interviewers. Like survey questionnaires, we need to pretest our non-survey instruments extensively with people who are similar to the participants in our study to make sure the logistics work and participants can understand and complete them easily. In many cases, we can add non-survey instruments into our baseline or follow-up surveys and train our team of enumerators or interviewers to conduct both.

This appendix gives some examples of different kinds of non-survey instruments that can be useful for measuring women's empowerment, their pros and cons, and tips for when to consider using them in addition to or instead of a set of survey questions. Like the rest of the guide, we primarily focus on examples of how researchers have used non-survey instruments in quantitative analysis, so it may be most relevant to researchers and practitioners interested in these uses. Most of the non-survey instruments described in this appendix can generate both quantitative and qualitative data and have many uses beyond the specific examples featured here.

LOCATION: INDIA. PHOTO: SURYA BANDA | J-PAL
COVER LOCATION: COLOMBIA. PHOTO: OLIVIA SEIGEL | J-PAL/IPA

1. DIRECT OBSERVATION AND STRUCTURED COMMUNITY ACTIVITIES

Instead of asking people to recall how they behaved in a particular situation, we can try to observe the behavior of interest directly as it happens. One form of direct observation that is useful for measuring power dynamics is observing a group of people interacting, having a discussion, and/or making a decision. For example, a member of our research team could attend a monthly community meeting hosted by a local government body and count how many women attend and how many times women contribute to the discussion. Seeing how people speak to and behave around each other can give us information about decision-making that is impossible to gather from self-reported information in a survey, which may be subject to recall error and reporting bias.

If there are no natural opportunities to observe a group decision, another option is to create a structured community activity that asks people to make a group or individual decision as part of the program or survey (see example below). We could similarly induce a decision between spouses or partners at the end of a questionnaire to observe intra-household decision-making dynamics. By inducing a decision, we can isolate the exact type of decision we hope to observe, like relative bargaining power between spouses. Unless structured community activities are a normal part of the program being evaluated, they are typically more expensive than observing events that would have happened anyway.

PROS:

This technique is good for outcomes about agency in group settings that are subject to misreporting or poor recall, such as people's participation in community meetings or joint household decisions. It is one of the tools we can use to measure women's voice and participation at the group or community level in addition to the household level. With well-trained enumerators or interviewers, it can also provide rich data on power dynamics, interactions between people, language and vocabulary, body language, and more.

CONS:

This approach may not be as useful when the presence of a research team member affects people's behavior, which may be less likely in large group meetings or activities. There is also a risk of capturing snapshots of a specific point in time that are not representative. It requires a large enough number of observations to mitigate this risk. Direct observation is also subject to enumerator or interviewer bias. Including enumerator fixed effects in our regressions can help address this. Direct observation requires highly trained enumerators or interviewers who can consistently and accurately report what is occurring, so it can in some cases be expensive relative to surveys. Because the data are so rich, coding, cleaning, and analyzing the data from direct observation can also be more time-consuming than completing the same tasks with survey data.

TIPS:

- Consider using direct observation or a structured community activity if the evaluation is trying to measure women's participation in community decision-making.
- Before implementing, think about whether or not the presence of a member of our research team is likely to affect the outcomes we want to measure. A regularly scheduled large group meeting is a good opportunity to use direct observation (i.e., a community meeting with local government officials). One person's presence is unlikely to affect most people's behavior in a large group, but it may in small group settings.
- When there are not existing opportunities to observe a group discussion or decision, another option is to prompt a standardized activity as part of the survey to observe a real decision taking place.
- Repeat the observations over time and/or in different locations to avoid misrepresenting community interactions with just one example.

