

SURVEY MODE EFFECTS IN A DEVELOPING COUNTRY: COMPARING PHONE AND FACE-TO-FACE SURVEYS IN COSTA RICA

*Efectos del modo de encuesta en un país en desarrollo:
Comparando encuestas telefónicas y presenciales en Costa Rica*

*Efeitos do modo de pesquisa em um país em desenvolvimento:
Comparando pesquisas por telefone e face a face na Costa Rica*

NOAM LUPU  noam.lupu@vanderbilt.edu ¹

ADAM WOLSKY  awolsky@princeton.edu ²

¹ Vanderbilt University

² Princeton University

Submission: 2022-08-24

Accepted: 2022-12-20

First View: 2022-12-28

Publication: 2022-12-31

Keywords:

surveys;
mode effects;
phone surveys;
face-to-face
surveys;
Costa Rica

Abstract

Responses to phone surveys tend to exhibit higher rates of social desirability bias and extreme responses when compared to face-to-face surveys. Yet, studies of mode effects typically compare either representative studies that implausibly assume comparability or experimental studies that rely on convenience samples. Our study compares two national probability samples but uses matching to address comparability. We study Costa Rica, a middle-income democracy, to see whether the conventional wisdom drawn from Western Europe and North America extends to the Global South. We analyze two nationally representative surveys, one fielded by phone and one face-to-face, allowing us to compare identically worded items we placed on both surveys. We find that phone respondents exhibited more socially desirable responding and were more likely to choose negative endpoints on scalar items. This suggests that survey researchers and practitioners should carefully assess the tradeoffs in shifting modes or employing mixed modes.

Palabras clave:

*encuestas;
efectos del modo;
encuestas
telefónicas;
encuestas
presenciales;
Costa Rica*

Resumen

Las respuestas a las encuestas telefónicas generalmente generan tasas más altas de sesgo de deseabilidad social y respuestas extremas en comparación con las encuestas cara a cara. Sin embargo, los estudios de los efectos de modo generalmente comparan estudios representativos que asumen una comparabilidad implausible o estudios experimentales que se basan en muestras de conveniencia. Nuestro estudio compara dos muestras probabilísticas nacionales, pero utiliza matching para abordar la comparabilidad. Estudiamos el caso de Costa Rica, una democracia de ingresos medios, para ver si la sabiduría convencional extraída de Europa Occidental y América del Norte se extiende al Sur Global. Analizamos dos encuestas nacionales representativas, una realizada por teléfono y la otra cara a cara, lo que nos permite comparar preguntas idénticas que colocamos en ambos cuestionarios. Encontramos que los encuestados por teléfono tienden más a dar la respuesta más socialmente deseable y a elegir valores extremos negativos en preguntas con escalas. Esto sugiere que los investigadores de encuestas y los profesionales deberían evaluar cuidadosamente las ventajas y desventajas de un cambio de modo o del uso de modos mixtos.

Palavras-chave:

*pesquisas;
efeitos do modo;
pesquisas
telefónicas;
pesquisas face a
face;
Costa Rica*

Resumo

As respostas a pesquisas por telefone geralmente geram taxas mais altas de viés de desejabilidade social e respostas extremas em comparação com pesquisas face a face. No entanto, estudos de efeitos do modo geralmente comparam estudos representativos que assumem comparabilidade implausível ou estudos experimentais baseados em amostras de conveniência. Nosso estudo compara duas amostras probabilísticas nacionais, mas usa matching para abordar a comparabilidade. Estudamos o caso da Costa Rica, uma democracia de renda média, para ver se a sabedoria convencional adquirida na Europa Ocidental e na América do Norte se estende ao Sul Global. Analisamos duas pesquisas nacionais representativas, uma realizada por telefone e outra face a face, permitindo-nos comparar perguntas idênticas que colocamos em ambos os questionários. Descobrimos que os entrevistados por telefone eram mais propensos a dar a resposta socialmente desejável e a escolher valores extremos negativos para perguntas com escalas. Isso sugere que pesquisadores e profissionais de pesquisa devem pesar cuidadosamente as vantagens e desvantagens de alternar modos ou usar modos mistos.

INTRODUCTION

Phone coverage has become nearly universal in many low- and middle-income countries with the expansion of cellphones. According to the International Telecommunication Union, in 2016, developing countries had 94 cellphone subscriptions per 100 individuals (Gibson *et al.*, 2017). In the 2018/19 round of the AmericasBarometer, 90 percent of respondents in the average Latin American and Caribbean country reported having access to a cellphone in their household. Given these trends, phone surveys are becoming a viable alternative to the face-to-face surveys that are considered the gold standard in developing contexts (e.g., Montalvo, Pizzolitto, and Plutowski, 2022).

Face-to-face surveys have some clear advantages over phone surveys, including higher response rates, longer questionnaires, the ability to use visual aids and cues, and better rapport between interviewer and respondent (e.g., Holbrook *et al.*, 2003; Hox and de Leeuw, 1994; Jäckle *et al.*, 2010; Roberts and Jäckle, 2012). But phone surveys have some advantages of their own, including potential cost savings, the ease of re-contacting respondents, and the potentially broader geographic coverage as compared to clustered area sampling. Phone surveys became even more vital during the COVID-19 pandemic, which made it challenging to carry out face-to-face surveys safely. Online surveys represent another alternative—with its own potential mode effects (e.g., Homola, Jackson, and Gill, 2016; Shino and Martinez, 2021)—but recent studies suggest that they fail to produce nationally representative samples in developing contexts like Latin America (Castorena *et al.* Forthcoming). Researchers wishing to draw nationally representative samples in developing contexts therefore have to choose between face-to-face and phone surveys.

One issue in considering phone surveys is the possibility of mode effects, such that data quality or responses themselves differ when people are interviewed over the phone rather than in-person. Social-scientific studies in the US and Europe generally find that responses to phone surveys tend to exhibit more social desirability bias when compared to face-to-face (e.g., Dahlberg and Persson, 2014; Holbrook *et al.*, 2003; Holbrook and Krosnick, 2010; Roberts and Jäckle, 2012) or online surveys (e.g., Kreuter *et al.*, 2008). Similarly, studies in public health find that respondents in phone surveys are less likely to admit drug and alcohol use (e.g., Aquilino, 1991; 1994).¹ Studies also show that for questions that ask respondents to use a numeric scale, purely aural survey modes like phone tend to elicit more extreme responses than more visual modes like face-to-face surveys with visual aids (Christian *et al.* 2009; Dillman *et al.* 2009). In particular, these studies find that respondents in phone surveys tend to choose the positive extreme of the scale more often, a result of primacy bias since the positive endpoint is typically mentioned first.

