EFFECTS OF INTERVIEWER GENDER ON SURVEY RESPONSES: FINDINGS FROM A HOUSEHOLD SURVEY IN MEXICO

Francisco Flores-Macías and Chappell Lawson

In the social sciences, most studies on gender-of-interviewer effects have focused on questions dealing with sexual behavior and gender roles, as well as on response rates. Results from these analyses have been mixed, sometimes showing little evidence of effects (Johnson & DeLamater, 1976; Johnson & Moore, 1993) and other times suggesting non-trivial biases (Landis, Sullivan & Sheley, 1973; Galla, Frisone, Jeffrey, & Gaer, 1981; Lueptow, Moser, & Pendleton, 1990; Catania et al., 1996; Tu & Liao, 2007). In addition, response effects have not always been found in the same direction. While Galla et al. (1981) argue that more 'non-traditional' responses occur for both male and female respondents only when facing a female interviewer, Lueptow et al. (1990) and Huddy et al. (1997) sustain that effects are stronger among female respondents who are interviewed by another female, and Kane and Macaulay (1993) claim that 'more egalitarian' answers are more pervasive among male respondents interviewed by a woman.

Only a few studies have explored the existence of bias attributable to the gender of the interviewer on political questions. Although these analyses have provided invaluable preliminary evidence on gender-of-interviewer effects, they have failed to yield statistically significant effects on the interaction between the gender of the interviewer and the gender of the respondent (Kane & Macaulay, 1993; Huddy et al., 1997). The one exception is the early study by Landis et al. (1973), who used a sample of 90 female college students in a town in the Western United States and found statistically significant evidence that female interviewers elicited more feminist responses.

Using a nationally representative telephone survey in the United States, Kane and Macaulay (1993) analyzed potential interviewer-gender effects for a variety of questions, including some dealing with public policy on women's rights. They found that, although responses to some items appeared to be influenced by the gender of the

interviewer, interaction effects between the gender of the interviewer and the gender of the respondent were not significant. Thus 'despite the tendency for bivariate interviewer-gender effects to be evident for different types of items for male and female respondents, and for the interviewer-gender/respondent-gender interactions to be in the expected directions, [the] multivariate analyses indicate that these interactions do not tend to be statistically significant in [the] sample' (p. 22).

Huddy et al. (1997), using two telephone surveys in selected areas of upstate New York, test for interviewer-gender effects on questions dealing with gender equality, activism of women, abortion, affirmative action, and on opinions on female political figures, among others. Out of the eleven items tested, the interaction between interviewer and respondent gender is significant only for one (dealing with the desirability of collective action by women in order to improve their conditions). In all, then, past research has called attention to the issue of interviewer gender but has yet to document large or significant effects that would be relevant for most political science research.

Our findings shed light on several unresolved controversies in survey research. First, they document effects that have long seemed conceptually plausible but had proven empirically elusive. Second, they underscore that gender-of-interviewer effects are likely to be asymmetric—that is, more likely to affect men than women on gender-sensitive items. Third, they show that interviewer effects may also depend on the social context in which the surveys take place. In Mexico City—a large, politically progressive metropolis—gender-of-interviewer effects are consistently significant; nationally, they are undetectable. Finally, our findings extend the debate over interviewer-induced bias to countries other than the United States, where research has so far been much less extensive.

DATA AND METHODS

Our analysis draws on the first (October 2005) wave of the Mexico 2006 Panel Study, a face-to-face household survey designed to measure public opinion during Mexico's 2006 presidential campaign. Respondents included 1,600 adults in a national sample, 500 in an over-sample in Mexico City, and 300 in an over-sample of rural areas in the states of Chiapas, Jalisco, and Oaxaca. Participants were selected within the households based on which eligible adult had the most recent birthday. In the national sample, primary sampling units (PSUs) were determined according to a probability-proportional-to-size criterion, while simple random sampling was used to select sites in the Mexico City and rural oversamples. Each PSU provided a quota of 20 interviews, with households selected through a systematic random process followed by the interviewer. The questionnaire included approximately 80 questions, mostly concerning political opinions. The response rate for the survey, according to the minimum rate described in the guidelines of the American Association of Public Opinion Research, was 34 percent, and the sampling margin of error within a 95 percent confidence interval for each question was ±2.5 percent for the national sample, ±4.4 percent for Mexico City, and ±5.7 percent for the rural areas. To ensure high reliability in the data collection process, 33 percent of the interviews were directly supervised at the time of the survey, 19 percent were supervised in person

