

DOCUMENTOS DE TRABAJO

Privacy and Measurement Error in Phone Surveys: The Case of Women's Agency

Jorge M. Agüero, Julián Pierino Pedrazzi,
Viviana M.E. Perego y Javier Romero

Documento de Trabajo Nro. 384

Octubre, 2026

ISSN 1853-0168

www.cedlas.econo.unlp.edu.ar

Cita sugerida: Agüero, J.M., J.P. Pedrazzi, V.M.E. Perego y J. Romero (2026). Privacy and Measurement Error in Phone Surveys: The Case of Women's Agency. Documentos de Trabajo del CEDLAS N° 384, Octubre, 2026, CEDLAS-Universidad Nacional de La Plata.

Privacy and Measurement Error in Phone Surveys: The Case of Women's Agency*

Jorge M. Agüero Julián Pierino Pedrazzi Viviana M.E. Perego
Javier Romero

October 17, 2026

*We dedicate this paper to the loving memory of Jorge Agüero,
our brilliant, kind, and fun colleague.*

Abstract

There is a growing literature focusing on measurement error in self-reported data. This paper studies the case of women's agency, a key outcome in economic development, and introduces a simple and cost-effective corrective measure to enhance the reliability of data collection around sensitive topics. The study draws from a survey experiment in rural Guatemala where 1,183 women received the same questions about their agency as part of a phone survey, but the privacy of their responses was varied. Women in the control group answered verbally. The first treatment group used code words to conceal answers from potential bystanders. The second treatment arm used the phone keypad to conceal answers from bystanders and the enumerator alike. The paper finds that women answering under higher levels of privacy report lower levels of agency, by 0.19 standard deviations on an aggregate index. The effects of the two privacy treatments cannot be statistically distinguished. The effect is concentrated among women with higher social standing, and qualitative interviews indicate that agency is socially valued in this setting. Both patterns point to social desirability bias as the most likely explanation for why reported agency responds to privacy.

JEL classification: C93, J16, O12

Keywords: Field experiments, measurement, survey design, gender, women's agency.

*The authors acknowledge generous funding from the Research Methods Initiative (RMI) at Innovations for Poverty Action (IPA), the InfoDev Trust Fund through its Entrepreneurship for Development call for proposals, the Agriculture and Food Global Practice of the World Bank, the Gender Innovation Lab for Latin America and the Caribbean of the World Bank, and the Umbrella Facility for Gender Equality (UFGE). The UFGE has received generous contributions from Australia, Canada, Denmark, Finland, Germany, Iceland, Ireland, the Netherlands, Norway, Spain, Sweden, Switzerland, the United Kingdom, the United States, the Bill and Melinda Gates Foundation, and the Wellspring Philanthropic Fund. The authors are grateful for useful comments received at the Measurement Conference coorganized by Innovations for Poverty Action and the Global Poverty Research Lab at Northwestern University, the International Economic Association World Congress, and the Congreso Anual de la Asociación Peruana de Economía. The authors thank Jony Ji Araujo, Sofia Collante, Guillermo Delgado, Arianna Locatelli, Angela R. Lopez, Danielle Marie Orihuela, Glenn Ortiz, Rudy Ramirez, Ignacio Rodriguez Hurtado, Rosa Rojas, Carolina Saavedra, and Hugo Salas for excellent research assistance and field coordination at various stages of this project. Kelly Montano provided valuable research management support. Andrea Tock led and supervised the qualitative fieldwork and conducted analysis over qualitative surveys. The study was reviewed and approved by Health Media Lab IRB (protocol 868TWBG21) and registered at the American Economic Association's registry for randomized controlled trials (RCT ID: AEARCTR-0008551). The authors thank Amparo Palacios-Lopez, Amber Peterman, and Astrid Pineda for their insightful comments. Contact information: Perego: The World Bank; email: vperego@worldbank.org. Pedrazzi (corresponding author): CEDLAS-Universidad Nacional de La Plata and CONICET; email: pedrazzi.julian@gmail.com; Romero: The World Bank and Lima School of Economics at Universidad de Piura; email: javierromero@worldbank.org. The findings, interpretations, and conclusions expressed in this work are entirely those of the authors and should not be attributed in any manner to The World Bank, its Board of Executive Directors, or the governments they represent. All errors are our own. 1

1 Introduction

Empowering women is fundamental to reducing inequality, protecting human rights, and boosting global well-being. Reflecting this, the fifth UN Sustainable Development Goal calls on countries to “achieve gender equality and empower all women and girls” (United Nations, 2024). This global agenda is supported by empirical findings that link women’s empowerment with positive outcomes in nutrition (Allendorf, 2007; van der Meulen Rodgers and Kassens, 2018), education (Deininger et al., 2019; Edmonds et al., 2023), and agricultural productivity (Dillon and Voena, 2018). Underlying these outcomes is women’s agency, typically measured through a woman’s participation in household decisions and her freedom to act (Kabeer, 1999).

However, measuring women’s agency is difficult for several reasons. Agency is a multidimensional concept that is inherently hard to quantify (Costa et al., 2023). Measurement can further be affected by who in the household reports on it (Ambler et al., 2021), how its indicators are constructed (Peterman et al., 2021), the characteristics of surveyors (Álvarez-Aragón and Champeaux, 2026; Di Maio and Fiala, 2020), and misreporting driven by the sensitivity of certain questions (Álvarez-Aragón and Champeaux, 2026; Kadam et al., 2025; Cullen, 2023; Iqbal et al., 2025).

Misreported levels of women’s agency can lead to flawed evidence about what empowers women, potentially undermining policy and program design (Agüero et al., 2026). Targeting programs using inaccurate agency measures may result in interventions that miss those most in need. Developing accurate survey techniques is thus critical, and recent studies have examined how surveys can be structured to capture women’s agency more reliably (Donald et al., 2020, 2024; Annan et al., 2021).

This paper uses a phone survey experiment to test whether women report their agency differently when the privacy of their responses varies. The study recruited 1,183 rural women in Western Guatemala. The experiment had three groups to which participants were randomly assigned. All groups received the same questions during a phone interview,

following a short questionnaire originally proposed by [Jayachandran et al. \(2023\)](#) and adapted to the Guatemalan context; what changed across the experimental groups was the way in which participants could provide their answers.

Specifically, while the control group responded directly to the enumerator, two treatment arms were designed to conceal responses from bystanders, and from bystanders and the enumerator, respectively. The control group received questions on agency directly from the enumerator, to which subjects had to reply verbally with one of the given options, like in a regular interview. For example, the interviewer asked “Who usually decides if you can go to visit the house of a friend or neighbor?”, to which the interviewee could respond “myself”, “my partner”, or “someone else”. In the first treatment arm, instead, respondents answered using code words, such as “one”, “two”, and “three” for answers such as “myself”, “my partner”, and “someone else”, respectively. Thus, in this arm, the respondent did not risk disclosing potentially sensitive information to individuals in her surroundings. In the second treatment arm, respondents replied through their phone keypads, pressing numbers to submit their responses, thereby avoiding disclosing their answers either to bystanders or the enumerator themselves.

The paper finds that building privacy into phone surveys alters how women report agency. Results show that adding privacy induces women to report *lower* levels of agency vis-à-vis direct reporting, reducing an aggregate index of agency by 0.19 standard deviations. Interviewees in groups with more privacy are more likely to acknowledge that they need permission to make small purchases and that they have less decision-making power over their children’s health. The results are consistent across specifications and robustness checks. Interestingly, tests of equality of the effects across the treatment arms cannot be rejected, suggesting that the additional privacy of concealing responses from the enumerator does not add much beyond concealing them from bystanders.

To explore why reported agency responds to privacy, this paper considers the two broad sources of misreporting identified in the literature on sensitive questions: social desirability, whereby respondents adjust their answers toward what they believe is socially

acceptable, and threat of disclosure, whereby they fear the consequences of a third party learning their answer (Tourangeau et al., 2000; Tourangeau and Yan, 2007). The evidence is most consistent with the former. Qualitative interviews with a random subsample of participants, coded in an approach that recent work finds informative about women’s agency (Clingain et al., 2026), show that agency is socially valued in this setting: 76 percent believe their community would approve of women having agency and their own voice (Blair et al., 2020). In line with this, the heterogeneity results show stronger effects of privacy for women with higher social standing: respondents who are the household head, completed primary education, or are single, with a similar but less precisely estimated pattern for landowners (Table A4). This is consistent with this group being under the most pressure not to deviate from what is normatively valued (Bernstein et al., 2001; Silver et al., 1986).

Threat of disclosure finds less support in the data. Fear of retaliation is a common concern in the domestic violence literature, where respondents report on potentially criminal conduct and the risk is immediate (Ellsberg et al., 2001; Vranda et al., 2018). In this setting, added privacy could analogously allow women to report low agency without fear of reprisal, for example, from a controlling partner. However, reprisal risk is usually higher among more vulnerable women (Mabena et al., 2025; Baruwa et al., 2025; Acharya, 2025), which contradicts the heterogeneity results.