1. DIRECT OBSERVATION AND STRUCTURED COMMUNITY ACTIVITIES

LOCATION: SIERRA LEONE. PHOTO: GLENNA GORDON | J-PAL/IPA

EXAMPLE:

In studying a community-driven development program in Sierra Leone, Casey, Glennerster, and Miguel (2012) wanted to evaluate if the program was effective in promoting democratic decision-making and involving marginalized populations in community decisions, including women. The researchers planned to observe a decision-making process at community meetings to measure these outcomes. To ensure that the same decision was being made across treatment and comparison groups, researchers prompted a situation in which the whole community would have to make a decision together. Local leaders gathered community members for a meeting during which the enumerators announced that they would like to give the community a gift of either a carton of batteries or packages of iodized salt to thank them for taking part in the study.[1] The community had to decide which gift it preferred; field research staff observed and recorded who engaged in the discussion, including the frequency with which women participated. They found that participation in the community-driven development program did not lead to increased women's participation in these community meetings. A script for the sessions along with charts for tracking community members' participation in the meetings are available on the Harvard Dataverse.[2]

CITATION:

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *The Quarterly Journal of Economics* 127 (4): 1755-1812. https://doi.org/10.1093/qje/qje027.

[1] These options should be based on formative research to determine which options are desirable in the specific community and what kinds of decisions will generate discussion.

[2] Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2013. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." Harvard Dataverse, V4. https://hdl.handle.net/1902.1/21708.

2. PURCHASE DECISIONS

People's decisions to purchase or use a product or service can often give us more accurate information about their willingness to pay for them than their stated responses during a survey. In a survey, respondents do not have any of their own resources on the line. So when they are asked about how likely they are to use or buy certain things, such as contraceptives, they may give a response that reflects what they hope to do but not what they would actually do, or give an answer that is in line with what they think the interviewer wants to hear. For instance, people may be hesitant to admit that they do not use contraceptives even though they are sexually active.

Observing real-life decisions to purchase or use products or services can help us measure women's access to resources and agency in purchasing and/or using products or services that matter to them. To observe these decisions, interviewers could give respondents numbered vouchers to purchase a product and record whether people redeemed the voucher and how much they purchased in a given timeframe using administrative data from the partner vendors or service providers. Randomly assigning different prices or discounts allows us to measure demand for the product at various prices and can be useful for helping implementing organizations set prices. Researchers have also used similar methods to measure a woman's willingness to pay to receive a smaller household cash transfer herself, rather than have her husband receive a slightly larger household cash transfer, as a proxy measure of women's bargaining power in the relationship.[3]

PROS:

Purchase decisions are better than survey questions at capturing people's actual purchasing behavior and preferences (compared to what they might say or intend to do but not actually do). They can also help us measure the demand for products or services at various prices.

CONS:

It can be challenging to eliminate the risk of token purchases that participants make to please the interviewer. In addition, a one-time purchase decision may not reflect a person's purchasing behavior in general, especially when the voucher used to purchase items is gifted. For things that people buy regularly, measuring repeat purchases can help disentangle this.

TIPS:

- Purchase decisions are useful when there is likely to be a large amount of reporting bias in people's responses to how much they buy or use a product or service, and when take-up and use of a product or service are main outcomes of interest in our evaluation.
- When possible, measure purchases after the survey has ended so we know that purchase decisions are not affected by the presence of the interviewer. Often researchers do this by gathering administrative sales data from partners like local pharmacies or retail shops.
- If purchases must be measured during the survey itself, try to offer meaningful amounts of the item for purchase to protect against token purchases made to please the interviewer.

EXAMPLE:

Thornton (2008) examined the impact of a program in Malawi that provided free HIV tests door-to-door and offered small cash incentives to people to collect their results at voluntary care and testing centers in their community. Once participants had the opportunity to collect their test results a couple months after the campaign, enumerators visited them at home and offered the chance to buy subsidized condoms. Researchers used this purchasing decision as a proxy measure of contraceptive use to understand how learning one's HIV status affects these purchase decisions.

CITATION:

Thornton, Rebecca L. 2008. "The Demand for, and Impact of, Learning HIV Status." *American Economic Review* 98 (5): 1829-1863. https://doi.org/10.1257/aer.98.5.1829.

[3] Almås, Ingvild, Alex Armand, Orazio Attanasio, and Pedro Carneiro. "Measuring and Changing Control: Women's empowerment and targeted transfers." National Bureau of Economic Research Working Paper No. w21717, November 2015. https://doi.org/10.3386/w21717.