Studies of mode effects involve methodological tradeoffs. Observational studies, which typically use generalizable samples, rarely address the possible differences between those samples, making it difficult to draw accurate comparisons. Experiments avoid this limitation by assigning survey mode after sampling, but typically rely on convenience samples or very local probability samples because of the difficulty of randomly assigning mode in a national probability survey—and this raises questions about generalizability. Taking a middle-ground approach, our

1. While this seems to be the modal finding, some studies also find no effect of mode on social desirability bias (Greenfield, Midanik, and Rogers, 2000; Vannieuwenhuyze, Loosveldt, and Molenberghs, 2010).

study relies on an observational comparison of two national probability samples but uses matching to address imbalances. This approach entails its own methodological limitations, which we return to in the final discussion below, but it provides us with the generalizability of national samples and helps us approximate an experimental ideal with observational comparisons.

What we know about mode effects, especially when it comes to surveys on politics, comes almost exclusively from studies conducted in the affluent democracies of Western Europe and North America. There may be reasons to expect similar findings in less-affluent contexts in the Global South, but we might also expect differences: surveys are far less common in developing countries, the typical respondent has less formal education and less familiarity with numeric scales, and differing cultural contexts may condition how respondents interact with enumerators across modes (see Lupu and Michelitch, 2018). This could mean that we would see stronger mode effects on extreme responses, given lower levels of overall familiarity with numeric scales in the population and the association between extreme responses and lower levels of formal education (e.g., Meisenberg and Williams, 2008). At the same time, it might also mean more attenuated mode effects with regard to social desirability bias, since those with more formal education appear to misreport more (e.g., Karp and Brockington, 2005). Would we see mode effects in developing contexts similar to those documented in more-affluent contexts?

To answer this question, we carried out two nationally representative surveys, one by phone and one face-to-face, in Costa Rica in 2018, several months after that year's two-round presidential election. Costa Rica is a Central American country that ranks in the middle third of countries by wealth and human development. Because of its comparatively long democratic history, it also has a history of survey research (see Seligson, 2020). Comparing effectively identical items we placed on the two surveys allows us to assess differences in social desirability bias and response styles for scalar questions.

This approach follows what many mode studies comparing national probability samples do, effectively treating the two independent samples as experimental groups. Of course, we know that phone and face-to-face surveys use different recruitment methods, potentially producing samples with important demographic and geographic differences (see discussion in Holbrook *et al.*, 2003). Rather than assume that there are no differences between our phone and face-to-face samples, as many prior studies do, we use coarsened exact matching to better approximate an experimental design. This makes us more confident that the differences we observe between our samples are driven by mode effects rather than sampling—a point we return to below.

Even after adjusting for observable covariates through coarsened exact matching, we find that phone respondents were significantly more likely to say that

they turned out to vote for both rounds of the 2018 presidential election. Phone respondents were also significantly less likely to agree that a woman should be able to have an abortion if her health is at risk and that same-sex couples should have the right to marry. Finally, phone respondents were significantly more likely to choose negative endpoints for three scalar questions included on both surveys. This suggests that Costa Ricans are more likely to engage in socially desirable responding for sensitive questions when interviewed by phone, consistent with prior work. Phone respondents also appear to rely more on endpoint responses, but unlike prior work, we find that they tend to choose the extreme negative options rather than the extreme positive ones.

DATA AND METHODS

The face-to-face survey we analyze in this paper is part of the 2018/19 round of the AmericasBarometer, a cross-national study fielded roughly every two years since 2004 across the Americas (more information about the AmericasBarometer can be found at www.vanderbilt.edu/lapop/). The Costa Rica survey was conducted between September 24 and October 31, 2018, included 1,501 respondents, and had a margin of error of $\pm 2.5\%$ and a response rate of 26% (AAPOR RR3). The survey employed a multistage clustered area sample, stratified by urban/rural residence and five geographical census regions. Like all AmericasBarometer surveys, respondent selection within the household relied on matching to gender and age group frequencies (see Lupu *et al.*, 2022). The fieldwork was conducted by Analítica, a well-respected local provider. Importantly, the face-to-face survey used show cards as visual aids for items with ordinal 1-7 or 1-10 response scales.

The phone survey was fielded by the Center for Political Research and Study at the University of Costa Rica. The study was carried out between August 13 and 16, 2018, included 720 respondents, and had a margin of error of $\pm 3.7\%$. Unfortunately, we do not have data for the response rate of this survey; a similar survey conducted November 2020 had a response rate of 8.8%. The phone survey randomly dialed numbers from a government-provided list of active cellphone numbers in Costa Rica, attempting each number up to three times.²

Both surveys only included residents or citizens of the country who were at least 18 years old and both instruments included nine effectively identically worded questions (in addition to demographics). Only the item on vote choice differed very slightly between modes (see appendix), but it was effectively identical.

2. See Alfaro Córdoba, Villareal Galera, and Navarro Cerdas (2010) and Montalvo, Pizzolitto, and Plutowski (2022) for discussions about the choice to use cellphones versus landlines.

However, the questionnaires of the two surveys were not identical, with the face-to-face questionnaire being significantly longer.

Table 1. Survey items included in both surveys

Item	Response scale	Social Desirability?
Internal efficacy	Ordinal (1-7)	No
Inequality	Ordinal (1-7)	No
Same-sex marriage	Ordinal (1-10)	Yes
Abortion	Binary	Yes
Turnout (round 1)	Binary	Yes
Turnout (round 2)	Binary	Yes
Vote choice (round 1)	Multicategorical	Yes
Vote choice (round 2)	Multicategorical	Yes

Note: See appendix for complete question wording in Spanish and English.

Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey

Table 1 lists the nine items included on both surveys, with the full wording of each one provided in English and Spanish in the online appendix. We included both standard political items that would seem uncontroversial to the average Costa Rican (internal efficacy and whether the government should reduce inequality) and others that could be viewed as sensitive (support for same-sex marriage legalization, support for abortions in cases where the mother's life is at risk, and vote choice) or subject to social desirability bias (turnout). Given the socially conservative cultural context in Costa Rica, we expect respondents may be reluctant to admit supporting same-sex marriage or abortion in certain instances. Same-sex marriage in particular had been a salient point of contention in the 2018 election following a decision by the Inter-American Court of Human Rights that required Costa Rica to recognize same-sex marriages (Pignataro and Treminio, 2019).