afterwards, 7 percent were supervised by telephone, and only 41 percent of the interviews went completely unsupervised.

Due primarily to safety concerns, female interviewers were not assigned in the rural over-sample. Consequently, the present analysis focuses on the national and Mexico City samples. The national sample included 1,600 respondents, clustered in 80 PSUs that were assigned to 75 interviewers (38 female and 37 male). The Mexico City sample included 500 respondents, clustered in 25 PSUs assigned to 33 interviewers (17 female and 16 male). In the urban areas of the national sample and the entire Mexico City sample, interviewers were assigned randomly by gender; some female interviewers were assigned to rural portions of the national sample, but their assignment was not consciously random. (The findings reported here hold when only urban areas are analyzed, as well as when both urban and rural areas of the national sample are included in the analysis.)

Although this design was intended to yield a field experiment, randomization proved only partially successful (as shown by chi-squared tests of basic socio-demographic characteristics). Lack of randomization is partly attributable to the fact that a disproportionate number of female interviewers was assigned to particular regions of the country (out of considerations of safety). Another complication was the fact that supervisors themselves were not assigned randomly by gender, and the presence of these individuals could have muddled the experiment. To address this potential source of contamination, we also collected data on the gender of the supervisor.

In the national sample, 527 of the 1,600 interviews were supervised concurrently, and 163 (approximately 10 percent of the total) were cases where the gender of the supervisor did not match the gender of the interviewer. In the Mexico City sample, 68 of the 500 interviews were supervised concurrently, and 49 (approximately 10 percent of the total) were instances where the gender of the interviewer and the gender of the supervisor were different. In all analyses presented here, if an interview was supervised concurrently by a woman, we coded 'interviewer gender' as female, even if the interviewer was male. We believe that this is justified since the mere presence of a female supervisor at the time of the interview, overlooking the process, should be enough to elicit any effects. In any case, we tested for robustness and obtained almost identical results to the ones presented here when conducting all regressions in two additional ways: (i) not considering the cases when there was a mismatch between the gender of the interviewer and the gender of the supervisor and (ii) coding only the gender of the interviewer and disregarding that of the supervisor.

Two questions in the survey instrument are considered gender-sensitive and thus most likely to be influenced by the gender of the interviewer. The first of these asked whether or not abortion should be legal in cases of rape (see Appendix for all question wordings). The second item was part of a battery in which respondents were asked to rate the urgency of different policy priorities for the next president. 'Women's rights' was included among these, mixed with 'poverty', 'corruption', 'education', 'modernization of the electricity sector', 'trade relations between Mexico and the United States', and others. The response options were 'very urgent', 'urgent', and 'not so urgent'.

As with most of the policy priorities mentioned, only a small number of respondents rated women's rights as 'not so urgent'; consequently, both 'urgent' and 'not so urgent' were coded as 0 and 'very urgent' was coded as 1. (Our conclusions do not change if we code the responses on a 1–3 scale and conduct our multivariate analysis using either ordinary least squares or ordered probit.) For the question on abortion, a zero meant that abortion should be illegal in cases of rape and a one that it should be legal. Thus a higher score indicated a more socially progressive response.