3. GAMES

Asking respondents to play a game can help measure qualities like altruism, cooperation, and trust. Some common examples are the dictator game (measuring altruism), the trust game, and the ultimatum game (measuring fairness or reciprocity). They can be useful for testing theories about how people respond to different incentives, as well as differentiating between different types of people along lines that are not easy to observe. For example, games can be used to classify people as more or less risk averse, or more or less cooperative. By observing and comparing how men and women make choices in games, we can approximate things like differences in bargaining power or preferences for contributing to public goods. Games also have important limitations, stemming from the fact that they do not have real world stakes.

PROS:

Compared to prompting more complex real-world scenarios, they are typically a more cost-effective way to see how people might respond to different incentives and choices.

CONS:

Games are not the real world, so the findings may not be generalizable to real life situations with real stakes. People may also play a certain way if they perceive they are being evaluated on their decisions. In addition, games can potentially be an intervention themselves, particularly when coupled with an intervention related to the skills used in the game (e.g., negotiation training). This may be a concern when the intervention being evaluated is relatively small in scope (e.g., a short information session), but may be less of a concern for large-scale interventions or programs.

TIPS:

- Games can be used to measure both outcomes (e.g., prevalence of cooperative behavior, bargaining power) or to identify profiles within your sample (e.g., risk loving or risk averse individuals, more or less altruistic individuals, individuals with more or less bargaining power).
- While a game is not the same as a real-world decision, to create a sense of real stakes, offering small cash or mobile phone credit incentives or an entry into a lottery for a larger incentive as part of the game can help better approximate how people would respond if the stakes were real.

EXAMPLE:

In an evaluation in western Kenya, couples were offered the opportunity to open individual or joint bank accounts. Schaner (2017) used a game to measure intra-household bargaining power to test how men and women's relative bargaining power affected saving behaviors. Enumerators played a game with husbands and wives at the end of the baseline survey. Husbands and wives, who were being surveyed separately, were asked to divide a small cash prize between themselves and their spouse. Each spouse recorded his or her allocation separately on cards and placed the amount they allocated to themselves in their tin and the amount allocated to the spouse in the spouse's tin. Then they came together to decide how to allocate the prize money between them and recorded it on cards, which they added to each of their tins. To ensure respondents' privacy, the husband and wife also each added an additional envelope to each of their tins with a randomly selected amount of money. The husband and wife then chose one card from each of their tins and were immediately given the cash amount allocated to them on the card they chose. This game allowed researchers to identify women with relatively low or high bargaining power and test whether the impact of the intervention was different for women with higher or lower bargaining power. Women with larger differences between their individual and joint preferences for allocating the money between themselves and their husbands were classified as having relatively lower bargaining power.

Results showed that providing ATM cards led to very different results between men and women once bargaining power was taken into account. In particular, women with low bargaining power opened fewer joint accounts on average when they had access to the ATM card, likely because they preferred to use individual accounts to keep savings for their own purposes and reduce the likelihood of the money being used by their husbands. The game script is available on Schaner's website.[4]

CITATION:

Schaner, Simone. 2017. "The Cost of Convenience? Transaction Costs, Bargaining Power, and Savings Account Use in Kenya." *Journal of Human Resources* 52(4): 919-945. https://muse.jhu.edu/article/674025.

[4] Schaner, Simone. "Bargaining Game Script: The Cost of Convenience? Transaction Costs, Bargaining Power, and Savings Account Use in Kenya." Research. Accessed June 8, 2018. https://sites.google.com/site/sschaner/home/research.

4. VIGNETTES

Vignettes are brief descriptions of hypothetical scenarios. They can be useful in many different aspects of research, from reducing the risk of social desirability bias in survey questions about sensitive topics by asking about a hypothetical scenario, to helping clarify the meaning of concepts in our survey questions and ensure that participants understand them in the same way. Experimental vignettes—in which study participants are randomly assigned to hear one of multiple versions of the same story with a key detail changed—can also be used to measure subconscious biases or prejudices based on gender, race, ethnicity, or other factors. To measure gender bias, for example, researchers could create two identical versions of a story in which only the gender of the subject changes and randomly assign respondents to hear one of the two versions. This allows researchers to isolate the extent to which the difference in responses was caused solely by the gender of the subject in the hypothetical scenario described in the vignette. Vignettes can also be used to measure attitudes about gender norms. For instance, researchers can describe a hypothetical scenario with fictional people (e.g., a family deciding not to send their daughter to school; a couple deciding to use contraception) and ask the respondent to what extent they agree or disagree with the decision made in the scenario.