Like respondents elsewhere in the world (e.g., Adida *et al.*, 2019; Holbrook and Krosnick, 2010), Costa Ricans may also be motivated to over-report participation in elections in order to cast themselves in a favorable light. Finally, since voting is secret, respondents may feel uneasy about sharing their past vote choice with an interviewer, especially if they think the interviewer has a particular preference (Cassell and Cohen Forthcoming). Costa Rica's 2018 election was unlike prior elections, characterized by the decline of the country's traditional political parties

and the rising political influence of religious cleavages (see Alfaro-Redondo and Alpízar Rodríguez, 2020; Alfaro-Redondo and Gómez Campos, 2021; Colburn and Cruz S., 2018; Díaz González and Cordero Cordero, 2020; Pignataro and Treminio 2019). Respondents who voted for Fabricio Alvarado, who lost the runoff, might be especially reluctant to admit doing so given that Carlos Alvarado won the runoff by a landslide and given the extreme positions that Fabricio Alvarado took on cultural issues during the course of the campaign.

Although both surveys employed probability samples targeting the national population, some important demographic and geographic differences did emerge between the two effective samples (see Table A1).³ In particular, the phone survey has a significantly lower proportion in the highest age category, landline owners, and residents of San José, but a higher proportion of university-educated respondents. Moreover, the two surveys had different interviewers, and the phone survey had a bigger percentage of interviews conducted by female enumerators (69%) compared to the face-to-face survey (46%). We employ matching in order to address these imbalances across the two surveys and increase our confidence that any substantive differences we find are driven by mode rather than differences between the samples. This allows us to better approximate an experimental design than the typical approach to comparing independent samples in which researchers simply assume random assignment to mode.

We employ coarsened exact matching in our analysis. This approach temporarily coarsens the data to produce an exact match between samples and then produces survey weights to incorporate into data analysis. This method has some advantages over exact matching, which often produces very few matches, and Mahalanobis distance or propensity score matching, which requires setting the matching *ex ante* and checking for balance *ex post*, resulting in multiple specifications until the user obtains balance (Blackwell *et al.*, 2009). We use many-to-one matching with replacement in order to avoid unnecessarily dropping observations. One-to-one matching produces treatment and control groups of the same size at the cost of substantially reducing the matched sample—a problem that is particularly acute when one sample is much larger than the other, as in our case. However, our substantive results are robust to one-to-one matching without replacement as well (see Tables A5 and A6).⁴

3. For reference, Table A1 also compares both samples to population benchmarks, either from the 2018 National Household Survey (ENAH, *Encuesta Nacional de Hogares*) fielded annually by the National Institute of Statistics and Census or the 2011 Census.

4. The exception to this is the result on same-sex marriage, which shrinks somewhat. This is likely due to the process of randomly pruning observations from the treatment and control groups, combined with the large reduction in sample size (from 1,716 to 1,092 observations). At the same time, using a different seed value produced a larger, statistically significant coefficient even with the smaller sample, underscoring the instability of one-to-one matching and reinforcing our preferred approach.

To account for demographic and geographic differences between the samples, we matched on gender, education, age group, residence in San José and Puntarenas provinces, ownership of household items that proxy for wealth (computer/tablet ownership, landline ownership, and internet access), and interviewer gender. Although this is not an exhaustive set of variables, they substantially reduce the differences between the samples (see Table A2). We do lose some observations in the process (416 face-to-face observations and 89 phone observations), but the multivariate L1 distance which measures the global imbalance drops from 0.53 to 0.27 and now there are no differences between the samples on these variables.

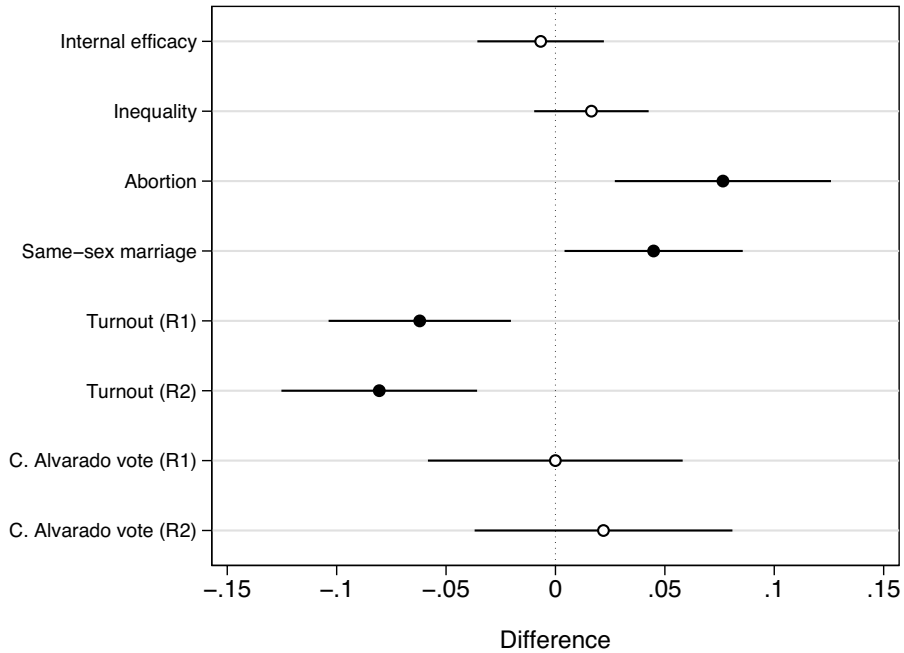
We compare the two samples by conducting difference-in-proportions tests (for binary variables) or difference-in-means tests (for ordinal variables). First, we assess differences in substantive responses between the two modes for both ostensibly sensitive and non-sensitive questions. To do this, we recode our ordinal and categorical variables. For vote choice, we create a dummy variable for those who said they voted in each round of the election for the eventual winner, Carlos Alvarado. Next, we evaluate differences across modes in endpoint responding by recoding each of our items with an ordinal response scale (internal efficacy, inequality, and same-sex marriage) into dummy variables that identify respondents who chose either the lowest or the highest value (figura 1).

FINDINGS

We first want to know whether survey mode affected how people responded, particularly when it comes to sensitive items or those subject to social desirability. Figure 1 plots the difference-in-proportions (for binary variables) and difference-in-means (for ordinal variables) between our face-to-face and phone survey samples. Here we do see a clear distinction between sensitive and non-sensitive items. The items we expected not to be sensitive (internal efficacy and inequality) show no statistically significant differences between the two survey modes.

Where we do see differences is for the sensitive items on abortion and same-sex marriage, and for the turnout questions that are subject to social desirability bias. On the two sensitive items, we see that Costa Rican respondents are more likely to support these contentious social policies in the face-to-face survey than in the phone survey, consistent with the findings in prior work that respondents are more forthcoming about sensitive topics in face-to-face surveys. We see similar patterns with turnout: the phone mode, where prior studies suggest social desirability bias will be higher, yields more over-reporting of electoral participation. Although some scholars consider our items on vote choice potentially sensitive, we find no differences on this item between survey modes. This could be

Figure 1. Comparing substantive responses across modes



Note: Dots represent estimated difference-in-proportions or difference-in-means for each item, comparing the face-to-face and phone surveys. Negative values indicate higher rates in phone as compared to face-to-face. Lines represent 95 % confidence intervals. Black circles represent differences that are statistically significant, open circles those that are not. Rates for each sample are provided in Table A3.