Other explanations identified in the literature also seem inconsistent with the evidence. One is survey miscomprehension, whereby treated respondents report less agency simply because they did not understand the response mechanisms (Park et al., 2025). However, a placebo designed in the experiment argues against survey miscomprehension: respondents in all three arms answered a non-sensitive question about the harvest season using the same response method as the agency questions, and reporting did not differ across arms. Surveyor characteristics could also lead to differential reporting if correlated with the treatment arms (Di Maio and Fiala, 2020; Huddy et al., 1997). The experiment randomized participants within surveyors, and results remain unchanged after controlling

for surveyor effects. Construct validity is another key topic that could lead to differential responses (Peterman et al., 2021), but by design the exact same questions were asked in the three arms; only the response mode was changed. The experiment, however, was designed to vary privacy rather than to isolate the drivers behind it. This analysis is thus suggestive rather than conclusive, and designs built for that purpose are left for future research.

This paper contributes to several efforts to develop frameworks and survey modules for measuring women’s agency and empowerment (Donald et al., 2020; Jayachandran et al., 2023; Malapit et al., 2019). Indeed, a recent review finds that while research on women’s agency is blossoming, no consensus has emerged on survey modules or delivery methodologies (Costa et al., 2023). Kadam et al. (2025) randomly vary the survey mode (phone or in-person) and the enumerator’s gender for a module on gender norms and women’s life experiences among ultra-poor women in rural Ethiopia, finding that responses are, on average, similar across arms. This paper is closest to Álvarez-Aragón and Champeaux (2026), who compare face-to-face interviews with a human-assisted self-administered mode in Benin and likewise find that women report less gender-equal outcomes under the more confidential method. Relative to that study, our design varies privacy while holding the rest of the interview fixed. Their comparison changes the setting, device, and enumerators at once, so the estimate captures these features jointly rather than separately.¹ Contributing to this result, ours is the first experimental evidence that the privacy of the response itself changes how women report their agency.

This paper also builds on research into survey methods for sexual violence and sexual behavior, where added privacy typically increases disclosure. A systematic review by Langhaug et al. (2010) finds that private, self-administered responses raise reporting throughout the developing world, and computer-assisted surveys have had this effect in Malawi, Zimbabwe, and Ethiopia (Kelly et al., 2013; Langhaug et al., 2011; Lindstrom

¹Álvarez-Aragón and Champeaux (2026) conducted fieldwork in 2021–2022; fieldwork for our study was conducted in 2021. Álvarez-Aragón and Champeaux (2026) cite the 2023 working-paper version of our study.

et al., 2012); Cullen (2023), Peterman et al. (2025), and Joseph et al. (2017) are recent examples from economics. Some studies, however, find that privacy does not increase sensitive reporting and may even lower it (Agüero and Frisanco, 2022; Gilligan et al., 2025). Several rely on list experiments, whose complex instructions participants may not follow (Chuang et al., 2021), but the concern is broader: privacy gains can come at the cost of comprehension in low-literacy settings (Park et al., 2025). This paper offers a simpler alternative easy to implement across a broad range of settings and populations.

Finally, this paper also contributes to the broader literature examining survey methods for sensitive topics, not just women’s issues. Tourangeau and Yan (2007) organize this literature around the sources of question sensitivity, distinguishing intrusiveness, threat of disclosure, and social desirability. Gnambs and Kaspar (2015) provide a meta-review mainly focused on the US, finding that more private computer-assisted surveys increase response rates for risky teen behaviors such as drug use. Höglinger et al. (2016) examine survey methods for sensitive student behavior such as cheating, while Makkai and McAllister (1992) and Corkrey and Parkinson (2002) examine the impact of private survey response methods on the study of drug and alcohol use. Tanton et al. (2023) use different survey methods to study violence occurring within schools, and Karlan and Zinman (2012) use a list experiment to study how microloan proceeds are used. This paper contributes evidence from a behaviour respondents have reason to display rather than conceal, and shows that privacy lowers rather than raises what they report.

The rest of the paper is organized as follows. Section 2 introduces the context of rural Guatemala and the setting for the survey experiment. Section 3 outlines the experimental design, sample, and data collection process. Section 4 describes the empirical strategy. Section 5 presents the main results, heterogeneous effects, robustness tests, and a discussion of the potential mechanisms. Section 6 discusses the implications of the paper and concludes.

2 Context: Women in Rural Guatemala

Guatemala ranks 127th out of 189 countries in the Gender Inequality Index of the United Nations' Development Programme ([UNDP, 2020](#)). Gender gaps in the country are widespread and span from decision-making, to social and political participation, access to resources, time-use, and gender-based violence. Levels of disempowerment are particularly acute in rural areas, where women act as secondary agricultural workers or helpers to their husbands, fathers, or brothers: while official figures see women as constituting only around 10 percent of agricultural employment, this percentage rises to 40 percent when counting women engaged in unpaid farm work or otherwise supporting their family farms ([INE, 2019](#)). Furthermore, 57 percent of women employed in agriculture do not earn a salary, and of those that do, 97 percent earn below the minimum wage ([INE, 2015](#)).

Women in agriculture also face barriers in terms of human capital and access to productive assets. The 2018 Guatemala National Census shows that only 53 percent of women in agriculture can read and write, versus 71 percent of men. At the same time, only 7.8 percent of landowners are women ([FAO, 2016](#)), and the percentage of women in land access programs has also been historically low (only 10.7 percent of beneficiaries of these programs were women in 2014, cf. [Moreno, 2020](#)).

The analysis in this paper builds on the World Bank DIGITAGRO pilot to improve market access for women smallholders in the ambit of Guatemala's School Feeding Program ([Lopez et al., 2025](#); [Perego et al., 2022](#)). DIGITAGRO operated in the department of San Marcos in Western Guatemala, and it was expected that, by providing greater awareness about the School Feeding market, information on how to participate in it, and information on how to achieve higher-quality agricultural produce, the pilot would increase market participation and improve sales and prices.

The project was implemented in 2020–2021 and relied heavily on digital technologies for information dissemination and data collection, including phone surveys. The survey

experiment described in this paper was conducted after the DIGITAGRO project. As explained later, the DIGITAGRO sample was complemented with an additional top-up sample to increase the power of the survey experiment.

3 Experiment Description

The survey experiment took advantage of the ongoing DIGITAGRO impact evaluation in San Marcos (Perego et al., 2022; Lopez et al., 2025). During the impact evaluation endline survey, a survey experiment was conducted to test whether building privacy into a survey affects the reporting of markers of agency among participants. The survey was conducted in August of 2021.

The survey experiment had three experimental arms. The first arm was the control group, which received questions on agency directly from the interviewer during the phone interview. The interviewee replied with one of the given options, like a regular interview. For example, the interviewer asked “Who usually decides if you can go to visit the house of a friend or neighbor?”, to which the interviewee could respond “myself”, “my partner”, or “someone else”.

In the second arm, the **“Code words”** treatment group, respondents received the same set of questions, but they answered using code words, such as “one”, “two”, and “three” for answers such as “myself”, “my partner”, and “someone else”, respectively. Thus, in this arm, the respondent did not risk disclosing potentially sensitive information to individuals in her surroundings.

In the third arm, the **“Keystroke”** treatment group, respondents received the exact same set of questions but they replied with their phone keypads, pressing numbers to submit their responses. Thus, in the third arm, interviewees did not risk disclosing their answers to bystanders and the enumerator alike. The interviews were audio recorded and then decoded by mapping the tone produced by each keystroke (i.e., tone dialing, also known as dual-tone, multi-frequency system or DTMF) to the options provided by the

question.

As shown later, analysis which separates out the treatment arms yields largely similar results. Thus, for the main analysis of the paper, the “Code words” and “Keystroke” treatments were pooled into a single variable.

3.1 Survey experiment structure

The treatment and control groups received the same set of six questions on agency. These questions were taken from the proposal of [Jayachandran et al. \(2023\)](#) for a short questionnaire to measure women’s agency, and adapted to the Guatemalan context when possible using the Project-level Women’s Empowerment in Agriculture Index (Pro-WEAI). Pro-WEAI is a survey-based index for measuring empowerment, agency, and inclusion of women in the agricultural sector in a standardized manner developed by the International Food and Policy Research Institute (IFPRI), the Oxford Poverty and Human Development Initiative (OPHI), and 13 partner projects in IFPRI’s GAAP2 portfolio ([Malapit et al., 2019](#)).²

The verbatim questions and possible answers in each experimental group were as follows:

1. Are you allowed to go alone to meet up with your girlfriends and friends for any reason (chat, have lunch, etc.)?
 - Possible answers for the Control Group: say “Yes” or “No”.
 - Possible answers for the Code words Group: say the code words “one” or “two” for the answers “yes” or “no”, respectively.
 - Possible answers for the Keystrokes Group: press the key “1” or “2” on the phone’s keypad for the answers “yes” or “no”, respectively.

²Pro-WEAI consists of 12 indicators of women’s empowerment in agriculture: autonomy in income, self-efficacy, attitudes about domestic violence, input in productive decisions, ownership of land and other assets, access to and decisions on credit, control over use of income, work balance, visiting important locations, group membership, membership in influential groups, and respect among household members.