RESULTS

Gender-Sensitive Items

With these issues in mind, we first report the bivariate results and then turn to the multivariate analysis, controlling for the main factors that might influence attitudes on gender-sensitive topics. As the data in Table 1 indicate, male respondents in Mexico City appear to have been more susceptible to gender-of-interview effects. For instance, approximately 30 percent of men interviewed by men identified women's rights as a 'very urgent' priority for the next president, compared to 40 percent who were interviewed by women. Likewise, men interviewed by men were about twelve percentage points more likely to favor criminalizing abortion in the case of rape than men interviewed by women. On the other hand, in the national sample there were virtually no differences among male respondents by gender of interviewer. Women appear to become more progressive when interviewed by men in the national sample; in Mexico City, however, they are almost equally progressive for male and female interviewers in the abortion question and less progressive when interviewed by a man in the women's rights question.

We next turn to the multivariate analysis: a binary probit model that controls for socio-demographic variables, as well as for the presence of a third party—other than the supervisor—at the time of the interview. (This third party was generally a family member.) Cluster-robust standard errors are used in order to take into account clustering in the sampling units. The few cases of item non-response are not included in the analysis. Deleting these cases does not affect the results, as shown by a binary probit model analogous to the general model with item non-response as the dependent variable (1 = doesn't know/refused and 0 = valid response). Item non-response to these two questions was not significantly related to the interviewer's gender.

The independent variables are described below (for descriptive statistics of the independent variables, see Table 2):

- *Age* of the respondent, ranging from 17 (for those eligible respondents who would turn 18 prior to the election day) to 92.
- *Educational level* of the respondent, on the following 9-point scale: none, incomplete primary, complete primary, incomplete secondary, complete secondary, incomplete preparatory, complete preparatory, incomplete college, complete college or more.
- Frequency of *church attendance* of the respondent, on a 5-point scale— more than once a week, once a week, once a month, only on special occasions,

TABLE 1 **Responses by gender of interviewer and gender of respondent**

	Male respondents		Female respondents	
	Female interviewer (%)	Male interviewer (%)	Female interviewer (%)	Male interviewer (%)
National sample				
Abortion question				
Legal in case of rape	53	55	56	62
Illegal in case of rape	38	37	36	31
Doesn't know/refused	8	8	9	7
N	450	343	459	348
	$\chi^2 = 0.16$; df $= 2$; $p = .92$		$\chi^2 = 3.21$; df $= 2$; $p = .20$	
Women's rights question				
Very urgent	36	34	43	51
Urgent/not so urgent	62	64	55	48
Doesn't know/refused	2	2	2	1
N	450	343	459	348
	$\chi^2 = 0.33$; df $= 2$; $p = .85$		$\chi^2 = 7.73$; df $= 2$; $p = .06$	
Mexico City sample				
Abortion question				
Legal in case of rape	76	68	78	80
Illegal in case of rape	18	30	17	17
Doesn't know/refused	5	3	5	3
N	93	142	106	159
	$\chi^2 = 4.42$; df $= 2$; $p = .11$		$\chi^2 = 0.43$; df $= 2$; $p = .80$	
Women's rights question				
Very urgent	40	30	42	36
Urgent/not so urgent	60	70	57	63
Doesn't know/refused	–	–	1	1
N	93	142	106	159
	$\chi^2 = 2.62$; df $= 1$; $p = .11$		$\chi^2 = 1.08$; df $= 2$; $p = .58$	

never—so that higher scores indicate less religiosity. This variable is included as control only for the abortion question.

- Interviewer's assessment of the *socioeconomic status* of the respondent's household, on a 5-point scale. (This measure correlates closely with an index of household items—whether the respondent owns a telephone, refrigerator, car, etc.—but offers slightly greater spread. Using one measure rather than the other does not affect our conclusions.)

- *Ideological self-placement* of the respondent on a 7-point scale, ranging from strong right (-3) to strong left (3). Many respondents (35 percent in the national sample and 24 percent in the Mexico City sample) were unable to place themselves on the ideological spectrum and are coded in the middle of the scale. (Treating these respondents as missing does not affect our conclusions.)