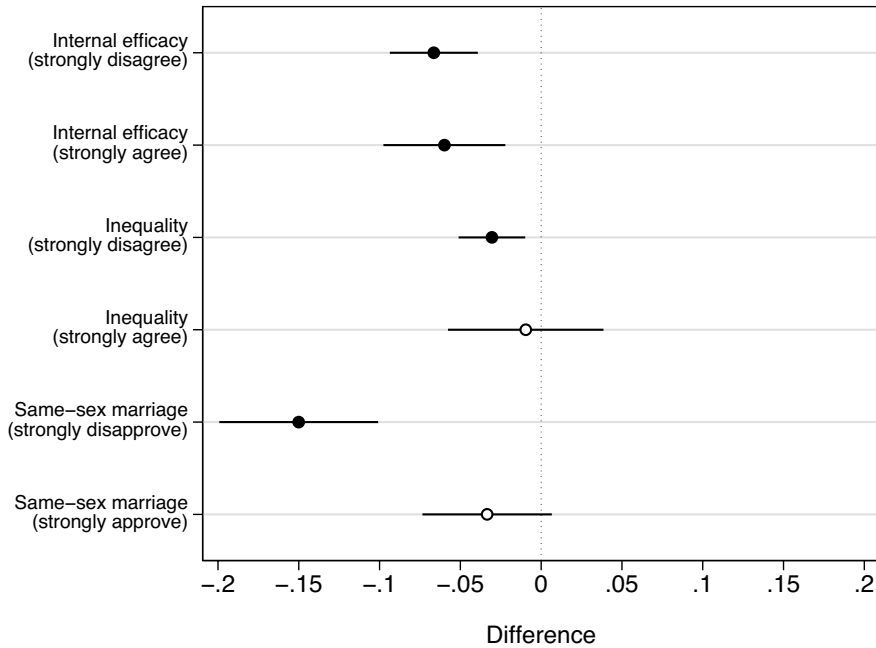
Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey

specific to this particular election, which was won overwhelmingly by Carlos Alvarado, making this question less sensitive. In addition, the question was asked 6-8 months after the election, and 3-5 months after Carlos Alvarado was sworn in as president. In the context of a more contested election, where such an item is more sensitive, we might expect to see mode effects like those we observe with our more sensitive questions.

Next, we want to examine whether our phone survey in Costa Rica yielded higher rates of responses at the extreme ends of scales, as previous work has found in other contexts. Here we focus only on the three items with ordinal response scales: internal efficacy, inequality, and same-sex marriage. In each case, the scale values are only labeled for respondents at the endpoints, but it should be highlighted that the face-to-face mode included visual aids. Figure 2 presents

the results of difference-in-proportions tests comparing the face-to-face mode to the phone mode for each endpoint of each of these scales.

Figure 2. Comparing endpoint responses across modes



Note: Dots represent estimated differences-in-proportions of endpoint responses for each item, comparing the face-to-face and phone surveys. Negative values indicate higher rates in phone as compared to face-to-face. Lines represent 95% confidence intervals. Black circles represent differences that are statistically significant, open circles those that are not. Rates for each sample are provided in Table A4.

Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey

The results indeed suggest that respondents in this context are more prone to choosing endpoint responses in the phone mode. But they further suggest that this bias tends to skew toward more extreme negative responses than toward extreme positive responses. That is, the underlying cause may be primacy bias—the response options respondents were exposed to first—rather than recency bias—the options they heard last. This primacy bias runs counter to some prior work showing that primacy effects are associated with self-administered and visual presentation of answers (Bowling, 2005; Dillman and Christian, 2005), and that phone surveys tend to have a higher incidence of recency effects. Our findings

suggest that the face-to-face survey with visual show cards had a significantly lower rate of primacy bias than the phone survey.⁵

DISCUSSION

While a substantial literature in survey methods finds compelling evidence of mode effects when comparing face-to-face and phone surveys, much of the work compares two independent samples without taking into consideration the differences in the composition of the samples or interviews. Experimental work that randomly assigns mode can improve causal estimates of the differences in response patterns, but typically has to rely on convenience samples that limit generalizability. By using coarsened exact matching, our analysis better approximates an experiment by artificially creating samples that are comparable on geography, demographics, and interviewer characteristics. At the same time, our findings have external validity because they uncover differences in response patterns across two probability-based national samples. Moreover, using a more rigorous method of assessing mode differences, we are able to examine whether mode differences in the literature travel to developing countries by comparing surveys in Costa Rica.

Our findings suggest that some of the conventional wisdom about mode effects obtain in our context as well. Sensitive items and those subject to social desirability bias, like turnout, do exhibit the usual pattern: respondents are more likely to choose the more socially acceptable response over the phone than they are face-to-face. We also find fairly consistent evidence of extreme responses, but they tend to be at the negative extreme of the scale—in contrast to a number of prior studies.

These results suggest that researchers working in the developing world should think carefully before adopting mixed-mode approaches to their surveys. Changing modes over the course of longitudinal surveys or combining data from surveys collected through different modes may introduce biases that subsequently affect

5. Homola, Jackson, and Gill (2016) suggest using entropy scores to compare the dispersion of responses across modes. Using their approach, we get fairly similar entropy scores for internal efficacy (1.83 for the face-to-face survey and 1.88 for the phone survey) and inequality (1.30 for the face-to-face survey and 1.34 for the phone survey) across the two modes, with slightly higher values in the phone survey. The same-sex marriage item yields a higher entropy score in the face-to-face survey (1.85) than in the phone survey (1.50), suggesting that responses were more spread out when administered in person. We think these results speak to the extent to which respondents utilize the full scale of responses equivalently across modes, which is an important methodological question. But they speak less directly to our focus—drawing on prior studies—on whether respondents are more likely to select an endpoint response in some modes.

inferences (see also de Leeuw, 2018; Dillman, 2009). Researchers should also be careful when studying changes over time in repeated cross-sectional surveys where the mode has changed.