2. Do you have to ask permission to buy vegetables or fruits; clothes for yourself; medications or personal supplies?
- Possible answers for the Control Group: say “Yes” or “No”.
 - Possible answers for the Code words Group: say the code words “one” or “two” for the answers “yes” or “no”, respectively.
 - Possible answers for the Keystrokes Group: press the key “1” or “2” on the phone’s keypad for the answers “yes” or “no”, respectively.
3. Who usually decides if you can go to visit the house of a friend or neighbor?
- Possible answers for the Control Group: say “myself”, “my partner”, or “another member of the household”.
 - Possible answers for the Code words Group: say the code words “one”, “two”, or “three” for the answers “myself”, “my partner”, or “another member of the household”, respectively.
 - Possible answers for the Keystrokes Group: press the key “1”, “2”, or “3” on the phone’s keypad for the answers “myself”, “my partner”, or “another member of the household”, respectively.
4. Who contributes the most to the decision to buy new assets (hoe, fumigation pump)?
- Possible answers for the Control Group: say “myself”, “my partner”, or “another member of the household”.
 - Possible answers for the Code words Group: say the code words “one”, “two”, or “three” for the answers “myself”, “my partner”, or “another member of the household”, respectively.
 - Possible answers for the Keystrokes Group: press the key “1”, “2”, or “3” on the phone’s keypad for the answers “myself”, “my partner”, or “another member of the household”, respectively.

5. Who decides when to give support to parents, in-laws, siblings, etc.?

- Possible answers for the Control Group: say “myself”, “my partner”, or “another member of the household”.
- Possible answers for the Code words Group: say the code words “one”, “two”, or “three” for the answers “myself”, “my partner”, or “another member of the household”, respectively.
- Possible answers for the Keystrokes Group: press the key “1”, “2”, or “3” on the phone’s keypad for the answers “myself”, “my partner”, or “another member of the household”, respectively.

6. When making decisions regarding children’s health care, who usually makes the decision?

- Possible answers for the Control Group: say “myself”, “my partner”, or “another member of the household”.
- Possible answers for the Code words Group: say the code words “one”, “two”, or “three” for the answers “myself”, “my partner”, or “another member of the household”, respectively.
- Possible answers for the Keystrokes Group: press the key “1”, “2”, or “3” on the phone’s keypad for the answers “myself”, “my partner”, or “another member of the household”, respectively.

Interviews followed a careful protocol to ensure comparability across arms. Before answering the module with the experimental questions, interviewees were asked to answer from a private place, away from other family members and bystanders, and not to use the phone’s speaker. The call was otherwise rescheduled. In all arms there were practice questions and guidelines before proceeding with the questions on agency. The practice questions were on agricultural practices, a topic related to the rest of the survey, and allowed participants to understand how to respond in the given arm before the agency

questions began. Separately, respondents were also asked a non-sensitive question, “Is this month the harvest season?”, using the same response method as the agency questions in their assigned arm. This item serves as a placebo in the analysis, since privacy should not affect how women report an innocuous fact about the agricultural calendar.

Moreover, surveyors explained the privacy gains in each treatment arm. In the code words arm, surveyors explained that nobody around the interviewee would know her answers. In the keystroke arm, surveyors explained that nobody around the interviewee nor the enumerator would know the interviewee’s answers. To make sure that participating in the study did not result in a cost for beneficiaries, airtime or a prepaid internet data plan were provided to each participant for free.

3.2 Data collection and balance

A total of 1,183 women completed the survey experiment: 627 as part of the DIGITAGRO impact evaluation endline survey, and 556 from an additional top-up sample. The top-up sample was interviewed during the same fieldwork effort as the DIGITAGRO endline in August 2021 and completed an identical questionnaire, with an added set of demographic questions on key characteristics (e.g., household composition, education level). This information had been gathered at baseline for the DIGITAGRO sample.

Both samples were drawn from a census conducted for the DIGITAGRO project in rural San Marcos, Guatemala, covering adult women (18+) engaged in agriculture and owning a cellphone. Women with smartphones were invited to participate in the DIGITAGRO intervention, while those with only basic cellphones were excluded. The survey experiment was cross-randomized across samples and DIGITAGRO’s treatment arms.³

Table 1 shows balance of the randomization in the assignment of participants to one of three treatment arms. Of the 1,112 respondents with complete covariate information (the sample in Table 1), 34.8% were assigned to “Code words,” 32% to “Keystroke,”

³As expected, DIGITAGRO participants were of higher socioeconomic status. As shown later, the results are robust to controlling for DIGITAGRO participation and treatment assignment (Table A5).

and 33.2% to conventional surveys.⁴ The first three pairs of columns of Table 1 show summary statistics across covariates in each experimental arm. The last three columns report p-values of t-tests of differences across experimental arms. None of the differences between groups is statistically significant, and the differences are not jointly significant as indicated in the bottom of the table. Overall, there is evidence of balance across arms due to the random allocation of participants into treatment. Due to technical difficulties, the tone of keystrokes was not recorded for 88 participants in the Keystroke arm on the first agency question; a further 24 participants across all arms are missing an answer to at least one of the other agency questions. However, excluding these observations did not result in sample imbalance in the resulting estimation sample, as shown in Appendix Table A1.⁵

4 Empirical Approach

Each of the six questions outlined in Section 3.1 was transformed into a binary outcome that takes value 1 to indicate more agency.⁶ Given one of these agency outcomes y , the regression equation for person i in village j is:

$$y_{ij} = \alpha + \beta \text{Added Privacy}_{ij} + e_{ij}, \quad (1)$$

where Added Privacy is a binary variable indicating if the participant received any of

⁴Sample sizes differ by table because of the exclusion criterion applied: Table 1 keeps the 1,112 of the 1,183 respondents with complete covariate data; Appendix Table A1 and Tables 2–4 require non-missing answers to all agency questions except child health, which is mean-imputed by arm (Section 4), yielding 1,000.

⁵In addition, Appendix Table A2 conducts a regression examining indicators for the missing outcomes. Keystrokes are more likely to be missing due to the technical errors. In contrast, codewords are less likely to be missing for the supporting relatives question, though this is only marginally significant at the 10% level. Codewords has an insignificant coefficient for all other outcomes. Reassuringly, the surviving sample remains balanced on observed characteristics (Appendix Table A1), and the results are unchanged across robustness checks as shown in Section 5.2.

⁶The six outcomes are: (i) allowed to go out alone = 1; (ii) does not have to ask permission to buy goods = 1; (iii) respondent decides whether she can visit friends = 1; (iv) respondent contributes the most to the decision to buy new assets = 1; (v) respondent decides when to give support to relatives = 1; and (vi) respondent usually makes decision regarding children’s health = 1.

the two privacy treatments. Thus, β is the parameter of interest. A positive (negative) parameter would imply that added privacy allowed women to report higher (lower) agency relative to the control group (i.e., responding verbally). The analysis was also conducted separating the effects by the two treatment arms. The random assignment of treatment arms allows us to interpret the relationship as causal. Standard errors are clustered at the village level.⁷

In order to deal with multiple hypothesis testing, since there are six measures of agency, the analysis is adjusted in two ways. First, an agency index is created using the inverse-covariance-weighted method from [Anderson \(2008\)](#).⁸ The index is standardized to have mean zero and standard deviation one in the control group, so that coefficients are expressed in control-group standard deviations; more positive values indicate more agency. It is constructed with the Stata `swindex` command from [Schwab et al. \(2020\)](#). Second, simultaneous inference is corrected based on the concept of a false discovery rate (FDR) in [Benjamini and Hochberg \(1995\)](#). Intuitively, the FDR is the expected share of rejected hypotheses that are false discoveries. An FDR-adjusted q-value of 0.01 implies that one percent of the tests declared significant are expected to be false positives, whereas an unadjusted p-value of 0.01 bounds at one percent the probability of rejecting a single true null hypothesis, without regard to how many hypotheses are tested.

To check the sensitivity of Equation 1, some specifications add controls for a set of pre-determined characteristics. These are: respondent age, marital status, log of land size, education level, partner education level, if respondent is the household head, land ownership status, number of children, number of family members, number of adult men and women in household, and participation in the DIGITAGRO program.

⁷Though this paper’s survey experiment was randomly assigned at the individual level, the DIGITAGRO program was randomly assigned at the village level. As a result, standard errors are conservatively clustered. Appendix Table A3 shows results with robust standard errors; these are virtually identical to the main results.

⁸Not all women in the sample have children, which creates missing data for the outcome on decision-making regarding child health. Missing observations are imputed with the mean by treatment arm for this outcome to avoid losing observations when creating the index.

4.1 Heterogeneity analysis

To conduct heterogeneity analysis for different covariates, the following equation was estimated:

$$y_{ij} = \alpha + \beta \text{Added Privacy}_{ij} + \gamma C_{ij} + \delta C_{ij} \times \text{Added Privacy}_{ij} + e_{ij}, \quad (2)$$

Where C_{ij} is the covariate of interest, a dummy variable (continuous covariates are split at the median). This framework allows us to explore differing effects for different groups within the sample. With this specification, the conditional average treatment effect for the group $C_{ij} = 1$ is given by $\beta + \delta$.

5 Results

5.1 Main findings and heterogeneous effects

The results of estimating Equation (1) are shown in Table 2. In column (1), focusing on the index, the results show that women exposed to higher privacy tend to report less agency. Added privacy decreases the agency index by 0.188 standard deviations. This result seems mainly driven by the question asking if respondents require permission to buy goods and the question on decisions around child health, although the negative coefficients are observed across all but two indicators.