PROS:

Vignettes have the potential to reveal information participants may not be willing to admit freely about themselves (bias or attitudes about gender).

CONS:

In experimental vignettes where two versions of the same story are randomly assigned to respondents, typically only one detail in the vignette can be changed in order to isolate the reason for differing perspectives. It can therefore be a relatively expensive way to measure one outcome compared to survey questions, though this is not necessarily the case if the vignette is short. In addition, people may perceive they are being tested and respond with a socially desirable answer rather than what they truly believe.

TIPS:

- Vignettes can be most useful when changing attitudes about or bias against women is one of the main outcomes being measured in the evaluation.
- Vignettes can also be used to help make sure respondents have a consistent understanding of a concept in a question.
- Conduct thorough piloting and work with partners to ensure the scenarios described in the vignettes are appropriate for the local context.

EXAMPLE:

In the evaluation of reservations for female leaders in village councils in India, researchers played a short recording of a speech by a local leader responding to a complaint from a villager. Respondents were randomly assigned to hear the same recording spoken by a man or woman. After the speech was over, they were asked to rate the leader's performance and effectiveness. This vignette allowed researchers to measure whether there was a subconscious bias that led people to rate female leaders as relatively less effective. They found that exposure to a female leader through the policy that reserved village council head positions for women improved men's evaluations of female leader effectiveness. Scripts for the vignettes are available on the Harvard Dataverse.[5]

CITATION:

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124 (4): 1497-1540. https://doi.org/10.1162/qjec.2009.124.4.1497.

[5] Beaman, Lori; Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2017. "Powerful Women and Aspirations in India." Harvard Dataverse, V2. https://doi.org/10.7910/DVN/PXV79W.

5. IMPLICIT ASSOCIATION TESTS (IATS)

LOCATION: INDIA. PHOTO: THOMAS CHUPEIN | J-PAL

Social psychologists developed IATs as a way to measure how strongly people associate two concepts with one another.[6] In an IAT, a respondent is shown a word or picture on a screen and asked to quickly sort it into one of two categories. The hypothesis is that respondents who pair a set of concepts more quickly during the test associate those concepts more strongly. IATs have therefore been used as a way to quantify implicit biases or stereotypes that people may not be willing to talk about openly with an interviewer, or may not even know they hold, including implicit gender biases.

However, there is an ongoing debate in psychology about the extent to which the speed of categorization is a good proxy for implicit bias and whether we can learn anything about people's behavior in the real world from this proxy measure (i.e. in some cases people may hold an implicit bias but not act on it). Results from IATs can also be sensitive to the framing and order of the questions, and collection of concepts referenced in the sorting exercise.[7] IATs need to measure response time down to the millisecond, and thus have to be administered using computers, tablets, or smartphones. You can see what it is like to take IATs about gender, race, ethnicity, and other topics at Harvard University's Project Implicit website.[8]

[6] Greenwald, Anthony G., Debbie E. McGhee, and Jordan LK Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74 (6): 1464-1480. https://doi.org/10.1037/0022-3514.74.6.1464.

[7] Researchers have also found low stability in people's IAT scores over time. Blanton, Hart, and James Jaccard. 2008. "Unconscious racism: A concept in pursuit of a measure." *Annual Review of Sociology* 34: 285-292. https://doi.org/10.1146/annurev.soc.33.040406.131632.
Payne, B. Keith, and Heidi A. Vuletich. 2017. "*Policy Insights From Advances in Implicit Bias Research*." Policy Insights from the Behavioral and Brain Sciences 5 (1): 49-56. 51. https://doi.org/10.1177/2372732217746190.