Nevertheless, our findings do not suggest that face-to-face surveys are always more desirable in a context like Costa Rica. Phone surveys have some clear advantages over face-to-face surveys and while our study points to potential disadvantages when it comes to measurement, there are also tools to address these issues. Scholars have developed a variety of techniques to elicit more truthful responses to sensitive questions, including the item-count technique, randomized response method, endorsement experiments, and numerous others (e.g., Blair *et al.*, 2020; Cassell and Cohen Forthcoming; Holbrook and Krosnick, 2010; Rosenfeld *et al.*, 2015). In cellphone surveys like ours, researchers could consider sending respondents visual aids for numeric scales via SMS to parallel the face-to-face experience. Or they might use branching rather than large numeric scales to improve accuracy and reduce endpoint responses (Malhotra, Krosnick, and Thomas, 2009; but see Gilbert, 2015). Finally, researchers fielding phone surveys could label the midpoint of numeric response scales, a practice that has been shown to reduce nonresponse (Courser and Lavrakas, 2012).

Our study also has its own limitations. While the ideal design would assign respondents to mode after sampling (see, e.g., Gooch and Vavreck, 2019), our study compares two independent samples. This raises the possibility that we might confound sampling effects with mode effects, and while we address this by matching on observables, that solution is imperfect as there could still be unobserved effects of sampling. Moreover, even though the two surveys were primarily asking about political issues, they did use different questionnaires. While we were able to place some identically worded items on both, our items were embedded among others, which could impact our comparisons (Schwarz *et al.*, 1991).

Still, the comparisons we make across survey modes in Costa Rica are instructive. Placing identically worded questions on two nationally representative surveys fielded very close together in combination with matching offers some advantages over experimental convenience samples. Using this methodology, we have been able to approximate the effect of survey mode on response patterns for sensitive and scalar survey items in a very different context compared to the majority of studies in predominantly affluent countries. This is of course only one such context among hundreds of other developing countries around the world. As more and more people around the developing world gain access to phones, we hope other researchers undertake similar studies of survey mode effects. Only with additional studies across a variety of contexts will we be able to know how much of our conventional wisdom about mode effects holds generally and how much is in fact circumscribed to particular social, cultural, or political contexts.

REFERENCES

- Adida, Claire, Jessica Gottlieb, Eric Kramon, and Gwyneth McClendon. 2019. «Response Bias in Survey Measures of Voter Behavior: Implications for Measurement and Inference.» *Journal of Experimental Political Science* 6(3): 192-8. <https://doi.org/10.1017/XPS.2019.9>
- Alfaro Córdoba, Marcela, María Paula Villareal Galera, and Gustavo Navarro Cerdas. 2010. «Comparación Metodológica de Encuestas vía Telefonía Fija y Celular en Costa Rica.» *Ciencias Económicas* 28(1): 51-68. <https://doi.org/10.15517/rce.v28i1.7064>
- Alfaro-Redondo, Ronald, and Felipe Alpízar Rodríguez, eds. 2020. *Elecciones 2018 en Costa Rica: Retrato de una Democracia Amenazada*. San José: Programa Estado de la Nación.
- Alfaro-Redondo, Ronald, and Steffan Gómez-Campos. 2021. «PLN and PAC: Two Costa Rican Parties with Constituencies Evolving in Opposite Directions.» In *Diminished Parties: Democratic Representation in Contemporary Latin America*, eds. Juan Pablo Luna, Rafael Piñeiro Rodríguez, Fernando Rosenblatt, and Gabriel Vommaro. New York: Cambridge University Press, pp. 111-28.
- Aquilino, William S. 1991. «Telephone Versus Face-to-Face Interviewing for Household Drug Use Surveys.» *International Journal of the Addictions* 27(1): 71-91. <https://doi.org/10.3109/10826089109063463>
- Aquilino, William S. 1994. «Interview Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment.» *Public Opinion Quarterly* 58: 210-40. <https://doi.org/10.1086/269419>
- Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. «cem: Coarsened Exact Matching in Stata.» *The Stata Journal* 9(4): 524-46. <https://doi.org/10.1177/1536867X0900900402>
- Blair, Graeme, Alexander Coppock, and Margaret Moor. 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): 1297-315. <https://doi.org/10.1017/S0003055420000374>
- Bowling, Ann. 2005. «Mode of Questionnaire Administration Can Have Serious Effects on Data Quality.» *Journal of Public Health* 27(3): 281-91. <https://doi.org/10.1093/pubmed/fdi031>
- Cassell, Kaitlen J., and Mollie J. Cohen. Forthcoming. «Reducing Item Non-Response to Vote Choice Questions: Evidence from a Survey Experiment in Mexico.» *Public Opinion Quarterly*. Available at <http://nebula.wsimg.com/4f954acc79f53e6189906a4b95ca82ab?AccessKeyId=6C80BA4D3A5A4F2D4389&disposition=0&alloworigin=1>
- Castorena, Oscar, Noam Lupu, Maita Schade, and Elizabeth J. Zechmeister. Forthcoming. «Online Surveys in Latin America.» *PS: Political Science & Politics*. Available at https://noamlupu.com/online_samples.pdf.
- Christian, Leah Melani, Nicholas L. Parsons, and Don A. Dillman. 2009. «Designing Scalar Questions for Web Surveys.» *Sociological Methods & Research* 37(3): 393-425. <https://doi.org/10.1177/0049124108330004>
- Colburn, Forrest D., and Arturo Cruz S. 2018. «Latin America's Shifting Politics: The Fading of Costa Rica's Old Parties.» *Journal of Democracy* 29(4): 43-53. <https://doi.org/10.1353/jod.2018.0061>

