FLETCHER
HENRY J. LEIR INSTITUTE FOR
MIGRATION & HUMAN SECURITY

REPORT | MARCH 2023

TECH-LED LISTENING: HOW WE LEARNED TO LISTEN FOR EVIDENCE OF CHANGING GENDER NORMS

Henry J. Leir Institute for Migration and Human Security

Our mission is to help policymakers and practitioners develop more equitable and sustainable responses to migration and its root causes by employing a human security approach. For more information, please visit go.tufts.edu/leir.

Decodis is a social research firm that creates tech-led, customized data capture and analysis to elevate the voices of vulnerable populations. For more information, please visit www.decodis.com.

Dr. Daryl Collins is the CEO and founder of Decodis and a Leir Senior Fellow.

Derry Moore is the Lead Linguist at Decodis and a Ph.D. candidate at SUNY Buffalo.

Pravarakya Reddy is the Director of Data Analytics at Decodis and a post-graduate of Purdue University.

Marcia Almeida-Mendes is the Director of Storytelling at Decodis and a master's graduate of The Fletcher School.

DECODIS

EXECUTIVE SUMMARY

As social science researchers, we seek to understand human behavior and interactions. However, the act of studying people itself is an engagement between two people and there is a small but growing literature on interview modes and social acceptance biases. Literature on social acceptance bias draws out some of the limitations of traditional social research interview modes between individuals, with data largely reliant on self-reporting by respondents. The propensity for respondents to represent themselves in a socially acceptable or desirable light potentially hampers the quality of the data elicited. This is especially so, as many papers argue, when researching subjects considered taboo like sexually transmitted diseases,[1] domestic violence, and other sensitive topics in certain sociocultural contexts. For instance, the need for social desirability is said to affect consistency[2] in responses, which in turn affects the quality of the data. Social scientists, especially at the intersection of psychology, have used techniques like 'cognitive loading' or the process of 'giving subjects more to think about or focus on as they perform a task, which therefore occupies some of their cognitive power'[3] as a means to offset social acceptance bias.

This article explores how tools such as Interactive Voice Recording (IVR) could both offer anonymity and reduce the power imbalances brought on by an in-person researcher-respondent dynamic, thus eliciting information free of the limitations posed by social acceptance bias. In this paper, we explore using a form of open-ended, IVR surveys to ask respondents in economic inclusion programs with a gender transformative approach in Paraguay and Colombia about attitudes towards gender issues as a means to not only increase our ability as researchers to accurately understand their thoughts, but also to assess whether the program had an effect on a comparison group of women not in the program.

1 Gregson et al (2002) "Methods to Reduce Social Desirability Bias in Sex Surveys in Low-Development Settings: Experience in Zimbabwe" *Sexually Transmitted Diseases* 29(10) pp. 568-575.

2 Mensch et al (2008) "Sexual Behavior and STI/HIV Status Among Adolescents in Rural Malawi: An Evaluation of the Effect of Interview Mode on Reporting" Stud Fam Plann 39(4) pp. 321-334.

3 Stodel (2015) "But what will people think? Getting beyond social desirability bias by increasing cognitive load" International Journal of Market Research 57(2) pp. 313-321.

TABLE OF CONTENTS

INTRODUCTION

Fundación Capital is a Bogota based social enterprise that aims to improve the economic and social lives of vulnerable populations around the world. In 2020, they launched economic inclusion programs in Colombia and Paraguay with a gender transformative approach. One of the gender component activities was to digitally distribute gender norms materials, such as skits and infographics over WhatsApp, to women in low-income communities across the country. Annex 1 attached to this report shows an example of one such infographic on domestic violence for Paraguay. *One of the objectives of both programs was to understand women's opinions on a host of social issues common in the region* ***and to reinforce norms that increase gender equality and decrease gendered violence.***

To understand if their message was getting across to their participants, Fundación Capital partnered with Decodis, a social research firm based in Boston, Massachusetts and associated with the Fletcher School at Tufts University's Leir Institute for Migration and Human Security. Decodis uses technology in its social research methods that aim to elicit robust responses from participants without the constraints present in traditional in-person survey methods. The aim of the partnership was to understand if the digitally mediated messaging strategy was effective in communicating its intended message, even during the COVID-19 lockdown.