[8] Project Implicit. 2011. https://implicit.harvard.edu/implicit/.

5. IMPLICIT ASSOCIATION TESTS (IATS)

PROS:

IATs are one of the few measurement tools that may help reveal subconscious biases, which traditional self-reported survey questions cannot typically pick up.

CONS:

IATs measure one specific approximation of implicit bias at a particular point in time. They do not speak to whether and how people act based on their subconscious stereotypes in real life. There is also an ongoing debate in psychology about whether they are a good measure of subconscious stereotypes.[9]

TIPS:

- IATs are most useful when prejudice or implicit bias against women is one of the main outcomes of interest and survey questions cannot be designed to adequately mitigate reporting bias.
- IATs may be most useful as a measure of implicit biases at an average or aggregate level rather than individual level.[10]
- Think through how the test framing could affect the test results. For example, one study found different amounts of measured bias with small changes in how the IAT was framed for test takers.[11]

EXAMPLE:

In an evaluation to measure whether exposure to female leaders in Indian village councils changed community members' perceptions about women's effectiveness as leaders, Beaman et al. (2009) administered an IAT to respondents during a household survey. In the test, respondents were shown two different screen configurations and asked to sort photographs to align with the configurations. In the "stereotypical" configuration, the screen showed a picture of a man politician combined with "good" attributes (happy faces, positive words) on one side and a picture of a woman politician combined with "bad" attributes (sad faces, negative words) on the other side. Respondents then saw a series of pictures of men or women engaging in political activities such as leading crowds or giving speeches. They were asked to click a button to select whether the picture belonged to the right- or left-hand side of the screen. Respondents completed the same test after being shown a "non-stereotypical" configuration: a picture of a woman politician combined with good attributes on one side of the screen and a picture of a man politician combined with negative attributes on the other. Their response times were recorded in milliseconds.

The test measured the relative strength of people's association of positive qualities with men leaders and negative qualities with women leaders by comparing response time across the stereotypical and non-stereotypical configurations. For example, if a respondent associated male politicians with positive qualities and women politicians with negative qualities, the sorting process would be faster under the "stereotypical" configuration. The researchers then calculated the difference in the average response times between the stereotypical and non-stereotypical configurations. This was one of the first examples of an IAT being adapted to a context with high rates of illiteracy.

CITATION:

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124 (4): 1497-1540. https://doi.org/10.1162/qjec.2009.124.4.1497.

---

[9] Payne and Vuletich 2017; Blanton and Jaccard 2008.

[10] Payne and Vuletich 2017, 49, 51.

[11] Van Nunspeet, Félice, Naomi Ellemers, Belle Derks, and Sander Nieuwenhuis. 2012. "Moral Concerns Increase Attention and Response Monitoring During IAT Performance: ERP Evidence." *Social Cognitive and Affective Neuroscience* 9 (2): 141-149. https://doi.org/10.1093/scan/nss118.

6. LIST RANDOMIZATION

List randomization can be used to create a proxy measure of the prevalence of a sensitive or socially undesirable behavior. When measuring women's empowerment in some contexts, for example, we may want to determine the prevalence of unprotected sex and women's control over decisions about contraceptive use. List randomization involves giving half of the randomly assigned study participants a short list of activities, including the sensitive behaviors being tracked and asking them how many activities they have engaged in without asking them to specify which ones. The other half of the study participants receive the same list without the sensitive behaviors being tracked and also respond how many activities they have engaged in. Researchers can take the average of the two groups; the prevalence of the sensitive behaviors will be indicated by the difference in the averages.

PROS:

List randomization enables researchers to measure behaviors that individuals might not typically self-report.