- Courser, Matthew, and Paul J. Lavrakas. 2012. «Item-Nonresponse and the 10-Point Response Scale in Telephone Surveys.» *Survey Practice* 5(4): 1-7. <https://doi.org/10.29115/SP-2012-0021>
- Dahlberg, Stefan, and Mikael Persson. 2014. «Different Surveys, Different Results? A Comparison of Two Surveys on the 2009 European Parliamentary Election.» *West European Politics* 37(1): 204-221. <https://doi.org/10.1080/01402382.2013.814961>
- de Leeuw, Edith D. 2018. «Mixed-Mode: Past, Present, and Future.» *Survey Research Methods* 12(2): 75-89. <https://doi.org/10.18148/srm/2018.v12i2.7402>
- Díaz González, José Andrés, and Stephanie Cordero Cordero. 2020. «Las Preferencias del Electorado en la Segunda Ronda Presidencial de 2018 en Costa Rica: Un Modelo de Socialización Política.» *Política y Gobierno* 27(1): 1-22. https://www.scielo.org.mx/scielo.php?pid=S1665-20372020000100200&script=sci_arttext
- Dillman, Don A. 2009. «Some Consequences of Survey Mode Changes in Longitudinal Surveys.» In *Methodology of Longitudinal Surveys*, ed. Peter Lynn. London: Wiley, pp. 127-40.
- Dillman, Don A., and Leah Melani Christian. 2005. «Survey Mode as a Source of Instability in Responses across Surveys.» *Field Methods* 17(1): 30-52. <https://doi.org/10.1177/1525822X04269550>
- Dillman, Don A., Glenn Phelps, Robert Tortora, Karen Swift, Julie Kohrell, Jodi Berck, and Benjamin L. Messer. 2009. «Response Rate and Measurement Differences in Mixed-Mode Surveys Using Mail, Telephone, Interactive Voice Response (IVR) and the Internet.» *Social Science Research* 38(1): 1-18. <https://doi.org/10.1016/j.ssresearch.2008.03.007>
- Gibson, Dustin G. Amanda Pereira, Brooke A. Farrenkopf, Alain B. Labrique, George W. Pariyo, and Adnan A. Hyder. 2017. «Mobile Phone Surveys for Collecting Population-Level Estimates in Low- and Middle-Income Countries: A Literature Review.» *Journal of Medical Internet Research* 19(5): 1-12. <https://doi.org/10.2196/jmir.7428>
- Gilbert, Emily E. 2015. «A Comparison of Branched Versus Unbranched Rating Scales for the Measurement of Attitudes in Surveys.» *Public Opinion Quarterly* 79(2): 443-70. <https://doi.org/10.1093/poq/nfu090>
- Gooch, Andrew, and Lynn Vavreck. 2019. «How Face-to-Face Interviews and Cognitive Skill Affect Item Non-Response: A Randomized Experiment Assigning Mode of Interview.» *Political Science Research and Methods* 7(1): 143-62. <https://doi.org/10.1017/psrm.2016.20>
- Greenfield, Thomas K., Lorrain T. Midanik, and John D. Rogers. 2000. «Effects of Telephone versus Face-to-Face Interview Modes on Reports of Alcohol Consumption.» *Addiction* 95(2). <https://doi.org/10.1046/j.1360-0443.2000.95227714.x>
- Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. «Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias.» *Public Opinion Quarterly* 67(1): 79-125. <https://doi.org/10.1086/346010>
- Holbrook, Allyson L., and Jon A. Krosnick. 2010. «Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique.» *Public Opinion Quarterly* 74(1): 37-67. <https://doi.org/10.1093/poq/nfp065>
- Homola, Jonathan, Natalie Jackson, and Jeff Gill. 2016. «A Measure of Survey Mode Differences.» *Electoral Studies* 44: 255-74. <https://doi.org/10.1016/j.electstud.2016.06.010>

- Hox, Joop J., and Edith D. de Leeuw. 1994. «A Comparison of Nonresponse in Mail, Telephone, and Face-to-Face Surveys: Applying Multilevel Modeling to Meta-Analysis.» *Quality & Quantity* 28: 329-44. <https://doi.org/10.1007/BF01097014>
- Jäckle, Annette, Caroline Roberts and Peter Lynn. 2010. «Assessing the Effect of Data Collection Mode on Measurement.» *International Statistical Review* 78(1): 3-20. <https://doi.org/10.1111/j.1751-5823.2010.00102.x>
- Karp, Jeffrey A., and David Brockington. 2005. «Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries.» *Journal of Politics* 67(3):825-40. <https://doi.org/10.1111/j.1468-2508.2005.00341.x>
- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. «Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity.» *Public Opinion Quarterly* 72(5): 847-65. <https://doi.org/10.1093/poq/nfn063>
- Lupu, Noam, and Kristin Michelitch. 2018. «Advances in Survey Methods for the Developing World.» *Annual Review of Political Science* 21: 195-214. <https://doi.org/10.1146/annurev-polisci-052115-021432>
- Lupu, Noam, J. Daniel Montalvo, Mitchell A. Seligson, Elizabeth J. Zechmeister, and Kirill Zhirkov. 2022. «Comparing Two Common Approaches to Within-Household Sampling: A Field Experiment in Costa Rica.» Unpublished manuscript. Available at https://noam-lupu.com/FM_vs_LB.pdf
- Malhotra, Neil, Jon A. Krosnick, and Randall K. Thomas. 2009. «Optimal Design of Branching Questions to Measure Bipolar Constructs.» *Public Opinion Quarterly* 73(2): 304-24. <https://doi.org/10.1093/poq/nfp023>
- Meisenberg, Gerhard, and Amandy Williams. 2008. «Are Acquiescent and Extreme Response Styles Related to Low Intelligence and Education?» *Personality and Individual Differences* 44(7): 1539-50. <https://doi.org/10.1016/j.paid.2008.01.010>
- Montalvo, J. Daniel, Georgina Pizzolitto, and Luke Plutowski. 2022. «Telephone Sampling in the 2021 AmericasBarometer.» LAPOP Methodological Note No. 009. Available at <https://www.vanderbilt.edu/lapop/insights/IMN009en.pdf>
- Perelló, Lucas, and Patricio Navia. 2021. «Abrupt and Gradual Realignment: The Case of Costa Rica, 1958-2018.» *Journal of Politics in Latin America* 13(1): 86-113. <https://doi.org/10.1177/1866802X20967733>
- Pignataro, Adrián, and Ilka Treminio. 2019. «Reto Económico, Valores y Religión en las Elecciones Nacionales de Costa Rica 2018.» *Revista de Ciencia Política* 39(2): 239-64. <http://dx.doi.org/10.4067/S0718-090X2019000200239>
- Roberts, Caroline, and Annette Jäckle. 2012. «Causes of Mode Effects: Separating Out Interviewer and Stimulus Effects in Comparisons of Face-to-Face and Telephone Surveys.» ISER Working Paper No. 2012-27. Available at <https://www.econstor.eu/handle/10419/91682>
- Rosenfeld, Bryn, Kosuke Imai, and Jacob N. Shapiro. 2016. «An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions.» *American Journal of Political Science* 60(3): 783-802. <https://doi.org/10.1111/ajps.12205>
- Schwarz, Norbert, Fritz Strack, Hans-J. Hippler, and George Bishop. 1991. «The Impact of Administration Mode on Response Effects in Survey Measurement.» *Applied Cognitive Psychology* 5: 193-212. <https://doi.org/10.1002/acp.2350050304>

- Seligson, Mitchell A. 2020. «The Seligson Political Culture Survey Archive: Costa Rica, 1973-2005.» LAPOP Insights Series No. 143. <https://www.vanderbilt.edu/lapop/insights/IO943en.pdf>
- Shino, Enrijeta, and Michael D. Martinez. 2021. «The Different Faces of Public Opinion: Is the American Voter Tinted by Mode?» *International Journal of Public Opinion Research* 33(4): 756-78. <https://doi.org/10.1093/ijpor/edab003>
- Vannieuwenhuyze, Jorre, Geert Loosveldt, and Geert Molenberghs. 2010. «A Method for Evaluating Mode Effects in Mixed-Mode Surveys.» *Public Opinion Quarterly* 74(5) 1027-45. <https://doi.org/10.1093/poq/nfq059>