Results of allowing the privacy effect to vary by arm are presented in Table 3. Again, column (1) focuses on the index and continues to find that adding privacy leads to less reporting of agency. The table also shows that the null hypothesis that both treatment arms (Code words and Keystrokes) have the same effect cannot be rejected. The p-value for this null hypothesis is never below 0.266. The inability to reject equality of impacts across the two treatment arms implies that code words alone might be an effective tool to elicit more truthful responses from interviewees, without having to resort to keystrokes

that are harder to capture for the researcher and to implement in the field.⁹

The average effect may mask differences across respondents. To examine this, a set of covariates are interacted one at a time as in Equation 2, where the outcome is the agency index. We choose these covariates based on availability in our data and previous research linking them to women’s agency and bargaining position within the household: dependency ratio, number of children, and marital status (Hanmer and Klugman, 2016); household headship (Goldman and Little, 2015); land tenancy (Mishra and Sam, 2016); and education, age, and ethnicity, the latter proxied by whether the respondent asked to be interviewed in Mam (O’Hara and Clement, 2018).¹⁰ Results are shown in Figure 1. Respondents who are the head of household, completed primary education, or are single report less agency when offered more privacy. Table A4 in the appendix formally tests whether the privacy effect differs between each pair of groups. The interaction is statistically significant for household headship, education, and marital status, but not for land ownership—even though the effect on its own is significant among landowners. These results are consistent with the effects being driven by respondents with higher social standing.¹¹

5.2 Robustness checks

This paper considers several robustness tests to the main analysis described in Section 5.1. As a preliminary check, Appendix Tables A5, A6, and A7 show the results of adding controls, adding fixed effects by municipality (but without other controls), and adding fixed effects and other controls, respectively. Appendix Table A8 includes fixed effects for the survey interviewers. The main finding persists: providing more privacy leads to lower measures of agency.

⁹Since the two arms cannot be distinguished statistically, the remainder of the analysis uses the pooled indicator.

¹⁰The dependency ratio counts children and people older than 65 relative to members aged 15–64. Continuous covariates are split at the median.

¹¹These characteristics are not randomly assigned and are correlated with one another, so each interaction is best read as a descriptive comparison across groups.

As a further check, different ways of constructing the aggregated agency index outcome were used. First, all outcome variables were averaged in two ways: imputing the average of each group to the missing data of the question regarding child health and not imputing. Second, a standardized version of the index following [Kling et al. \(2007\)](#) was considered, where each outcome was standardized into a Z-score by subtracting the mean and dividing by the standard deviation from the control group before averaging across variables. Third, an index averaging across the first five outcomes only was constructed. Finally, the first principal component of the same questions used to create the main index was used. The Appendix shows the results of this exercise without controls (Appendix Table [A9](#)), with controls (Appendix Table [A10](#)), with municipality fixed effects (Appendix Table [A11](#)), and with municipality fixed effects and controls (Appendix Table [A12](#)). The results are not affected by changes in the specification. Overall, these unchanged results provide additional evidence about the quality of the randomization process and the choice of the procedure to aggregate responses.

Finally, we show that survey miscomprehension is not driving a spurious correlation between the treatment arms and reporting. Treated respondents may report less agency because they did not understand the code-word or keystroke response mechanisms. To test for this, the survey asked a placebo, non-sensitive question, “Is this month the harvest season?”, using the same response methodology as the agency questions. As displayed in Table [4](#), there are no differences in reporting across experimental arms, suggesting that respondents understood the keystroke and code-word response mechanisms and that the differences in reporting are due to privacy.

5.3 Discussion

Sensitive questions can produce misreporting under two broad mechanisms: social desirability (when the question elicits answers that are socially (un)desirable) and threat of disclosure (fear of third-party consequences) ([Tourangeau et al., 2000](#); [Tourangeau and Yan, 2007](#)). This subsection discusses the potential drivers behind the main finding that

added privacy results in less reported agency.

Under social desirability, respondents adjust their answers toward what they believe is socially acceptable (Krumpal, 2013; Tourangeau et al., 2000; Tourangeau and Yan, 2007). Under this interpretation, women could over-report having agency when responding directly if women’s agency is socially desirable, and answer more truthfully when offered increased privacy (Álvarez-Aragón and Champeaux, 2026).

The evidence of this paper seems consistent with social desirability bias. A set of semi-structured interviews was carried out over the phone with a random subsample of participants a few weeks after the survey experiment. Relying on broad prompts to spark discussion and taking time to build trust, semi-structured interviews are seen as less prone to misreporting than standard structured surveys (Jayachandran et al., 2023). The answers were subsequently coded to quantify women’s opinions on different topics, a strategy that recent work suggests can recover meaningful signal about women’s agency from interview responses (Clingain et al., 2026).¹² Out of 104 responses, 76 percent of interviewees think that their community would approve and support women having agency and their own voice, indicating that agency is a socially desirable trait in this setting (Blair et al., 2020).¹³ Moreover, the heterogeneity results show stronger effects of privacy for women with higher social standing, which would be consistent with this group being under the most pressure not to deviate from what is normatively valued (Bernstein et al., 2001; Silver et al., 1986).

The results of the study are less supportive of threat of disclosure. For example, a commonly cited safety and reporting concern in the domestic violence literature is fear of retaliation, as respondents report on potentially criminal conduct and the risk is immediate (Ellsberg et al., 2001; Sikweyiya and Jewkes, 2012; Vranda et al., 2018). In our setting, added privacy could allow women to report low levels of agency without fear of reprisal, for example, from a controlling partner. However, this would be at

¹²The interviews were conducted between September and October of 2021. All of them were audio recorded and transcribed.

¹³This finding would be inconsistent with agency being a taboo topic and thus generating differential misreporting across arms (Tourangeau and Yan, 2007).

odds with the earlier result from the qualitative surveys that agency, and not lack of it, is a desirable trait. In addition, reprisal risk and threats of disclosure are usually higher among vulnerable women (Mabena et al., 2025; Baruwa et al., 2025; Acharya, 2025), which contradicts the heterogeneity results, where added privacy benefits those with higher social standing.

Other explanations identified in the literature also seem inconsistent with the evidence. Survey miscomprehension—whereby treated respondents spuriously report less agency because they did not understand the code-word or keystroke response mechanisms—is inconsistent with the placebo results in Table 4. Surveyor characteristics could also lead to differential reporting if correlated with the treatment arms (Di Maio and Fiala, 2020; Huddy et al., 1997), but our results remain unchanged after controlling for surveyor effects (Appendix Table A8). Construct validity is another key topic that could lead to differential responses (Peterman et al., 2021), but by design in our experiment the exact same questions were asked in the three arms; only the response mode was changed.

Although the leading explanation seems to be social desirability bias, the experiment was designed to vary privacy rather than to separate the mechanisms behind it. Designs built for that purpose are left for future research.

6 Conclusion

This paper shows that self-reported markers of women’s agency respond to the privacy of the response, evidence of measurement error in standard survey conditions. Using a survey experiment applied to a sample of women in rural Western Guatemala, the paper shows that enhancing privacy in surveys leads to *lower* reporting of women’s agency. Methodologically, the study introduces a simple and low-cost approach to provide higher privacy that allows sensitive issues to be analyzed at the individual level, similar to Audio Computer-Assisted Self-Interview (ACASI) survey techniques, and unlike indirect methods such as list experiments, which recover prevalence only at the population or

group level. The results show that phone interviews allowing participants to use code words or keystrokes can trigger different responses, and the pattern of those differences points to less overreporting.

The study also finds that overreporting of agency is most likely driven by social desirability bias. This interpretation is supported by heterogeneous effects as well as by qualitative evidence gathered from semi-structured interviews with participants in the experiment. Other explanations, such as threat of disclosure, survey miscomprehension, surveyor effects, and construct validity, find little support in the evidence.

These findings have implications for program and policy design. First, standard surveys may inflate agency levels, which could lead to imprecise estimates of the true status of women. Second, programs that use self-reports to target support may be overlooking key groups that incorrectly appear more empowered. Third, measurement error might lead to wrong policy conclusions as it can bias the estimates of impact evaluations ([Agüero et al., 2026](#)). Using simple privacy methods, like those proposed in this paper, can help correct these distortions.

Finally, the paper presents questions for future research. First, replication in other contexts is needed. The finding that women overreport their agency in our setting could imply that agency levels are lower than expected in the larger population. Second, further research could explore how additional agency and empowerment questions vary with privacy. Third, more evidence is needed on the level of privacy that affects reporting. The study finds no empirical difference between the keystroke and codeword methods, despite the expectation that keystrokes might offer greater privacy. Similarly, the paper finds effects on yes/no questions. While the act of verbalizing a response may engage self-affirmation dynamics consistent with social desirability bias, further research is needed to disentangle these effects.