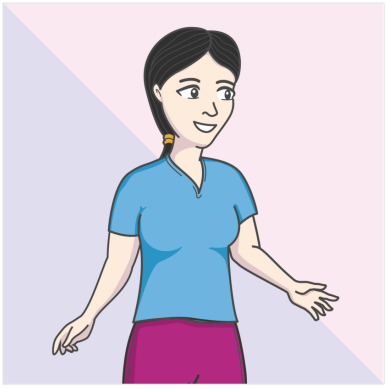
Using a combination of Interactive Voice Recording (IVR), Natural Language Processing (NLP) and sociolinguistics, Decodis has created a method to deploy qualitative surveys and analyze speech signals to elicit nuanced insights in a cost-effective and efficient way. The process is intended to help prioritize programming changes to increase impact for budget-constrained programs.

Using these methods, Decodis analyzed over 300 women's responses to Fundación Capital's gender norms modules in Colombia and Paraguay. The tech-led methods helped get in-depth and 'honest' responses from participants as it afforded both anonymity and a sense that they were speaking to a 'peer' rather than a researcher. Most importantly, because it can be done quickly and inexpensively, it can be used as an early (and continued) assessment of social programming, allowing for course correction.

Designing remote and scalable qualitative research

For this research exercise, Decodis assembled a survey with questions about the topics covered in the distributed materials. Examples are shown below in Diagram 1.

Diagram 1: Examples of Topics and Related Questions



- Yoli was designed as one of the characters of the materials, she was introduced as one of their friends.
- She was voiced by different actresses to account for regional differences in colloquial language. In Paraguay, she spoke Guaraní.
- She asked questions using voice notes, making it easier for low-literacy respondents to participate.
- Respondents often thanked Yoli for asking their opinion.

Ctrl + click on the speaker to listen to Yoli's introduction.

Diagram 2: Yoli, the Digital Interviewer

The respondents included both women who had received the materials about gender norms and a (much smaller) comparison group who had not received the materials but were in the same social programs and had similar socioeconomic characteristics to the other respondents. This study was not a Randomized Control Trial (RCT) and thus was not seeking to prove impact or attribution of differences in respondents' gender norms attitudes to the program. However, it did provide Fundación Capital an early and agile indicator of whether the gender norms material was being reflected in the responses of women who were exposed to the materials versus those who were not.

The data collection for this project was fully remote and happened in two phases: December 2020 in Paraguay and August 2021 in Colombia. We took advantage of an existing communication channel between Fundación Capital and the respondents: WhatsApp. To participate in the survey, a respondent would receive an invite over WhatsApp to speak to Yoli, a digital interviewer. Yoli is a bot powered by a third-party automated survey system. She asked questions using voice notes, and the respondent was expected to answer also using voice notes. Sending voice notes is common among the study population, thus leveraging a familiar practice while eliminating barriers for low-literacy or illiterate respondents. Diagram 2 below shows a drawing of Yoli and discusses some of the attributes of the interview experience.

Each survey consisted of 10-12 questions which respondents answered with an open-ended response via a voice note. To analyze these responses, Decodis transcribed and translated the audio responses and then used Natural Language Processing (NLP) to analyze the transcriptions. The NLP process identified and categorized topics that emerged from each question's responses. Diagram 3 below shows the two-step process. In Stage A, exploratory research leverages topic modeling, which generates key words in a block of text. For example, responses to the question *"Is it ok for your partner to check your phone?"* the word *trust* is a key word. However, knowing that respondents used the word trust is not enough to gain real insights as it can be used in different ways. Stage B therefore uses topic classification to find key phrases – not just words – that surround the word *trust.* For example, one woman might say *"If I trust them, I don't need to check their phone"* which is the type of gender-positive response that reflects a woman's strong sense of confidence in a partner, as well as her own sense of privacy. Another woman might say "We check each other's phones because it builds trust," which is less gender-positive and confident. Using Stage A to first find key words and then Stage B to find key phrases generates