APPENDIX A

Question Wording

Internal efficacy:

Usted siente que entiende bien los asuntos políticos más importantes del país. ¿Hasta qué punto está de acuerdo o en desacuerdo con esta frase? (1) Muy en desacuerdo – (7) Muy de acuerdo

You feel that you understand the most important political issues of this country. How much do you agree or disagree with this statement? (1) Strongly disagree – (7) Strongly agree

Government should reduce inequality

El Estado costarricense debe implementar políticas firmes para reducir la desigualdad de ingresos entre ricos y pobres. ¿Hasta qué punto está de acuerdo o en desacuerdo con esta frase? (1) Muy en desacuerdo – (7) Muy de acuerdo

The Costa Rican government should implement strong policies to reduce income inequality between the rich and the poor. To what extent do you agree or disagree with this statement? (1) Strongly disagree – (7) Strongly agree

Same-sex marriage

¿Con qué firmeza aprueba o desaprueba que las parejas del mismo sexo puedan tener el derecho a casarse? (1) Desaprueba firmemente – (10) Aprueba firmemente

How strongly do you approve or disapprove of same-sex couples having the right to marry? (1) Strongly disapprove – (10) Strongly approve

Abortion when mother's health at risk

¿Cree usted que se justificaría la interrupción del embarazo, o sea, un aborto, cuando peligr la salud de la madre? (1) Sí, (0) No

Do you think it's justified to interrupt a pregnancy, that is, to have an abortion, when the mother's health is in danger? (1) Yes, (0) No

Turnout (round 1)

Face-to-face: ¿Votó usted en la primera ronda de las últimas elecciones presidenciales de 2018? (1) Sí, (0) No

Phone: ¿Votó usted en las elecciones de febrero de 2018 (primera ronda)? (1) Sí votó, (2) No votó

Did you vote in the first round of the last presidential elections of 2018? (1) Yes, (0) No

Turnout (round 2)

Face-to-face: *Votó usted en la segunda ronda de las últimas elecciones presidenciales de 2018?* (1) Sí, (0) No

Phone: *¿Votó usted en la s elecciones de abril de 2018 (segunda ronda)?* (1) Sí votó, (2) No votó

Did you vote in the second round of the last presidential elections of 2018? (1) Yes, (0) No

Vote choice (round 1)

Face-to-face: *¿Por quién votó para Presidente en la primera ronda de las últimas elecciones presidenciales de 2018?* (0) Ninguno (fue a votar pero dejó la boleta en blanco), (97) Ninguno (anuló su voto), (601) Carlos Alvarado Quesada (PAC), (602) Fabricio Alvarado Muñoz (PRN), (603) Antonio Álvarez Desanti (PLN), (604) Rodolfo Piza Rocafort (PUSC), (605) Juan Diego Castro Fernández (PIN), (606) Rodolfo Hernández Gómez (PRSC), (607) Otto Guevara Guth (ML), (608) Edgardo Araya Sibaja (FA), (677) Otro

Who did you vote for in the first round of the last presidential election of 2018? (0) No one (blank vote), (97) No one (null vote), (601) Carlos Alvarado Quesada (PAC), (602) Fabricio Alvarado Muñoz (PRN), (603) Antonio Álvarez Desanti (PLN), (604) Rodolfo Piza Rocafort (PUSC), (605) Juan Diego Castro Fernández (PIN), (606) Rodolfo Hernández Gómez (PRSC), (607) Otto Guevara Guth (ML), (608) Edgardo Araya Sibaja (FA), (677) Other

Phone: *¿Por quién votó usted?* (0) No votó, (97) Ninguno (anuló el voto), (1) Rodolfo Piza (PUSC), (2) Antonio Álvarez (PLN), (3) Carlos Alvarado (PAC), (4) Otto Guevara (ML), (5) Edgardo Araya (FA), (6) Rodolfo Hernández (RSC), (7) Juan Diego Castro (PIN), (8) Sergio Mena (NG), (9) John Vega (PT), (11) Fabricio Alvarado (RN), (10) Otro

For whom did you vote? (0) Didn't Vote, (97) No one (cast a null ballot), (1) Rodolfo Piza (PUSC), (2) Antonio Álvarez (PLN), (3) Carlos Alvarado (PAC), (4) Otto Guevara (ML), (5) Edgardo Araya (FA), (6) Rodolfo Hernández (RSC), (7) Juan Diego Castro (PIN), (8) Sergio Mena (NG), (9) John Vega (PT), (11) Fabricio Alvarado (RN), (10) Other

Vote choice (round 2)

Face-to-face: *¿Y por quién votó para Presidente en la segunda vuelta de las elecciones presidenciales del 2018?* (0) Ninguno (fue a votar pero dejó la boleta en blanco), (97) Ninguno (anuló su voto), (601) Carlos Alvarado Quesada (PAC), (602) Fabricio Alvarado Muñoz (PRN)

For whom did you vote in the second round of the presidential elections in 2018? (0) No one (blank vote), (97) No one (null vote), (601) Carlos Alvarado Quesada (PAC), (602) Fabricio Alvarado Muñoz (PRN)

Phone: ¿Por quién votó usted el pasado 01 de abril? (3) Carlos Alvarado, (11) Fabricio Alvarado, (96) Se abstuvo, (97) Anuló el voto, (98) Votó en blanco
For whom did you vote on April 1st? (3) Carlos Alvarado, (11) Fabricio Alvarado, (96) Abstained, (97) Null vote, (98) Cast a blank vote

APPENDIX B

Additional Tables

Table A1. Comparing demographic distributions

Variable	Face-to-Face Sample	Phone Sample	ENAH0 2018
Gender			
Female	50.0%	52.9%	52.3%
Male	50.0%	47.1%	47.7%
Age group			
18-25	22.1%	21.7%	18.1%
26-35	25.7%	24.3%	20.4%
36-45	16.4%	19.4%	17.9%
46-55	14.2%	16.9%	16.6%
56-65	11.1%	10.6%	14.0%
65+	10.5%	7.1%	13.0%
Education level			
Primary or less	32.3%	27.5%	41.8%
Secondary	47.9%	38.8%	39.1%
Tertiary	19.8%	33.8%	19.1%
Household assets			
Computer	52.8%	56.8%	48.9%
Landline phone	33.0%	28.4%	
Internet access	67.9%	69.8%	75.6%

NOAM LUPU AND ADAM WOLSKY
SURVEY MODE EFFECTS IN A DEVELOPING COUNTRY

Variable	Face-to-Face Sample	Phone Sample	ENAH0 2018
Province			Census 2011
San José	40.0%	32.8%	32.7%
Alajuela	17.3%	20.4%	20.0%
Cartago	10.0%	10.7%	10.7%
Heredia	9.9%	9.7%	10.2%
Guanacaste	6.4%	7.9%	7.7%
Puntarenas	6.3%	9.5%	9.7%
Limón	9.9%	8.9%	9.0%

Note: Bolded values identify statistically significant ($p < 0.05$) differences between the two survey samples. Italicized values identify statistically significant ($p < 0.05$) differences between a survey and the benchmark (ENAH0 or Census).

Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey

Table A2: Comparing distributions with and without matching

	Face-to-Face Sample (No Matching)	Phone Sample (No Matching)	Face-to-Face Sample (CEM Weights)	Phone Sample (CEM Weights)
College	0.20 (0.01)	0.34 (0.02)	0.31 (0.01)	0.31 (0.02)
65+	0.10 (0.01)	0.07 (0.01)	0.06 (0.01)	0.06 (0.01)
Female	0.50 (0.01)	0.53 (0.02)	0.52 (0.02)	0.52 (0.02)
Landline	0.33 (0.01)	0.28 (0.02)	0.27 (0.01)	0.27 (0.02)
Computer	0.53 (0.01)	0.57 (0.02)	0.58 (0.02)	0.58 (0.02)

NOAM LUPU AND ADAM WOLSKY
SURVEY MODE EFFECTS IN A DEVELOPING COUNTRY

	Face-to-Face Sample (No Matching)	Phone Sample (No Matching)	Face-to-Face Sample (CEM Weights)	Phone Sample (CEM Weights)
Internet	0.68 (0.01)	0.70 (0.02)	0.73 (0.01)	0.73 (0.02)
San Jose	0.40 (0.01)	0.33 (0.02)	0.34 (0.01)	0.34 (0.02)
Puntarenas	0.06 (0.01)	0.09 (0.01)	0.05 (0.01)	0.05 (0.01)
Female	0.46 (0.01)	0.69 (0.02)	0.69 (0.01)	0.69 (0.02)
Interviewer	(0.01)	(0.02)	(0.01)	(0.02)
<i>N</i>	1,501	720	1,085	631

Note: Standard errors in parentheses.

Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey

Table A3. Comparing substantive responses across modes

Variable	Face-to-Face Sample	Phone Sample	Difference
Internal efficacy	0.609 (0.008)	0.615 (0.012)	0.007 (0.015)
Inequality	0.855 (0.007)	0.838 (0.011)	0.016 (0.013)
Abortion	0.659 (0.015)	0.582 (0.020)	0.077 (0.025)
Same-sex marriage	0.380 (0.012)	0.336 (0.017)	0.045 (0.021)
Turnout (R1)	0.724 (0.014)	0.786 (0.016)	-0.062 (0.021)
Turnout (R2)	0.653	0.734	-0.081

NOAM LUPU AND ADAM WOLSKY
SURVEY MODE EFFECTS IN A DEVELOPING COUNTRY

Variable	Face-to-Face Sample	Phone Sample	Difference
	(0.014)	(0.018)	(0.023)
C. Alvarado vote (R1)	0.385	0.385	-0.000
	(0.019)	(0.023)	(0.029)
C. Alvarado vote (R2)	0.658	0.636	0.022
	(0.019)	(0.024)	(0.030)

Note: Standard errors in parentheses. Bolded differences are statistically significant ($p < 0.05$).

Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey

Table A4. Comparing endpoint responses across modes

Variable	Face-to-Face Sample	Phone Sample	Difference
Internal efficacy	0.042	0.108	-0.066
(strongly disagree)	(0.006)	(0.012)	(0.013)
Internal efficacy	0.140	0.200	-0.059
(strongly agree)	(0.011)	(0.016)	(0.019)
Inequality	0.026	0.057	-0.030
(strongly disagree)	(0.005)	(0.009)	(0.011)
Inequality	0.613	0.622	-0.010
(strongly agree)	(0.015)	(0.020)	(0.025)
Same-sex marriage	0.397	0.547	-0.150
(strongly disapprove)	(0.015)	(0.020)	(0.025)
Same-sex marriage	0.183	0.217	-0.033
(strongly approve)	(0.012)	(0.017)	(0.020)

Note: Standard errors in parentheses. Bolded differences are statistically significant ($p < 0.05$).

Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey

Table A5. Comparing substantive responses across modes (one-to-one matching)

Variable	Face-to-Face Sample	Phone Sample	Difference
INTERNAL EFFICACY	0.610 (0.012)	0.620 (0.013)	-0.010 (0.017)
INEQUALITY	0.844 (0.011)	0.842 (0.012)	0.003 (0.016)
ABORTION	0.658 (0.021)	0.579 (0.022)	0.079 (0.030)
SAME-SEX MARRIAGE	0.367 (0.017)	0.331 (0.018)	0.035 (0.025)
TURNOUT (R1)	0.705 (0.020)	0.775 (0.018)	-0.070 (0.026)
TURNOUT (R2)	0.629 (0.021)	0.723 (0.019)	-0.095 (0.028)
C. ALVARADO VOTE (R1)	0.398 (0.027)	0.389 (0.025)	0.009 (0.037)
C. ALVARADO VOTE (R2)	0.633 (0.027)	0.633 (0.025)	0.001 (0.037)

Note: Standard errors in parentheses. Bolded differences are statistically significant ($p < 0.05$).

Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey

Table A6. Comparing endpoint responses across modes (one-to-one matching)

Variable	Face-to-Face Sample	Phone Sample	Difference
Internal efficacy	0.040	0.097	-0.057
(strongly disagree)	(0.008)	(0.013)	(0.015)
Internal efficacy	0.139	0.211	-0.072
(strongly agree)	(0.015)	(0.018)	(0.023)
Inequality	0.029	0.056	-0.027
(strongly disagree)	(0.007)	(0.010)	(0.012)
Inequality	0.607	0.627	-0.020
(strongly agree)	(0.021)	(0.021)	(0.030)
Same-sex marriage	0.412	0.552	-0.139
(strongly disapprove)	(0.021)	(0.022)	(0.030)
Same-sex marriage	0.174	0.211	-0.037
(strongly approve)	(0.016)	(0.018)	(0.024)

Note: Standard errors in parentheses. Bolded differences are statistically significant ($p < 0.05$).

Source: 2018 AmericasBarometer, University of Costa Rica 2018 phone survey