CONS:

The instrument does not provide individual-level information on a behavior, only aggregate-level estimates; we can only compare average responses in treatment and comparison groups. This makes analysis that relies on individual-level outcomes difficult. In addition, the technique requires participants to count or add as part of the response, which could introduce noise in the data if people accidentally miscount. Additionally, for the participants who engage in all the behaviors in the list including the sensitive one, list randomization does not make their answers private so they may still feel uncomfortable answering honestly. Allowing participants to enter their responses into a tablet privately without the interviewer being able to see their answers may help address this challenge.[12]

TIPS:

- Include this section in the middle of the survey, after respondents have gotten more comfortable with the enumerator.
- Consider allocating a larger proportion of the sample to the group being asked about the sensitive behaviors, since we lose statistical power by splitting the sample (i.e., not everyone is asked about the sensitive behaviors).[13]

EXAMPLE:

In an evaluation in Uganda, researchers wanted to understand how text messages delivering health education information affected the prevalence of certain sexual behaviors, such as condom usage and number of sexual partners. Jamison, Karlan, and Raffler (2013) randomly divided the respondents into two groups. Individuals in the first group received a list of four neutral statements and reported how many of the statements were true for them. Individuals in the second group received a list of five statements—four neutral and one sensitive—and reported how many were true for them. For example, one of the sensitive statements included, "I did not use a condom the last time I had sex." The difference in the average number of statements respondents reported as true for them between the two groups represented the prevalence of the sensitive behaviors being tracked, such as condom usage and number of sexual partners. The study found that the texting service did not decrease risky behavior, and in fact, the number of sexual partners increased in the treatment group.

CITATION:

Jamison, Julian C., Dean Karlan, and Pia Raffler. 2013. "Mixed Method Evaluation of a Passive mHealth Sexual Information Texting Service in Uganda." *Information Technologies and International Development* 9, 1-28. http://dev.itidjournal.org/index.php/itid/article/view/1087.

[12] Audio Computer-Assisted Self-Interview (ACASI) technology can make this possible in contexts with low literacy.

[13] For more on this and other factors to consider when using list randomization techniques, see: Ozler, Berk. 2017. "List Experiments for Sensitive Questions: A Methods Blog." World Bank Development Impact Blog. Accessed April 23, 2018. https://blogs.worldbank.org/impactevaluations/list-experiments-sensitive-questions-methods-bleg

7. PARTICIPATORY RESEARCH METHODS

Participatory research methods involve discussions and interactive exercises with groups of community members in order to understand local needs and opportunities from their perspectives. Interactive exercises include mapping or creating community timelines. There are many versions of this methodology with slightly varying focus areas, including participatory resource appraisal, rapid resource appraisal, rapid rural appraisal, participatory learning and action, among others.

Asking a group to collectively provide information on a topic can sometimes yield better information than simply surveying individuals. This is particularly useful when each individual may only know some of the information, but conversation and prompts from others elicits more detail. For example, the Participatory Resource Appraisal (PRA) approach can be used to understand and map the resources available in a community and how these interact with gender. They can be used to identify which girls, women, or families may be eligible for some programs. Participatory approaches are also good for gathering historical information or mapping social connections. For more information on the many different ways participatory research methods can be used and how to design and carry out participatory research, see the Institute of Development Studies' website.[14] While these methods can help increase marginalized groups' participation and voice in research, some have critiqued participatory research for not fully accounting for the ways in which local and international power dynamics affect who contributes and how.[15]

PROS:

Participatory methods can generate richer and more accurate data than surveying individuals for some kinds of community-level data like the existence and condition of infrastructure in a given place, mapping networks, and collecting data on a community's history. When used in formative research and needs assessments, they can also create space for people participating in our study to contribute ideas to the diagnosis of the problem, the design of the program, and/or the outcomes we measure in the evaluation.

CONS:

Power dynamics can affect the conversation, and some people may be unwilling to say certain things in front of others. The presence of the community elite, for example, may alter the content of the conversation. People may find it difficult or inappropriate to admit to not knowing something when asked, or contradict what another person says.

TIPS:

- It is important to be mindful of local power dynamics when selecting the participants. There may be a risk of powerful individuals dominating the direction of the conversation. In addition, women may sometimes hesitate to share their opinions when men are present, so it can be useful to repeat the exercise with different gender compositions. It is also important to consider which individuals may be most knowledgeable about the specific issue being explored.
- One of the main goals of participatory methods is to generate discussion to capture multiple perspectives. We should be sure to draw out the conversation and probe areas of disagreement and consensus.