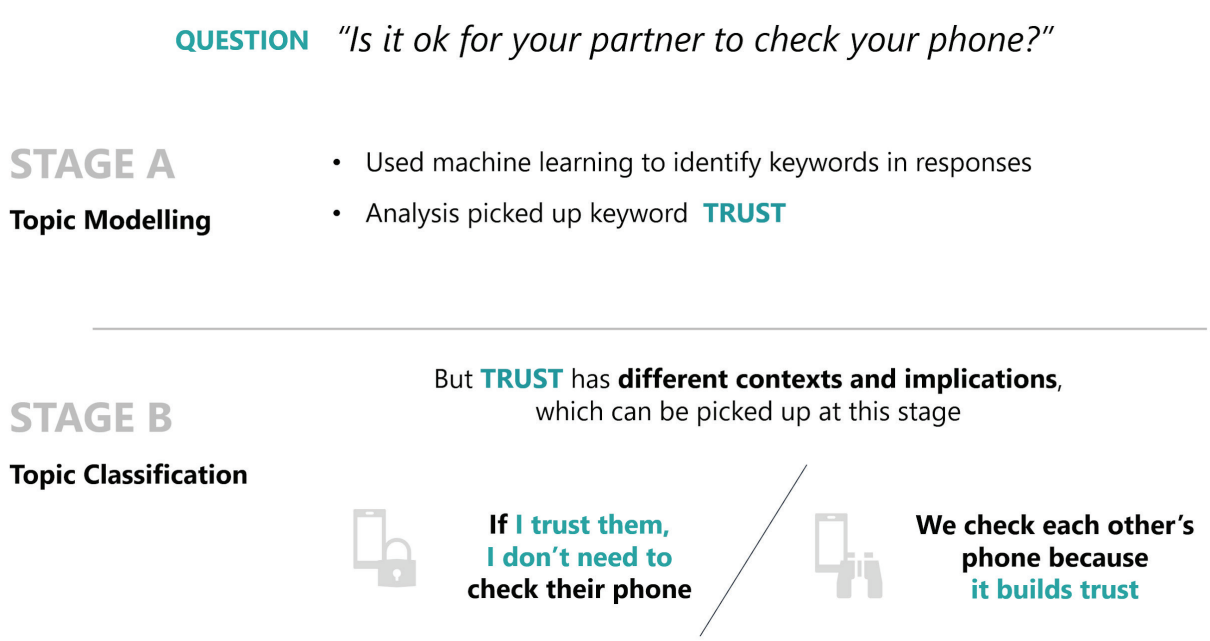


Diagram 3: Natural Language Processing

faster insights from open-ended responses than a typical qualitative researcher. The Decodis technology analyzed a total of over 30 hours of audio recordings for over 300 respondents within two days. We estimate that the amount of time a single qualitative researcher would have taken to analyze this amount of data would have taken closer to two months.[4]

In addition, Decodis also analyzed the audio recordings themselves, using software Praat to extract speech signals which helped provide a sociolinguistic interpretation of the emotional content of the responses. Diagram 4 shows some of the details of sociolinguistic analysis. There are approximately 28 different speech signals that researchers have developed to extract from audio recordings. These speech signals are then compared across different languages, population segments and contexts to test their universal robustness. There are an estimated 20,000 peer-reviewed academic articles across many countries and languages that inform how speech signals can be interpreted in what contexts.[5] In essence, the NLP analysis generates data on *what* respondents say while speech signals generate data on *how* they say it. On a practical level, the speech signal analysis also helped identify instances where women's answers were being dictated by a male voice or a man was answering the survey (because of the low pitch of the voice) in lieu of the woman invited to take the survey.

"Thanks for Asking"

When thinking about remote data collection on qualitative research with such sensitive topics, one might imagine that digital interviews are distant and removed from participants' reality. However, this method did the opposite – respondents spoke longer for each response.[6] The method allows for customization and offers a sense of freedom for respondents to speak openly, which would otherwise not be possible in the presence of an interviewer and the

4 Typically, there are two words spoken in every second of audio, and there are typically 250 words on a page. Using these assumptions, this amount of audio would be over 860 pages of text. Some estimates suggest that 20 pages of text could take one day to analyze. This suggests that it might take 43 days to complete analysis on this much text.