TABLE 2 **Descriptive statistics of the control variables (non-responses not included)**

	N	Mean	SD	Minimum	Maximum
National sample					
Age	1599	40.24	16.07	17	92
Education	1593	4.93	2.48	1	9
Church attendance	1579	2.85	1.14	1	5
Socioeconomic status	1598	2.13	0.85	1	5
Ideology	1600	0.03	1.43	−3	3
Third person present	1597	0.39	0.49	0	1
Mexico City sample					
Age	500	40.11	15.71	17	84
Education	500	6.09	2.34	1	9
Church attendance	492	3.09	1.14	1	5
Socioeconomic status	500	2.44	0.89	1	5
Ideology	500	0.28	1.45	−3	3
Third person present	500	0.25	0.43	0	1

- Presence of a *third person*, other than the interviewer, during the interview. Normally, the person in question was another member of the respondent's household.
- Dummy variable for the *gender of the interviewer* (1 = female and 0 = male). As explained above, if the supervisor was female and present during the interview, the interviewer is coded as 'female,' even if the person asking the questions was male.

Once these confounding factors are taken into account, the effect of interviewer gender for male respondents living in Mexico City is highly significant for both questions. In keeping with previous empirical research on the topic, the results are not statistically significant for male respondents in the national sample. Likewise, interviewer gender has no significant effect on female respondents on either question, in either sample. Table 3 presents these results, disaggregated by type of sample and by gender of respondent in order to make the interactions among these dummy variables more clear.

The control variables generally operate in the expected direction. Education is highly significant and positively associated with favoring legal abortion in the event of rape in all the models run; church attendance is also significant in the expected direction in the national sample. The presence of a third party is never significant, and socioeconomic status and ideology are significant at the 5-percent level in only one instance each.

To give a sense of the magnitude of the effect of interviewer gender, we follow the standard practice of holding all coefficients at their means (or modes for binary variables) while manipulating the value of one independent variable. For a male living in Mexico City, the difference between being interviewed by a woman rather than by

TABLE 3 **Multivariate analysis for responses to gender-sensitive questions (binary probit)**

	Male respondents		Female respondents	
	Abortion question	Women's rights question	Abortion question	Women's rights question
National sample				
Female interviewer	−0.147	0.155	−0.198	−0.100
	[0.126]	[0.133]	[0.113]	[0.109]
Age	−0.003	−0.001	−0.008	−0.008*
	[0.004]	[0.003]	[0.005]	[0.003]
Socioeconomic status	0.011	0.071	0.056	0.128
	[0.070]	[0.065]	[0.077]	[0.071]
Ideology	0.048	0.011	−0.015	−0.003
	[0.035]	[0.032]	[0.040]	[0.031]
Third person present	0.034	0.087	−0.151	0.048
	[0.109]	[0.091]	[0.099]	[0.093]
Education	0.159**	−0.020	0.106**	−0.011
	[0.025]	[0.023]	[0.027]	[0.025]
Church attendance	0.108*	NA	0.189**	NA
	[0.042]		[0.046]	
(Constant)	−0.749*	−0.501	−0.242	0.074
	[0.339]	[0.274]	[0.310]	[0.252]
Log pseudo-likelihood	−435.91	−501.83	−440.51	−538.33
N	711	770	733	787
Mexico City sample				
Female interviewer	0.355*	0.455**	−0.040	0.211
	[0.150]	[0.165]	[0.141]	[0.187]
Age	−0.010*	−0.003	0.000	−0.006
	[0.007]	[0.005]	[0.007]	[0.005]
Socioeconomic status	−0.012	−0.240*	−0.059	0.011
	[0.094]	[0.105]	[0.122]	[0.084]
Ideology	−0.102	−0.071	0.149*	−0.017
	[0.055]	[0.071]	[0.062]	[0.058]
Third person present	−0.034	−0.341	−0.063	0.111
	[0.222]	[0.264]	[0.194]	[0.189]
Education	0.138**	0.056	0.139**	−0.019
	[0.04]	[0.042]	[0.051]	[0.044]
Church attendance	0.058	NA	0.152	NA
	[0.108]		[0.096]	
(Constant)	−0.055	−0.186	−0.167	−0.010
	[0.566]	[0.376]	[0.511]	[0.359]
Log pseudo-likelihood	−115.53	−143.39	−106.29	−174.60
N	224	235	249	263