References

- Acharya, S. R. (2025), ‘Exposure to gender-based violence and the risk of hypertension, obesity, and anemia: a multilevel analysis of the 2017 Tajikistan demographic and health survey’, *BMC Public Health* **25**(1), 2165.
- Agüero, J. M., Aldana, Ú., Field, E., Frisancho, V., Rodriguez Hurtado, I. and Romero, J. (2026), ‘Measurement error in self-reported data: Experimental evidence for intimate partner violence’. Working Paper.
- Agüero, J. M. and Frisancho, V. (2022), ‘Measuring violence against women with experimental methods’, *Economic Development and Cultural Change* **70**(4), 1565–1590.
- Allendorf, K. (2007), ‘Do women’s land rights promote empowerment and child health in nepal?’, *World development* **35**(11), 1975–1988.
- Álvarez-Aragón, P. and Champeaux, H. (2026), ‘Measuring norms and enumerator effects: Survey method matters’, *The World Bank Economic Review* . Advance access, lhaf043.
- Ambler, K., Doss, C., Kieran, C. and Passarelli, S. (2021), ‘He says, she says: Spousal disagreement in survey measures of bargaining power’, *Economic Development and Cultural Change* **69**(2), 765–788.
- Anderson, M. L. (2008), ‘Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects’, *Journal of the American statistical Association* **103**(484), 1481–1495.
- Annan, J., Donald, A., Goldstein, M., Martinez, P. G. and Koolwal, G. (2021), ‘Taking power: women’s empowerment and household well-being in sub-saharan africa’, *World Development* **140**, 105292.
- Baruwa, O. J., Biney, E., Asuma, B. S. and Amoateng, A. Y. (2025), ‘Factors associated

- with intimate partner violence among ever-partnered women in two Southern African countries: a multilevel analysis', *BMC Women's Health* **25**(1), 304.
- Benjamini, Y. and Hochberg, Y. (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society: Series B (Methodological)* **57**(1), 289–300.
- Bernstein, R., Chadha, A. and Montjoy, R. (2001), 'Overreporting voting: Why it happens and why it matters', *Public Opinion Quarterly* **65**(1), 22–44.
- Blair, G., Coppock, A. and Moor, M. (2020), 'When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments', *American Political Science Review* **114**(4), 1297–1315.
- Chuang, E., Dupas, P., Huillery, E. and Seban, J. (2021), 'Sex, lies, and measurement: Consistency tests for indirect response survey methods', *Journal of Development Economics* **148**, 102582.
- Clingain, C., Donald, A., Hernández-de Benito, M., Nasirumbi Sanya, L. and Rizzi, M. (2026), Can text measure agency? comparing word counts, semantic methods and ai models against expert-coded interviews. Unpublished manuscript, World Bank.
- Corkrey, R. and Parkinson, L. (2002), 'A comparison of four computer-based telephone interviewing methods: Getting answers to sensitive questions', *Behavior Research Methods, Instruments, & Computers* **34**(3), 354–363.
- Costa, J. C., Saad, G. E., Hellwig, F., Maia, M. F. S. and Barros, A. J. (2023), 'Measures of women's empowerment based on individual-level data: a literature review with a focus on the methodological approaches', *Frontiers in Sociology* **8**, 1231790.
- Cullen, C. (2023), 'Method matters: The underreporting of intimate partner violence', *The World Bank Economic Review* **37**(1), 49–73.

- Deininger, K., Jin, S., Nagarajan, H. K. and Xia, F. (2019), ‘Inheritance law reform, empowerment, and human capital accumulation: Second-generation effects from india’, *The Journal of Development Studies* **55**(12), 2549–2571.
- Di Maio, M. and Fiala, N. (2020), ‘Be wary of those who ask: A randomized experiment on the size and determinants of the enumerator effect’, *The World Bank Economic Review* **34**(3), 654–669.
- Dillon, B. and Voena, A. (2018), ‘Widows’ land rights and agricultural investment’, *Journal of Development Economics* **135**, 449–460.
- Donald, A., Doss, C., Goldstein, M. and Gupta, S. (2024), ‘Sharing responsibility through joint decision-making and implications for intimate-partner violence: evidence from 12 sub-saharan african countries’, *Review of Economics of the Household* **22**(1), 35–66.
- Donald, A., Koolwal, G., Annan, J., Falb, K. and Goldstein, M. (2020), ‘Measuring women’s agency’, *Feminist Economics* **26**(3), 200–226.
- Edmonds, E., Feigenberg, B. and Leight, J. (2023), ‘Advancing the agency of adolescent girls’, *Review of Economics and Statistics* **105**(4), 852–866.
- Ellsberg, M., Heise, L., Pena, R., Agurto, S. and Winkvist, A. (2001), ‘Researching domestic violence against women: methodological and ethical considerations’, *Studies in family planning* **32**(1), 1–16.
- FAO (2016), Guatemala: Voz a las mujeres sobre seguridad alimentaria y nutrición, Technical report, Food and Agriculture Organization of the United Nations, Guatemala. <https://www.fao.org/in-action/gender-equality-guatemala/es/>.
- Gilligan, D. O., Hidrobo, M., Leight, J. and Tabet, H. (2025), ‘Using a list experiment to measure intimate partner violence: cautionary evidence from Ethiopia’, *Applied Economics Letters* **32**(11), 1594–1600.

- Gnambs, T. and Kaspar, K. (2015), ‘Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis’, *Behavior research methods* **47**, 1237–1259.
- Goldman, M. J. and Little, J. S. (2015), ‘Innovative grassroots ngos and the complex processes of women’s empowerment: An empirical investigation from northern tanzania’, *World Development* **66**, 762–777.
- Hanmer, L. and Klugman, J. (2016), ‘Exploring women’s agency and empowerment in developing countries: Where do we stand?’, *Feminist economics* **22**(1), 237–263.
- Höglinger, M., Jann, B. and Diekmann, A. (2016), Sensitive questions in online surveys: An experimental evaluation of different implementations of the randomized response technique and the crosswise model, *in* ‘Survey Research Methods’, Vol. 10, pp. 171–187.
- Huddy, L., Billig, J., Bracciodieta, J., Hoeffler, L., Moynihan, P. J. and Pugliani, P. (1997), ‘The effect of interviewer gender on the survey response’, *Political Behavior* **19**(3), 197–220.
- INE (2015), República de Guatemala: Encuesta nacional de condiciones de vida 2014. ENCOVI 2014, Technical report, Instituto Nacional de Estadística, Guatemala.
- INE (2019), XII Censo Nacional de Población y Vivienda – Resultados Censo 2018, Technical report, Instituto Nacional de Estadística, Guatemala.
- Iqbal, N., Jalal, A., Mahmud, M. and Vyborny, K. (2025), Collecting accurate data on intimate partner violence: Learnings from pakistan, Technical Report Policy Research Working Paper 11077, World Bank.
- Jayachandran, S., Biradavolu, M. and Cooper, J. (2023), ‘Using machine learning and qualitative interviews to design a five-question survey module for women’s agency’, *World Development* **161**, 106076.

- Joseph, G., Javaid, S. U., Andres, L. A., Chellaraj, G., Solotaroff, J. and Rajan, S. I. (2017), ‘Underreporting of gender-based violence in kerala, india: An application of the list randomization method’, *World Bank Policy Research Working Paper* (8044).
- Kabeer, N. (1999), ‘Resources, agency, achievements: Reflections on the measurement of women’s empowerment’, *Development and Change* **30**(3), 435–464.
- Kadam, A., McCullough, E. B., McGavock, T. J. and Magnan, N. (2025), ‘Who is asking and how? Effects of survey mode and enumerator gender on measuring women’s life experience’, *World Development* **195**, 107078.
- Karlan, D. and Zinman, J. (2012), ‘List randomization for sensitive behavior: An application for measuring use of loan proceeds’, *Journal of Development Economics* **98**(1), 71–75.
- Kelly, C. A., Soler-Hampejsek, E., Mensch, B. S. and Hewett, P. C. (2013), ‘Social desirability bias in sexual behavior reporting: evidence from an interview mode experiment in rural malawi’, *International perspectives on sexual and reproductive health* **39**(1), 14–21.
- Kling, J. R., Liebman, J. B. and Katz, L. F. (2007), ‘Experimental analysis of neighborhood effects’, *Econometrica* **75**(1), 83–119.
- Krumpal, I. (2013), ‘Determinants of social desirability bias in sensitive surveys: A literature review’, *Quality & Quantity* **47**(4), 2025–2047.
- Langhaug, L. F., Cheung, Y. B., Pascoe, S. J., Chirawu, P., Woelk, G., Hayes, R. J. and Cowan, F. M. (2011), ‘How you ask really matters: randomised comparison of four sexual behaviour questionnaire delivery modes in zimbabwean youth’, *Sexually transmitted infections* **87**(2), 165–173.
- Langhaug, L. F., Sherr, L. and Cowan, F. M. (2010), ‘How to improve the validity

- of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries’, *Tropical medicine & international health* **15**(3), 362–381.
- Lindstrom, D. P., Hattori, M. K., Belachew, T. and Tessema, F. (2012), ‘Lifting the curtain on the conditions of sexual initiation among youth in ethiopia’, *Journal of adolescent health* **50**(6), 614–620.
- Lopez, R. A., Perego, V. M. E. and Romero, J. (2025), ‘Investing in digital technology to increase market access for smallholder farmers: Experimental evidence from rural guatemala’, *Working Paper* .
- Mabena, K., Tsabedze, W. and Mazibuko, X. (2025), ‘Socio-economic determinants of intimate-partner physical violence among women in South Africa’, *Frontiers in Psychology* **16**, 1499263.
- Makkai, T. and McAllister, I. (1992), ‘Measuring social indicators in opinion surveys: A method to improve accuracy on sensitive questions’, *Social Indicators Research* **27**, 169–186.
- Malapit, H., Quisumbing, A., Meinzen-Dick, R., Seymour, G., Martinez, E. M., Heckert, J., Rubin, D., Vaz, A., Yount, K. M., Phase, G. A. A. P. et al. (2019), ‘Development of the project-level women’s empowerment in agriculture index (pro-weai)’, *World development* **122**, 675–692.
- Mishra, K. and Sam, A. G. (2016), ‘Does women’s land ownership promote their empowerment? empirical evidence from nepal’, *World Development* **78**, 360–371.
- Moreno, A. (2020), Cadenas agroalimentarias modernas y resilientes - análisis de brechas de género, Unpublished, The World Bank.
- O’Hara, C. and Clement, F. (2018), ‘Power as agency: A critical reflection on the measurement of women’s empowerment in the development sector’, *World Development* **106**, 111–123.