5 Examples are El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern recognition,* 44(3), 572-587; Scherer, Klaus R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication,* vol. 40, no. 1-2, p. 227-256; and Feraru, S. M., D. Schuller & B. Schuller. (2015). Cross-language acoustic emotion recognition: An overview and some tendencies. *2015 International Conference on Affective Computing and Intelligent Interaction* (ACII), pp. 125-131.

6 In a Decodis study on farmers in northern Nigeria, a cohort analysis between in-person qualitative interviews and IVR interviews showed that respondents spoke at least as long in their IVR interviews as in the in-person qualitative interviews. Next, Decodis plans to do a pairwise comparison between the two methods and compare not only duration of responses but also content richness, i.e. distinct themes that can be extracted from responses. See www.decodis.com for more information

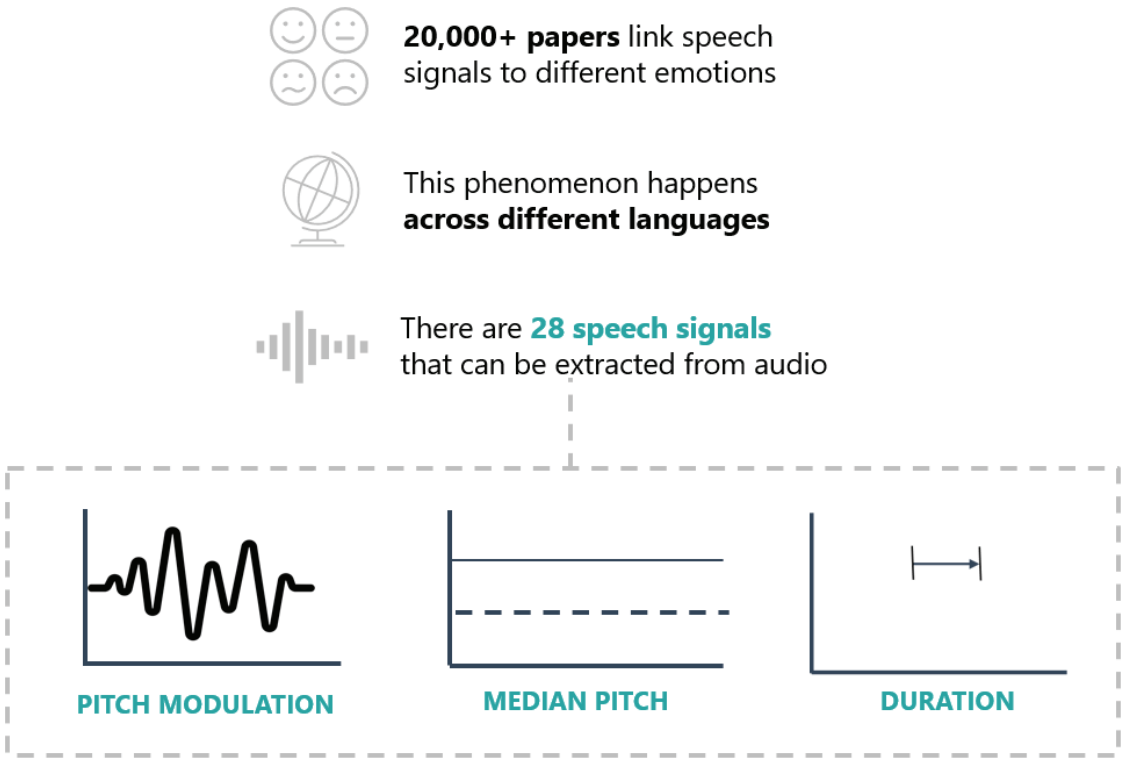


Diagram 4: Sociolinguistic analysis

various power dynamics that might be at play. Women were more comfortable talking to who they saw as a "peer" asking their opinion rather than a "researcher." This method allowed even women who were quiet in groups the opportunity to express themselves fully.[7] During the survey, Yoli encouraged respondents to speak for as long as they wanted. They answered at length and had none of the interruptions that creep in if a physical interviewer was collecting the data.