Note: Cluster-robust standard errors in brackets. *$p < .05$; **$p < .01$.

a man on the issue of abortion rights is roughly equivalent to the difference between being 25 years old and being 60, between having had a few years of high school and having completed only primary school, between attending church regularly and never attending at all, or between being a leftist and a rightist (all other things equal). Although these effects are not dramatic, neither are they trivial. The effects on attitudes toward women's rights are somewhat greater for the same variables. For instance, in terms of education, being interviewed by a woman rather than a man is roughly equal to the difference between not having completed primary school and having finished college.

ITEMS INSENSITIVE TO GENDER AND RESPONSE RATES

The next step in our analysis addressed whether the effects observed for male respondents in Mexico City are present only on the gender-sensitive questions or whether they also affect responses to other items—in other words, whether the effects of interviewer gender are pervasive or restricted to particular topics. Accordingly, we ran binary probit regressions for all other survey questions on issues not considered to be gender-sensitive, using the same statistical model as for the women's rights question. Out of the 14 questions examined, the coefficient for the gender of the interviewer was never significant at the 5-percent level. Gender-of-interviewer effects thus appear to be limited to gender-sensitive items.

To further clarify the scope of gender-of-interviewer effects, we tested whether the gender of the interviewer was a significant determinant of the total count of item non-responses in the whole survey. The interaction between interviewer and respondent gender was not significantly related to item non-response throughout the survey. This result is consistent with previous findings by Groves and Fultz (1985) and Kane and Macaulay (1993).

DISCUSSION

Our findings suggest two conclusions and one area for future research. First, in the research of interviewer-induced bias, the use of national samples may conceal or 'cancel out' differences that exist across different regions or between large cities and other areas. Conversely, samples from a particular region may not be readily generalizable to other settings. In this case, there are reasons why male respondents in Mexico City react differently to female interviewers than do their counterparts in other areas of the country. Mexico City is a cosmopolitan metropolis whose residents hold markedly different political views from most other Mexicans (see, for example, Klesner, 2004; Flores-Macias & Lawson, 2006); and it is the bastion of the leftist Party of the Democratic Revolution (PRD), other leftist splinter parties, and progressive social movements (including feminist groups). By contrast, other large urban centers—Guadalajara, Monterrey, Puebla, Mérida—are known for their conservative politics. What applies to one region is unlikely to hold true in another. Presumably, the same conclusion would hold for many large countries, and we speculate, for example, that a similar experiment in the United States, comparing

a national sample with a sample in a large, politically progressive metropolis such as New York City, could produce results similar to the ones presented here.

Second, on a practical level, our results suggest that survey researchers should take into account potential gender-of-interviewer effects where they might reasonably be suspected. The gold standard for survey research would be to randomize the assignment of interviewers by gender in situations where gender-of-interviewer effects seem likely. Additionally, and far easier in practice than randomization, the gender of the interviewer and the supervisor should be recorded so that researchers can control for potential biases after the fact in their analyses. Indeed, we believe that collecting information on the basic demographic characteristics of the interviewer and the supervisor should be standard practice in all surveys, given the possibility of differentiated effects.

One issue not addressed by our research concerns the cognitive processes that underlie interviewer-induced effects (for competing theories on this issue, see Zaller & Feldman, 1992, and, *inter alia*, Holbrook, Green, & Krosnick, 2003). Are these effects the product of certain respondents providing answers that they expect will please the interviewer, thus avoiding potential social sanctions? Or are they the product of the fact that certain characteristics of the interviewer may render more salient considerations respondents already hold, and thus how they approach the question that confronts them? This research note is agnostic as to the mechanism involved, the investigation of which we leave to future research.