- Park, D. S., Aggarwal, S., Jeong, D., Kumar, N., Robinson, J. and Spearot, A. (2025), ‘Private but misunderstood? Evidence on measuring intimate partner violence via self-interviewing in rural Liberia and Malawi’, *The World Bank Economic Review* **39**(4), 907–923.
- Perego, V., Romero, J., Freeman, K., Lopez, A., Ortiz, G., Salas, H., Ramirez, R., Locatelli, A., Orihuela, D. and de Ferrari, C. (2022), Digitagro: Investing in digital technology to increase market access for women agribusinesses in Guatemala, Report ACS34060, The World Bank, Washington, DC.
- Peterman, A., Dione, M., Le Port, A., Briaux, J., Lamesse, F. and Hidrobo, M. (2025), ‘Disclosure of violence against women and girls in Senegal’, *The World Bank Economic Review* **39**(3), 614–631.
- Peterman, A., Schwab, B., Roy, S., Hidrobo, M. and Gilligan, D. O. (2021), ‘Measuring women’s decisionmaking: Indicator choice and survey design experiments from cash and food transfer evaluations in Ecuador, Uganda and Yemen’, *World Development* **141**, 105387.
- Schwab, B., Janzen, S., Magnan, N. P. and Thompson, W. M. (2020), ‘Constructing a summary index using the standardized inverse-covariance weighted average of indicators’, *The Stata Journal* **20**(4), 952–964.
- Sikweyiya, Y. and Jewkes, R. (2012), ‘Perceptions and experiences of research participants on gender-based violence community based survey: Implications for ethical guidelines’, *PloS one* **7**(4), e35495.
- Silver, B. D., Anderson, B. A. and Abramson, P. R. (1986), ‘Who overreports voting?’, *American Political Science Review* **80**(2), 613–624.
- Tanton, C., Bhatia, A., Pearlman, J. and Devries, K. (2023), ‘Increasing disclosure of school-related gender-based violence: lessons from a systematic review of data collection methods and existing survey research’, *BMC public health* **23**(1), 1012.

- Tourangeau, R., Rips, L. J. and Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge University Press, Cambridge.
- Tourangeau, R. and Yan, T. (2007), ‘Sensitive questions in surveys’, *Psychological Bulletin* **133**(5), 859–883.
- UNDP (2020), Gender Inequality Index, Technical report, United Nations Development Programme.
- United Nations (2024), ‘Goal 5: Achieve gender equality and empower all women and girls’. Accessed: 2024-09-10.
URL: <https://sdgs.un.org/goals/goal5>
- van der Meulen Rodgers, Y. and Kassens, A. L. (2018), ‘Women’s asset ownership and children’s nutritional status: evidence from papua new guinea’, *Social Science & Medicine* **204**, 100–107.
- Vranda, M. N., Kumar, C. N., Muralidhar, D., Janardhana, N. and Sivakumar, P. T. (2018), ‘Barriers to disclosure of intimate partner violence among female patients availing services at tertiary care psychiatric hospitals: A qualitative study’, *Journal of Neurosciences in Rural Practice* **9**(3), 326–330.

Tables

Table 1: Balance Table - All Observations with Complete Covariate Data

Variable	(1) Control Mean/(SE)	(2) Code words Mean/(SE)	(3) Keystroke Mean/(SE)	(1)-(2) P-value	(1)-(3) Pairwise t-test P-value	(2)-(3) P-value
<i>Interviewee Characteristics</i>						
Interviewee Age	39.694 (0.637)	40.421 (0.677)	39.778 (0.646)	0.417	0.926	0.456
Interviewee finished primary education	0.469 (0.028)	0.473 (0.024)	0.480 (0.028)	0.912	0.752	0.827
Interviewee is single	0.274 (0.023)	0.225 (0.022)	0.270 (0.025)	0.116	0.904	0.167
Interviewee asked to interview in Mam (Mayan dialect)	0.225 (0.038)	0.207 (0.033)	0.222 (0.032)	0.567	0.930	0.598
Interviewee is HH head	0.163 (0.019)	0.134 (0.017)	0.146 (0.020)	0.267	0.576	0.616
Interviewee in DIGITAGRO Evaluation	0.566 (0.029)	0.537 (0.028)	0.556 (0.031)	0.478	0.813	0.650
<i>Household Characteristics</i>						
Land Ownership	0.488 (0.028)	0.468 (0.028)	0.520 (0.028)	0.576	0.411	0.185
Log Land Size (in cuerdas)	1.590 (0.053)	1.599 (0.051)	1.600 (0.056)	0.903	0.893	0.987
Household head finished primary education	0.455 (0.028)	0.450 (0.025)	0.441 (0.030)	0.875	0.694	0.824
Number of household members	6.230 (0.145)	6.382 (0.132)	6.298 (0.139)	0.405	0.729	0.634
F-test of joint significance (P-value)				0.796	0.998	0.934
Number of observations	369	387	356	756	725	743
Number of clusters	198	210	203	296	290	295

Notes: The value displayed for t-tests and F-tests are p-values. Standard errors are clustered at the village level.

* p< 0.10, ** p< 0.05, *** p< 0.01.

Table 2: Experimental Results Pooling Privacy Options

	Index	Out Alone	Small Purchases	Visit Friends	Asset Purchases	Support Relatives	Child Health
Added Privacy	-0.188*** (0.066) [0.019]	-0.045 (0.031) [0.275]	-0.120*** (0.029) [0.000]	-0.043 (0.031) [0.275]	0.015 (0.029) [0.741]	0.009 (0.035) [0.805]	-0.076** (0.038) [0.123]
Control Mean	0.000	0.620	0.690	0.681	0.227	0.404	0.574
Obs.	1000	1000	1000	1000	1000	1000	807
R-squared	0.009	0.002	0.014	0.002	0.000	0.000	0.005
Clusters	338	338	338	338	338	338	308

Notes: Added Privacy displays the pooled coefficient combining the code words and keystroke treatments.

The dependent variables are related to the next questions: Out alone: “Are you allowed to go alone to meet your friends for any reason (chatting, lunch, etc.)?”; Small Purchases: “Do you have to ask permission to buy vegetables or fruit; clothes for yourself; medicines or personal supplies?”; Visit Friends: “Who usually decides whether you can visit a friend’s/neighbour’s house?”; Asset Purchases: “Who contributes most to the decision to buy new assets (e.g. hoe or spray pump)?”; Support relatives: “Who decides whether to support parents, in-laws, siblings, etc.”; Child health: “When making decisions regarding children’s health care, who usually makes the decision?”

Index is a summary index of all agency questions, using the method in ([Anderson, 2008](#)), standardized using the control-group mean and standard deviation.

Clustered standard errors at the village level in parentheses. FDR-q values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Experimental Results Separating Privacy Options

	Index	Out Alone	Small Purchases	Visit Friends	Asset Purchases	Support Relatives	Child Health
Codewords	-0.212*** (0.072) [0.029]	-0.041 (0.036) [0.453]	-0.134*** (0.033) [0.001]	-0.045 (0.034) [0.382]	0.009 (0.030) [0.800]	-0.010 (0.038) [0.800]	-0.081* (0.042) [0.224]
Keystroke	-0.152* (0.084) [0.227]	-0.050 (0.037) [0.382]	-0.099*** (0.038) [0.048]	-0.040 (0.040) [0.509]	0.025 (0.039) [0.668]	0.037 (0.044) [0.598]	-0.069 (0.047) [0.370]
Control Mean	0.000	0.620	0.690	0.681	0.227	0.404	0.574
Obs.	1000	1000	1000	1000	1000	1000	807
R-squared	0.009	0.002	0.015	0.002	0.001	0.001	0.005
Pval Diff. between treatment	0.458	0.827	0.390	0.889	0.665	0.266	0.802
Clusters	338	338	338	338	338	338	308

Notes: Each column shows the coefficient of regressing each answer on privacy separately by treatment arm.