One of our indicators that women felt empowered by this method of data collection is that some of the women thanked Yoli for asking them their opinion in the survey. This reaction suggests that respondents did not see the survey as an "evaluation" of them but rather a consultation of their opinion during a program. In turn, the detailed responses generated, coupled with speech signal analysis methods, generated the type of rich insights that can be used for decision-making.

Insights: Not just interesting, but a tool for turning up the impact

The insights that emerge are certainly interesting in themselves. However, the objective of this evaluation exercise was not simply to show whether what Fundación Capital was doing "worked" or "did not work" but rather to provide recommendations on which modules were having an impact versus those which might either need to be adjusted or re-designed altogether. This was the purpose of conducting an evaluation only 6 months after the program had launched. Fundación Capital suspected that some of the modules might need to be adjusted and they wanted to know which ones and how. Therefore, this analysis had to prioritize its recommendations to make best use of limited resources.

Decodis arrived at recommendations by considering two criteria. First, for each question, the Fundación Capital team reflected on what they would consider a "good" response, as defined

7 We were not able to do a rigorous analysis to show this, but the Fundacion Capital had observations such as suggesting that one group or another were "shy" and didn't talk. In the study, we did see that the women in these groups spoke just as much as others.

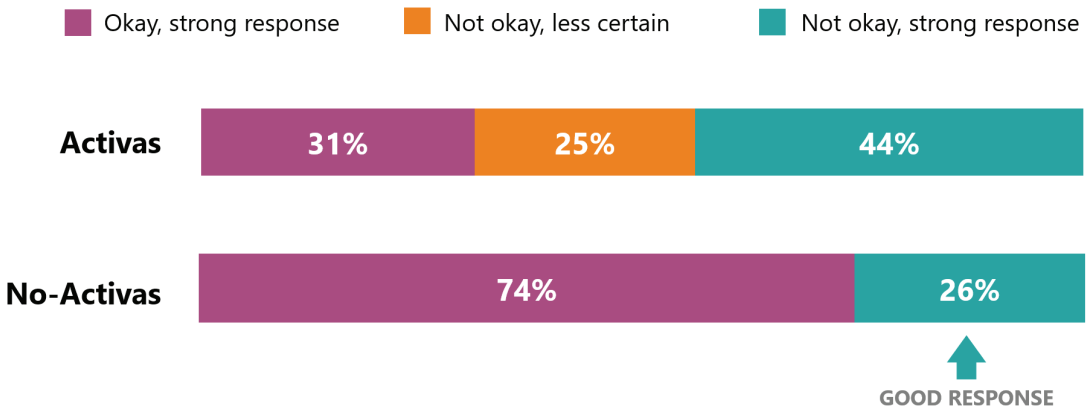


Diagram 5: Types of responses to checking partner's phone in Paraguay

by a combination of what a respondent said and how she said it. Second, the respondents who were involved in the program ("Activas") were compared to respondents who were not in the program ("No-Activas") as a means of assessing whether the gender attitudes of those in the program were much different from those who were not. The example below demonstrates our analysis process.

A. Example of a common norm: Checking a woman's phone

In many cultures, between men and women partners, men (and women) can exert control over the partner by checking each other's phone to see if they are in contact and perhaps flirting or seeing another person.[8] The Fundación Capital materials suggest that this is not necessary if the couple has trust – to what extent had the *activas* had internalized this message?

We asked the question below:

Si a una mujer su compañero o su pareja le revisa su celular, están los que piensan que su pareja hace eso por controlador, por argel, y están los que piensan que oî porâ upéa. ¿Cómo es en tu experiencia?

An English translation would be: "Some people in your neighborhood might think a man checking his partner's cell phone is too controlling, while others might think he has the right to do this. What do you think?"