EXAMPLE:

In India, Chattopadhyay and Duflo (2004) tested how quotas for women in local politics affected the provision of public goods. In particular, they assessed to what extent the number and/or quality of wells differed in communities where the village council head position was reserved for a woman. To find out about the history of wells in each community, the researchers gathered groups of people in a discussion using the PRA approach. The PRA included drawing maps of the community, locating the wells, and discussing when the wells were built and repaired. Different individuals volunteered different pieces of information, which generated a more complete history of the wells. In addition, the enumerators selected some of the community members to go on a walk around the community and double-check the accuracy of the maps. The guide used to facilitate the PRA is available on the Harvard Dataverse.[16]

CITATION:

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409-1443. https://doi.org/10.1111/j.1468-0262.2004.00539.x.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics* 124 (4): 1497-1540. https://doi.org/10.1162/qjec.2009.124.4.1497

[14] Institute of Development Studies. N.d. "Participatory Research Methods." Accessed April 25, 2018. http://www.participatorymethods.org.

[15] Cooke, Bill, and Uma Kothari. 2001. *Participation: The New Tyranny?*. London: Zed Books.

[16] Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2017. "Powerful Women and Aspirations in India." Harvard Dataverse, V2. https://doi.org/10.7910/DVN/PXV79W.

8. SOCIAL INTERACTION AND NETWORK EFFECTS

LOCATION: INDIA. PHOTO: THOMAS CHUPEIN | J-PAL

Social capital can be an important component of empowerment. The size and influence of one's social network can contribute to the amount and types of opportunities and choices a person can access. Researchers can collect basic information on whom various individuals in a given community know and talk to in order to map the social relationships in their community. This instrument can enable researchers to understand whether there are distinct subgroups in communities' social networks and whether and how they are connected to each other and communicating. In addition, mapping respondents' social networks can help provide a proxy measure of an individual's social capital. It can also be used to measure how ideas or attitudes spread through communities and identify which types of people tend to be more influential in spreading them. While surveys are typically used to collect social network data, the analysis and interpretation processes are different from those in traditional surveys.

For in-depth insights on social network analysis, see Breza's 2016 overview of how social network analysis has been used in randomized evaluations.[17] In addition, Banerjee et al. (2013) analyzed social networks to understand the diffusion of microcredit in southern India, offering additional insights on the methodology and applications.[18]

[17] Breza, Emily. 2016. "Field Experiments, Social Networks, and Development." *The Oxford Handbook of the Economics of Networks*, Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199948277.013.25.

[18] Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2013. "The Diffusion of Microfinance." *Science* 341 (6144): 363-370. https://doi.org/10.1126/science.1236498.

8. SOCIAL INTERACTION AND NETWORK EFFECTS

PROS:

Building social relationships can be an important part of the empowerment process—especially when it involves participating in collective action, joining new groups, or accessing new resources. Understanding women's interactions and connections can help us understand the relationship between social capital, empowerment, and the spread of ideas.

CONS:

Collecting social network data is time intensive and expensive, as every social link in the community must be accounted for. Small errors can have large consequences: if a participant misremembers one name during a survey it can dramatically change the entire map. Things like nicknames or multiple people that have the same name can be challenging to account for.

TIPS:

- Social network data can be collected with surveys or participatory appraisals.
- If we plan to use social network analysis in our research, it may be useful to consult with an expert on the topic early on while we are still designing the evaluation to plan the data collection and analysis properly.

EXAMPLE:

In a study in Bangladesh, Miller and Mobarak (2014) researched how various marketing interventions affected take-up of health-protecting clean cookstoves (an appliance primarily used by women), including the role that social networks play. The study measured demand for both stoves designed to protect health and stoves that were just more fuel-efficient. The researchers conducted a first round of the study promoting improved stoves using several marketing strategies. They then tracked improved stove adoption among the first round of households and gathered extensive information on their social networks. The researchers then offered stoves to the social network members of the first round of households and monitored how social relationships affected the decision to purchase an improved stove. The study found that first round households' social network members were less likely to purchase a stove when offered. The researchers conclude this was likely because the improved stoves were prone to break, and the news of faulty stoves discouraged social network members from purchasing it.