APPENDIX: QUESTION WORDING

Abortion question: 'In your opinion, should abortion in the case of rape be legal or illegal?' (Authors' translation.) Note: Abortion is currently legal in all Mexican states in the case of rape and to save the life of the mother. Abortion is normally illegal under other circumstances, though the prohibition is not systematically enforced.

Women's rights question: 'I am going to read to you several issues that can be addressed by the next president of the country, and I would like you to tell me how urgent each one is. Maybe they are all urgent, but some might be more urgent than others. Would you say that [read issue] is very urgent, urgent, or not so urgent?' (Authors' translation.) The policy issues mentioned were: (a) crime and public safety; (b) government corruption; (c) poverty; (d) unemployment; (e) economic stability; (f) water supply; (g) education; (h) modernization of the electricity sector; (i) trade relations between Mexico and the United States; and (j) women's rights. Note: The order of items was rotated by the interviewer.

REFERENCES

American Association for Public Opinion Research. (2006). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (4th ed.). Lenexa, KS: AAPOR.

Catania, J., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on

responses to questions concerning sexual behavior. *Public Opinion Quarterly, 60*, 345–375.

Flores-Macias, F., & Lawson, C. (2006). Mexican democracy and its discontents. *Review of Policy Research, 23*, 287–293.

Galla, J. P., Frisone, J. D., Jeffrey, L. R., & Gaer, E. P. (1981). Effect of experimenter's gender on responses to a sex-role attitude questionnaire. *Psychological Reports, 49*, 935–940.

Groves, R. M., & Fultz, N. H. (1985). Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods and Research, 14*, 31–52.

Holbrook, A., Green, M., & Krosnick, J. (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*, 79–125.

Huddy, L., Billig, J., Bracciodieta, J., Hoeffler, L., Moynihan, P. J., & Pugliani, P. (1997). The effect of interviewer gender on the survey response. *Political Behavior, 19*, 197–220.

Johnson, T. P., & Moore, R. W. (1993). Gender interactions between interviewer and survey respondents: Issues of pornography and community standards. *Sex Roles, 28*, 243–261.

Johnson, W., & DeLamater, J. (1976). Response effects in sex surveys. *Public Opinion Quarterly, 40*, 165–181.

Kane, E., & Macaulay, L. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly, 57*, 1–28.

Klesner, J. (2004). The structure of the Mexican electorate: Social, attitudinal, and partisan bases of Vicente Fox's victory. In J. I. Dominguez & C. Lawson (Eds.), *Mexico's pivotal democratic election: Candidates, voters, and the presidential campaign of 2000* (pp. 91–122). Stanford: Stanford UP.

Landis, J., Sullivan, D., & Sheley, J. (1973). Feminist attitudes as related to sex of the interviewer. *Pacific Sociological Review, 16*, 305–314.

Lueptow, L. B., Moser, S. L., & Pendleton, B. F. (1990). Gender and response effects in telephone interviews about gender characteristics. *Sex Roles, 22*, 29–42.

Tu, S. H., & Liao, P. S. (2007). Social distance, respondent cooperation and item nonresponse in sex survey. *Quality and Quantity, 41*, 177–199.

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science, 36*, 579–616.

BIOGRAPHICAL NOTE

Francisco Flores-Macias is a doctoral candidate in the Department of Political Science at the Massachusetts Institute of Technology. Previously he was project director for the Transnational Working Families Survey at the Harvard School of Public Health. He holds a B.A. from Harvard University.

Chappell Lawson is an Associate Professor in the Department of Political Science at the Massachusetts Institute of Technology. He is the principal investigator of the Mexico 2006 Panel Study.

Address correspondence to Francisco Flores-Macias, Department of Political Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room E53-448, Cambridge, MA 02139, USA, E-mail: fflores@mit.edu