Diagram 3 provides the types of responses to the question in Paraguay. There were three different types of responses:

1. Some said that checking each other's phones was perfectly okay, and they had speech signals that showed that they felt quite strong about that opinion;
2. At the other end of the spectrum, there were those who responded who said that checking each other's phone was not okay, and their speech signals also showed that they felt strongly about that opinion;
3. Lastly, there were those who responded who said that it was not okay but their speech signals were uncertain. They sounded as though they knew what they were supposed to say, but they didn't quite believe it.

The Fundación Capital team then listened to the different types of responses and asked themselves "How do you feel the responses in categories 1, 2, or 3 were in terms of the success of the program?" The Fundación Capital materials that these questions built upon were intended to address and shift gender normative behavior indicative of controlling relationships (checking partner's cell phones). Therefore, the team was very pleased with the responses in category 2. It was considered a success for these women to describe their phone as their "private space." The transcriber in Paraguay pointed out that their use of the word "privacy" seemed new to them; this was not a word they were used to using, and they were getting used to speaking this word with respect to themselves. The respon-

8 For example, see Belknap J., A. Chu, and A. DePrince. 2012. *The Roles of Phones and Computers in Threatening and Abusing Women Victims of Male Intimate Partner Abuse,* 19 Duke Journal of Gender Law & Policy. 373 - 406. Spring 2012.https://scholarship.law.duke.edu/djglp/vol19/iss2/4.

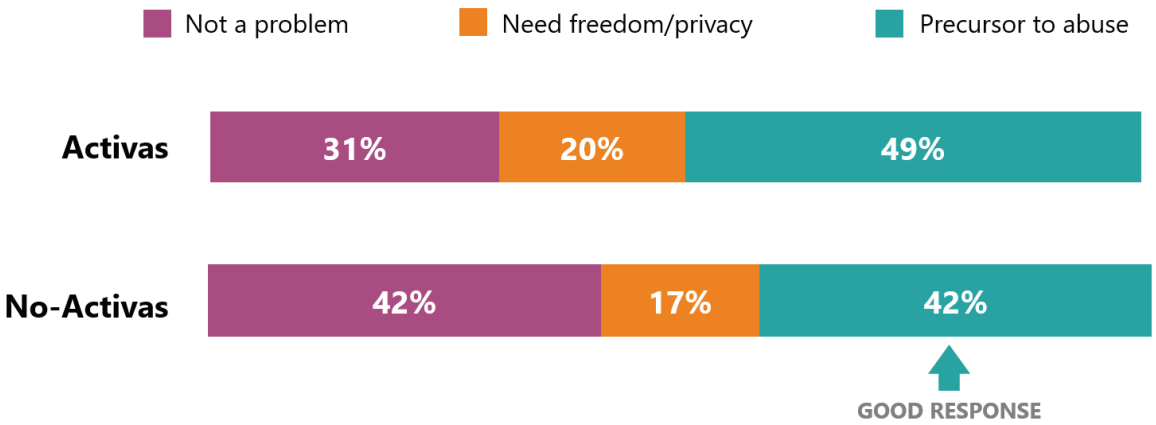


Diagram 6: Types of responses to checking each other's phone in Colombia

dents said that it was not necessary for phones to be checked by either party **because they have trust.**

On the other hand, those who had a less progressive attitude spoke in a completely different way, saying that checking phones **builds trust.** A common reasoning was that if he could check her phone, it meant she could check his, which she wanted to do.

The results of those who had received the Fundación Capital materials were certainly better than those who had not. Among those who had not received materials, 74% said confidently that it was okay to check each other's phones, while only 26% said, confidently, that it was not okay. In Activas, however, only 31% said that checking the phone was okay, and 44% confidently stated it was not okay. Even though the responses of the Activas were better than the no-Activas, a quarter of Activas respondents (25%) were less certain in stating it was not okay. The results indicate that the materials are effective, though may need more circulation and "resending" to achieve greater reach.

Responses to this question were more positive in Colombia than in Paraguay. A greater percentage of those who had seen the gender norms material expressed a strong view that this was not only not okay but that it was in fact a precursor to abuse. Moreover, the word privacy was used more easily than in Paraguay. That said, almost a third of the sample believes that checking phones is not a problem.