CITATION:

Miller, Grant, and A. Mushfiq Mobarak. 2014. "Learning about New Technologies through Social Networks: Experimental Evidence on Nontraditional Stoves in Bangladesh." *Marketing Science* 34 (4): 480-499. https://doi.org/10.1287/mksc.2014.0845.

9. BIOMARKERS

Since biomarkers cannot be manipulated or changed by the respondent, they offer an objective method for measuring many health outcomes. Biomarkers can be used to measure both direct effects of a program (e.g., an evaluation of a girls' nutrition intervention may measure body mass index, weight for height, and/or height for age) and indirect effects (e.g., an evaluation of a women's savings group program may measure women's cortisol levels to track stress). Biomarkers that might be relevant in programs measuring women's empowerment include STI and HIV tests, height, weight, BMI, arm circumference, saliva swabs for cortisol (a stress hormone), and diagnostic tests for illnesses. Certain biomarkers such as cortisol can potentially serve as proxy measures for elements of women's empowerment. For example, experiencing a lack of control or autonomy in the work place is a primary source of work-related stress. Thus, measuring stress through salivary cortisol swabs could signal woman's feeling of agency in her work. When combined with responses to survey questions, this could help us gain a fuller understanding of the process of women's empowerment in labor market choices.

Different kinds of biomarkers require different levels of ethical clearance and actions by the research team. For example, we may be ethically obligated to refer individuals to treatment to support those who are diagnosed with conditions through the biomarker data collection. It is important to consult our IRB's guidelines and resources on collecting biological data before designing our study to ensure compliance before submitting our IRB application for review.

PROS:

Biomarkers are less biased than self-reported information and can be used to objectively measure many important indicators related to health and well-being. They can also help triangulate outcomes that we are measuring with survey questionnaires.

CONS:

Collecting biomarkers often requires more in-depth training for field research staff. Some biomarkers can be very expensive to collect. For instance, diagnostic tests that have to be processed in a lab tend to be more expensive than biomarkers that can be measured during a survey. Refusal rates for intrusive biomarkers (e.g., those that involve drawing blood) can be high and response rates may be different in treatment and comparison groups, so we should think carefully about whether they are critical for our analysis before using them. This is generally not a problem with less intrusive biomarkers like nutrition indicators.

TIPS:

- We should weigh the costs and benefits of using biomarkers carefully and use them primarily when they are critical to our plan for analysis (i.e., a main outcome of interest).
- We need to consider whether we can reasonably expect there to be sufficient variation in the biomarker to detect differences between the treatment and comparison groups to justify collecting it.

EXAMPLE:

In an ongoing study in Bangladesh, Lopez-Peña et al. (ongoing) are evaluating the impact of Cognitive Behavioral Therapy (CBT) on women garment sector workers' stress levels and well-being. Many of the women migrated from rural areas and have limited support networks in their new homes in the city. The researchers will measure salivary cortisol along with self-reported symptoms of anxiety to understand the impact of the CBT program on stress levels.

CITATION:

Lopez-Peña, Paula, Md. Kamruzzaman Mozumder, Atonu Rabbani, and Christopher Woodruff. 2016. "Worker Well-Being and Productivity in the Bangladesh Garment Sector." AEA RCT Registry. https://www.socialscienceregistry.org/trials/738/history/7330.

ABOUT J-PAL

The Abdul Latif Jameel Poverty Action Lab (J-PAL) is a global research center working to reduce poverty by ensuring that policy is informed by scientific evidence. Anchored by a network of more than 160 affiliated professors at universities around the world, J-PAL draws on results from randomized impact evaluations to answer critical questions in the fight against poverty.

povertyactionlab.org

ABOUT THE GENDER SECTOR

Gender norms and biases continue to constrain human potential around the world. J-PAL's Gender sector produces cross-cutting insights on promoting gender equality and women's and girls' empowerment and on how social norms related to gender affect the outcomes of social programs.

povertyactionlab.org/gender