The dependent variables are related to the next questions: Out alone: “Are you allowed to go alone to meet your friends for any reason (chatting, lunch, etc.)?”; Small Purchases: “Do you have to ask permission to buy vegetables or fruit; clothes for yourself; medicines or personal supplies?”; Visit Friends: “Who usually decides whether you can visit a friend’s/neighbour’s house?”; Asset Purchases: “Who contributes most to the decision to buy new assets (e.g. hoe or spray pump)?”; Support relatives: “Who decides whether to support parents, in-laws, siblings, etc.”; Child health: “When making decisions regarding children’s health care, who usually makes the decision?”

Index is a summary index of all agency questions, using the method in ([Anderson, 2008](#)), standardized using the control-group mean and standard deviation.

Clustered standard errors at the village level in parentheses. FDR-q values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

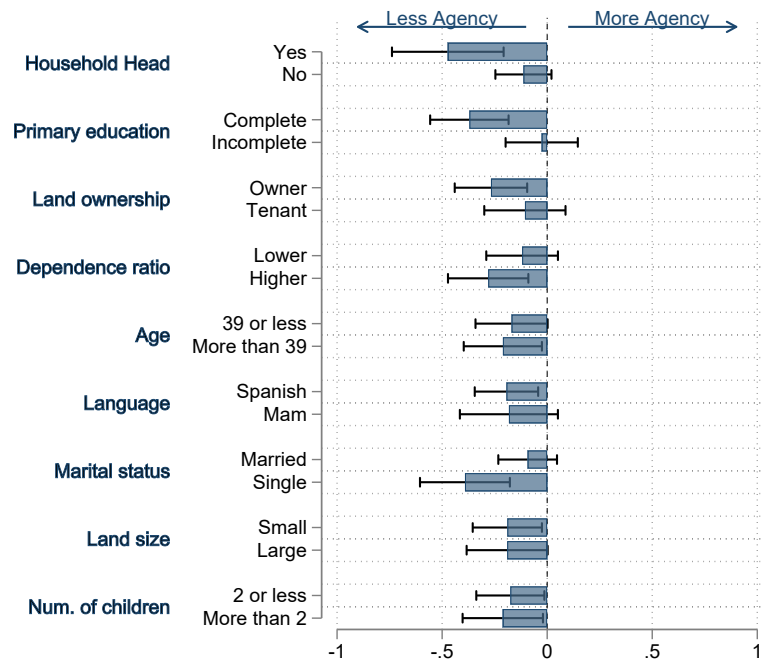
Table 4: Placebo Question

	Pooled	By Arm
Added Privacy	-0.016 (0.035) [0.643]	
Codewords		-0.011 (0.040) [0.791]
Keystroke		-0.025 (0.040) [0.791]
Control Mean	0.587	0.587
Obs.	1000	1000
R-squared	0.000	0.000
Clusters	338	338

Notes: Added Privacy indicates the pooled treatment. Codewords and Keystroke separate the effects for the two treatment arms.

The placebo question is “Is this month harvest season?”

Figure 1: Effect of adding privacy on the Agency Index across covariate groups



Notes: Each row reports the effect of Added Privacy on the Agency Index for the two groups defined by the covariate; continuous covariates are split at the median. 95% confidence intervals are displayed.

Appendix:

Table A1: Balance Table - Estimation Sample

Variable	(1) Control Mean/(SE)	(2) Code words Mean/(SE)	(3) Keystroke Mean/(SE)	(1)-(2) Pairwise t-test P-value	(1)-(3) P-value	(2)-(3) P-value
<i>Interviewee Characteristics</i>						
Interviewee Age	39.465 (0.646)	40.436 (0.681)	39.055 (0.750)	0.279	0.679	0.131
Interviewee finished primary education	0.474 (0.028)	0.473 (0.024)	0.496 (0.032)	0.979	0.569	0.531
Interviewee is single	0.274 (0.023)	0.223 (0.022)	0.280 (0.029)	0.103	0.891	0.111
Interviewee asked to interview in Mam (Mayan dialect)	0.224 (0.038)	0.208 (0.033)	0.197 (0.031)	0.608	0.454	0.706
Interviewee is HH head	0.161 (0.019)	0.132 (0.017)	0.138 (0.023)	0.266	0.471	0.839
Interviewee in DIGITAGRO Evaluation	0.571 (0.029)	0.535 (0.028)	0.594 (0.036)	0.384	0.613	0.192
<i>Household Characteristics</i>						
Land Ownership	0.493 (0.028)	0.465 (0.028)	0.520 (0.032)	0.435	0.529	0.187
Log Land Size (in cuerdas)	1.595 (0.054)	1.604 (0.051)	1.575 (0.066)	0.889	0.797	0.700
Household head finished primary education	0.454 (0.028)	0.449 (0.025)	0.476 (0.032)	0.892	0.568	0.505
Number of household members	6.241 (0.148)	6.390 (0.133)	6.201 (0.171)	0.421	0.852	0.345
F-test of joint significance (P-value)				0.710	0.993	0.610
Number of observations	361	385	254	746	615	639
Number of clusters	193	209	163	292	268	271

Notes: The value displayed for t-tests and F-tests are p-values. Standard errors are clustered at the village level.
* p< 0.10, ** p< 0.05, *** p< 0.01.

Table A2: Missing Outcomes

	Out Alone	Small Purchases	Visit Friends	Asset Purchases	Support Relatives	Child Health
Codewords	0.000 (.) [.]	0.000 (.) [.]	-0.003 (0.006) [0.620]	-0.005 (0.004) [0.200]	-0.013* (0.007) [0.117]	-0.031 (0.030) [0.338]
Keystroke	0.241*** (0.023) [0.000]	0.249*** (0.024) [0.000]	0.236*** (0.023) [0.000]	0.223*** (0.022) [0.000]	0.207*** (0.024) [0.000]	0.205*** (0.036) [0.000]
Control Mean	0.000	0.000	0.008	0.005	0.018	0.217
Obs.	1183	1183	1183	1183	1183	1183
R-squared	0.177	0.183	0.162	0.160	0.136	0.054
Pval Diff. between treatment	0.000	0.000	0.000	0.000	0.000	0.000
Clusters	361	361	361	361	361	361

Notes: Each column shows the coefficient of regressing each answer on privacy separately by treatment arm.

Each outcome is a binary variable indicating if the survey response is missing.

No respondent in the Control or Code words arms has a missing answer to Out Alone or Small Purchases, so the Codewords coefficient in those columns is zero by construction; these two cells are excluded from the FDR adjustment.

Clustered standard errors at the village level in parentheses. FDR-q values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Experimental Results Pooling Privacy Options, with Robust Standard Errors

	Index	Out Alone	Small Purchases	Visit Friends	Asset Purchases	Support Relatives	Child Health
Added Privacy	-0.188*** (0.064) [0.014]	-0.045 (0.032) [0.267]	-0.120*** (0.031) [0.001]	-0.043 (0.031) [0.267]	0.015 (0.028) [0.711]	0.009 (0.032) [0.788]	-0.076** (0.036) [0.098]
Control Mean	0.000	0.620	0.690	0.681	0.227	0.404	0.574
Obs.	1000	1000	1000	1000	1000	1000	807
R-squared	0.009	0.002	0.014	0.002	0.000	0.000	0.005

Notes: Each column shows the coefficient of regressing each answer on privacy separately by treatment arm.

The dependent variables are related to the next questions: Out alone: “Are you allowed to go alone to meet your friends for any reason (chatting, lunch, etc.)?”; Small Purchases: “Do you have to ask permission to buy vegetables or fruit; clothes for yourself; medicines or personal supplies?”; Visit Friends: “Who usually decides whether you can visit a friend’s/neighbour’s house?”; Asset Purchases: “Who contributes most to the decision to buy new assets (e.g. hoe or spray pump)?”; Support relatives: “Who decides whether to support parents, in-laws, siblings, etc.”; Child health: “When making decisions regarding children’s health care, who usually makes the decision?”

Index is a summary index of all agency questions, using the method in ([Anderson, 2008](#)), standardized using the control-group mean and standard deviation.

Robust Standard Errors in parentheses. FDR-q values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Formal Test of Differential Effects Across Covariate Subgroups

	$C = 1$ ($\beta + \delta$)	$C = 0$ (β)	Difference (δ)
Household head ($C = 1$)	-0.473*** [0.001]	-0.113* [0.097]	-0.360** [0.016]
Completed primary education ($C = 1$)	-0.370*** [0.000]	-0.026 [0.766]	-0.344** [0.010]
Land owner ($C = 1$)	-0.268*** [0.002]	-0.106 [0.281]	-0.162 [0.219]
Dependence ratio above median ($C = 1$)	-0.281*** [0.004]	-0.119 [0.169]	-0.162 [0.205]
Age above 39 ($C = 1$)	-0.211** [0.026]	-0.169* [0.053]	-0.042 [0.738]
Interview in Mam ($C = 1$)	-0.182 [0.125]	-0.194** [0.012]	0.012 [0.935]
Single ($C = 1$)	-0.391*** [0.000]	-0.093 [0.192]	-0.299** [0.019]
Large land size ($C = 1$)	-0.190* [0.053]	-0.190** [0.024]	-0.001 [0.995]
More than 2 children ($C = 1$)	-0.211** [0.030]	-0.175** [0.034]	-0.036 [0.765]
Obs.	1000	1000	1000
Clusters	338	338	338

Notes: Coefficients from Equation (2), estimated separately for each covariate C . Each row is labeled with the category coded $C = 1$; $C = 0$ is the complementary category (not household head, incomplete primary, tenant, dependence ratio at or below the median, age 39 or less, interview in Spanish, married, small land size, 2 or fewer children). The first column reports the effect of Added Privacy for the $C = 1$ group ($\beta + \delta$), the second column for the $C = 0$ group (β), and the third column their difference δ ($C = 1$ minus $C = 0$). Brackets show the p-value for each coefficient being different from zero and, for the Difference column, the p-value that the two group-specific effects differ from each other. The dependence ratio row uses 997 observations (337 clusters) because the ratio is missing for three respondents. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level.