Table 2 shows that some women used very strong language, but most referred to controlling. A significant number used the word *celo,* which carries a connotation of subjugation (or subjection) though its use often implies a sense of "caring."

Table 2: Common word usage about partners checking phones in Colombia

Words for abuse	English comparable	Activas	No-Activas
Violación	Violation (of privacy)	3%	0%
Violencia	Violence	2%	0%
Abuso	Abuse	2%	4%
Maltrato	Abusive attitude	2%	0%
Controlador	Controlling	21%	8%
Celo(s)	Jealousy	11%	21%

Program materials: revamped, recycled or retired?

These results can then be used to rank the effectiveness of materials' intended topics, or "hitting at the level intended." Fundación Capital had a fixed budget for revamping materials, and thus had to prioritize which topics to reinforce. Topics were therefore split into three different categories.

The first were those where a revamping of materials should be considered. These are topics where less than 50% of the Activas had the type of response that Fundación Capital considered "good." The topic of managing work and home life particularly stands out. For example, very few felt they could ask for help; most respondents had speech signal that were tired or emphasized trying to be "superwoman" rather than asking for support. But this was not much different from prevailing gender attitudes shown by the No-Activas.

A second category recommended recycling existing materials where low "good" response rates were recorded but Activas demonstrated evolving attitudes toward gender norms when compared to No-Activas. Results from the phone checking question above, for example, demonstrated this effect. Although fewer than 50% of Activas showed "good" responses, their attitudes were better than No-Activas, indicating that more could be brought into a "good" response through further exposure to the existing content. Therefore, for these topics, the materials should remain in circulation with Activa women.

The last category included topics where a high percentage of women indicated a "good" response, and there was little to no difference among the attitudes of Activas and no-Activas respondents. This suggests that prevailing gender attitudes were already "good," meaning materials should be de-prioritized and therefore retired. At an appropriate time, when the program has greater resources, the team should consider expanding these topics' materials to "reach higher." For example, in the "women in leadership" materials the question simply asked if a women could handle leadership as well as a man, in principle. It does not ask, for example, whether women might actually have advantages to men, or whether the women themselves might assume a position in local leadership, which is a step up in promoting more progressive social norms.

Table 1: Recommendations to re-vamp, recycle or retire

Topic	Recommendation
Managing work and home life for women	Revamp
Violent words towards children	Revamp
Inappropriate attention towards young woman	Recycle
Checking partner's phone	Recycle
Telling children of intimate parts	Recycle
Sharing household work	Recycle
Contraception in marriage	Recycle
Married sex	Recycle
Toys for girls vs boys	Retire
Women going out alone	Retire
Women leadership	Retire

CONCLUSION

Impact measurement for resource-limited institutions

This impact measurement clearly does not stand up to the "gold standard" of a Randomized Control Trial (RCT). It does not come even close to adequately assigning attribution to Fundación Capital's gender norms program, which leaves open the possibility that some other outside and unknown influence is changing gender norms rather than the program. Therefore, Fundación Capital cannot claim that their program alone created a change in gender norms.

However, this measurement has allowed Fundación Capital to do something that is, for them, perhaps more helpful. During COVID, when they could not deliver their program in person, they shifted to a digital delivery of their materials. This shift to digital would not only allow them to continue reach women when they most needed it during the pandemic, but it would also allow them to deliver materials cheaper and at a greater scale even after the pandemic. A significant, if not the largest, cost of any program or business is staff resources – especially staff travel. By using digital means for content delivery, Fundación Capital can then use staff for higher-value activities, such as creating impactful program materials.

The key question Fundación Capital had when they made this shift was "Is this material still impactful if it isn't delivered in person?" They asked this question early in their digital delivery program to understand whether their materials seemed to be "hitting at the right level" so they could adjust them if not. Moreover, this type of evaluation embraced a notion that Fundación Capital already knew – that shifting human behavior and opinions, such as gender norms, does not happen like an on/off switch, but rather a journey over time.

ANNEX 1: EXAMPLE OF FUNDACIÓN CAPITAL INFOGRAPHIC IN PARAGUAY ABOUT REPORTING DOMESTIC VIOLENCE