Table A5: Experimental Results Pooling Privacy Options, With Controls

	Index	Out Alone	Small Purchases	Visit Friends	Asset Purchases	Support Relatives	Child Health
Added Privacy	-0.165*** (0.056) [0.013]	-0.046 (0.031) [0.219]	-0.119*** (0.027) [0.000]	-0.041 (0.031) [0.243]	0.034 (0.023) [0.219]	0.016 (0.031) [0.640]	-0.067* (0.035) [0.148]
Control Mean	0.000	0.620	0.690	0.681	0.227	0.404	0.574
Obs.	1000	1000	1000	1000	1000	1000	807
R-squared	0.167	0.031	0.051	0.060	0.283	0.132	0.121
Clusters	338	338	338	338	338	338	308

Notes: Each column shows the coefficient of regressing each answer on the pooled Added Privacy treatment.

The dependent variables are related to the next questions: Out alone: “Are you allowed to go alone to meet your friends for any reason (chatting, lunch, etc.)?”; Small Purchases: “Do you have to ask permission to buy vegetables or fruit; clothes for yourself; medicines or personal supplies?”; Visit Friends: “Who usually decides whether you can visit a friend’s/neighbour’s house?”; Asset Purchases: “Who contributes most to the decision to buy new assets (e.g. hoe or spray pump)?”; Support relatives: “Who decides whether to support parents, in-laws, siblings, etc.”; Child health: “When making decisions regarding children’s health care, who usually makes the decision?”

Index is a summary index of all agency questions, using the method in (Anderson, 2008), standardized using the control-group mean and standard deviation.

Regressions additionally control for: respondent age, marital status, log of land size, education level, partner education level, if respondent is the household head, land ownership status, number of children, number of family members, number of adult men and women in household, and DIGITAGRO status (control arm, treatment arm, or not enrolled).

Clustered standard errors at the village level in parentheses. FDR-q values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Result Table - Fixed Effects by Municipality - No Controls

	Index	Out Alone	Small Purchases	Visit Friends	Asset Purchases	Support Relatives	Child Health
Added Privacy	-0.176*** (0.064) [0.024]	-0.044 (0.032) [0.326]	-0.114*** (0.028) [0.001]	-0.037 (0.030) [0.366]	0.019 (0.028) [0.664]	0.005 (0.035) [0.895]	-0.072* (0.037) [0.144]
Control Mean	0.000	0.620	0.690	0.681	0.227	0.404	0.574
Obs.	1000	1000	1000	1000	1000	1000	807
R-squared	0.008	0.002	0.013	0.001	0.000	0.000	0.005
Clusters	338	338	338	338	338	338	308

Notes: Clustered Standard Errors at the village level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01. FDR-q values in brackets.

Table A7: Result Table - Fixed Effects by Municipality - Controls

	Index	Out Alone	Small Purchases	Visit Friends	Asset Purchases	Support Relatives	Child Health
Added Privacy	-0.155*** (0.054) [0.018]	-0.046 (0.032) [0.252]	-0.111*** (0.027) [0.000]	-0.034 (0.030) [0.346]	0.031 (0.022) [0.252]	0.011 (0.031) [0.729]	-0.063* (0.035) [0.189]
Control Mean	0.000	0.620	0.690	0.681	0.227	0.404	0.574
Obs.	1000	1000	1000	1000	1000	1000	807
R-squared	0.156	0.030	0.035	0.052	0.273	0.120	0.115
Clusters	338	338	338	338	338	338	308

Notes: Clustered Standard Errors at the village level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01. FDR-q values in brackets.

Table A8: Result Table - Interviewer fixed effects, no controls

	Index	Out Alone	Small Purchases	Visit Friends	Asset Purchases	Support Relatives	Child Health
Added Privacy	-0.187*** (0.068) [0.027]	-0.046 (0.034) [0.274]	-0.133*** (0.031) [0.000]	-0.047 (0.031) [0.261]	0.025 (0.031) [0.558]	0.019 (0.036) [0.682]	-0.069* (0.040) [0.225]
Control Mean	0.000	0.620	0.690	0.681	0.227	0.404	0.574
Obs.	1000	1000	1000	1000	1000	1000	807
R-squared	0.067	0.064	0.091	0.072	0.070	0.057	0.073
Clusters	338	338	338	338	338	338	308

Notes: Each column shows the coefficient of regressing each answer on the pooled Added Privacy treatment. Regressions include interviewer fixed effects.

The dependent variables are related to the next questions: Out alone: “Are you allowed to go alone to meet your friends for any reason (chatting, lunch, etc.)?”; Small Purchases: “Do you have to ask permission to buy vegetables or fruit; clothes for yourself; medicines or personal supplies?”; Visit Friends: “Who usually decides whether you can visit a friend’s/neighbour’s house?”; Asset Purchases: “Who contributes most to the decision to buy new assets (e.g. hoe or spray pump)?”; Support relatives: “Who decides whether to support parents, in-laws, siblings, etc.”; Child health: “When making decisions regarding children’s health care, who usually makes the decision?”

Index is a summary index of all agency questions, using the method in ([Anderson, 2008](#)), standardized using the control-group mean and standard deviation.

Clustered standard errors at the village level in parentheses. FDR-q values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Indices - No Controls

	Index	Average Index, w/ Imput.	Average Index, No Imput.	Standardized Index	Index from Q1-Q5	Index PC
Added Privacy	-0.188*** (0.066) [0.029]	-0.043** (0.018) [0.033]	-0.045** (0.020) [0.042]	-0.094** (0.039) [0.033]	-0.037** (0.018) [0.049]	-0.172* (0.097) [0.078]
Control Mean	0.000	0.533	0.528	0.000	0.525	0.095
Obs.	1000	1000	807	1000	1000	1000
R-squared	0.009	0.007	0.007	0.006	0.004	0.004
Clusters	338	338	308	338	338	338

Notes: Clustered Standard Errors at the village level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01. Index and Standardized Index are standardized using the control-group mean and standard deviation.

Table A10: Indices - Controls

	Index	Average Index, w/ Imput.	Average Index, No Imput.	Standardized Index	Index from Q1-Q5	Index PC
Added Privacy	-0.169*** (0.056) [0.016]	-0.038** (0.015) [0.022]	-0.042** (0.017) [0.023]	-0.082** (0.032) [0.022]	-0.033** (0.015) [0.035]	-0.137* (0.078) [0.080]
Control Mean	0.000	0.533	0.528	0.000	0.525	0.095
Obs.	1000	1000	807	1000	1000	1000
R-squared	0.166	0.189	0.184	0.195	0.167	0.229
Clusters	338	338	308	338	338	338

Notes: Clustered Standard Errors at the village level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01. Index and Standardized Index are standardized using the control-group mean and standard deviation.

Table A11: Indices - Fixed Effects by Municipality

	Index	Average Index, w/ Imput.	Average Index, No Imput.	Standardized Index	Index from Q1-Q5	Index PC
Added Privacy	-0.176*** (0.064) [0.036]	-0.040** (0.017) [0.038]	-0.044** (0.020) [0.038]	-0.087** (0.037) [0.038]	-0.034** (0.017) [0.056]	-0.156* (0.093) [0.093]
Control Mean	0.000	0.533	0.528	0.000	0.525	0.095
Obs.	1000	1000	807	1000	1000	1000
R-squared	0.008	0.006	0.007	0.006	0.004	0.003
Clusters	338	338	308	338	338	338

Notes: Clustered Standard Errors at the village level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01. Index and Standardized Index are standardized using the control-group mean and standard deviation.

Table A12: Indices - Fixed Effects by Municipality, With Controls

	Index	Average Index, w/ Imput.	Average Index, No Imput.	Standardized Index	Index from Q1-Q5	Index PC
Added Privacy	-0.155*** (0.054) [0.027]	-0.035** (0.014) [0.029]	-0.039** (0.017) [0.029]	-0.074** (0.031) [0.029]	-0.030** (0.014) [0.049]	-0.122 (0.075) [0.106]
Control Mean	0.000	0.533	0.528	0.000	0.525	0.095
Obs.	1000	1000	807	1000	1000	1000
R-squared	0.156	0.179	0.173	0.185	0.158	0.219
Clusters	338	338	308	338	338	338

Notes: Clustered Standard Errors at the village level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01. Index and Standardized Index are standardized using the control-group mean and standard deviation.