



Citizenship and Surveys: Group Conflict and Nationality-of-Interviewer Effects in Arab Public Opinion Data

Justin J. Gengler¹ · Kien T. Le¹ · Jill Wittrock²

© The Author(s) 2019

Abstract

More research than ever before uses public opinion data to investigate society and politics in the Middle East and North Africa (MENA). Ethnic identities are widely theorized to mediate many of the political attitudes and behaviors that MENA surveys commonly seek to measure, but, to date, no research has systematically investigated how the observable ethnic category(s) of the interviewer may influence participation and answers given in Middle East surveys. Here we measure the impact of one highly salient and outwardly observable ascriptive attribute of interviewers—nationality—using data from an original survey experiment conducted in the Arab Gulf state of Qatar. Applying the total survey error (TSE) framework and utilizing an innovative nonparametric matching technique, we estimate treatment effects on both nonresponse error and measurement error. We find that Qatari nationals are more likely to begin and finish a survey, and respond to questions, when interviewed by a fellow national. Qataris also edit their answers to sensitive questions relating to the unequal status of citizens and noncitizens, reporting views that are more exclusionary and less positive toward out-group members, when the interviewer is a conational. The findings have direct implications for consumers and producers of a growing number of surveys conducted inside and outside the Arab world, where migration and conflict have made respondent-interviewer mismatches along national and other ethnic dimensions more salient and more common.

Keywords Interviewer effects · Interviewer nationality · Ethnic politics · Nonresponse error · Measurement error · Survey research · Experimental methods · Middle East

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11109-019-09583-4>) contains supplementary material, which is available to authorized users.

✉ Justin J. Gengler
jgengler@qu.edu.qa

Extended author information available on the last page of the article

Over the past 15 years, the Arab world has experienced a critical transition in the availability of nationally representative public opinion data (Tessler 2011).¹ The proliferation of surveys being implemented across the Middle East and North Africa (MENA) has shifted focus from merely procuring data to identifying and correcting problems affecting data quality. While this literature has not been grounded explicitly in the total survey error (TSE) paradigm of survey methodology, it shares the latter's aim of identifying "all the myriad ways in which survey measurement can go wrong" (Smith 2019, p. 14). Indeed, the MENA setting is known to pose particular challenges for survey research, due largely to conservative cultural norms and pervasive authoritarianism (Benstead 2018), and scholars have given special attention to how these and other contextual factors may introduce bias in Arab opinion surveys. For instance, previous research has examined the effects of systematic nonresponse when Arab respondents distrust a survey sponsor (Corstange 2014, 2016; Gengler et al. 2019); refusal to answer survey questions due to doubts about survey confidentiality (Benstead 2018); and measurement error on sensitive items due to a lack of survey privacy during the interview (Diop et al. 2015; Mneimneh et al. 2015).

Perhaps the most frequent methodological concern of survey researchers working in the MENA region, however, has been the impact of interviewers' observable characteristics on the answers of respondents—or interviewer effects (for a recent review, see West and Blom 2017). The Arab world features high levels of religiosity and gender inequality relative to many other survey contexts, and, likely as a result, most research on interviewer effects in Arab opinion data has sought to understand the impact of these two aspects of the MENA survey climate. Perceived religiosity—signaled through manner of dress—and gender of the interviewer have been found to impact responses on various topics in surveys conducted in North Africa, such as adherence to Islamic norms in Egypt and Morocco (Benstead 2014b; Blaydes and Gillum 2013); the role of religion in politics in Tunisia (Mneimneh et al. 2018); gender equality in Morocco (Benstead 2014a); and vote choice in Tunisia (Benstead and Malouche 2016).

However, notably few studies have examined the effects of other interviewer attributes, and how these might relate to other important aspects of the survey-taking environment in Arab states. In particular, one question that has received little attention is how the ethnic² identity of the enumerator may impact survey response behavior. In the MENA context, ethnicity—descent-based membership in racial, tribal, confessional, or other ascriptive groupings—is an interviewer characteristic that is usually outwardly observable based on name, dress, skin color, dialect, and other cues. This omission is significant not only because an extensive literature has investigated ethnicity-of-interviewer effects elsewhere, but also because ethnic identification and competition are widely theorized to mediate important socio-political orientations and behaviors that surveys of Arab populations commonly seek to

¹ This study was made possible by a grant (NPRP 6-086-5-014) from the Qatar National Research Fund, a member of The Qatar Foundation. The statements herein are solely the responsibility of the authors.

² The term "ethnic" is used here broadly to signify any ascriptive, descent-based category, including, among others, race, tribe, religious denomination, language group, and nationality.

measure. These include, among many others, social trust (Inglehart et al. 2006), voting (Corstange 2012; Gao 2016; Lust 2009), support for governments (Gengler 2015), and even views on international relations (Zogby et al. 2012).

A problem thus arises in Middle East surveys if a subject's behavior is affected by the readily-inferred and socially-salient ethnic category of the interviewer. Where ethnic-based cleavages exist, respondents may be less likely to participate in surveys administered by a non-coethnic. Such variance can bias a survey sample in favor of certain respondent types whose opinions and behaviors differ systematically from those of the overall population. Interviewer ethnicity may also cause respondents to edit their answers given in a survey after agreeing to an interview. Finally, studies have demonstrated an interaction between interviewer identity and questionnaire topics (e.g., Adida et al. 2016; Samii 2013), with interviewer ethnicity associated with response error particularly on survey items germane to inter-ethnic relations. All three types of errors can bias estimates of variable means, distributions, and relationships (Groves et al. 2009) and lead to flawed substantive conclusions about mass attitudes and behaviors in the Arab world.

This study extends work on ethnicity-of-interviewer effects geographically, substantively, and methodologically. Utilizing rare original data from the Arab Gulf state of Qatar, we evaluate the impact of a highly salient yet previously neglected ascriptive attribute of interviewers: nationality. Applying the TSE framework, we distinguish between an interviewer effect due to differences in the types of respondents recruited (nonresponse error) versus differences in survey responses given and item nonresponse (measurement error). We also examine the combined effect of measurement error from the interviewer and question topic. To estimate the effects of interviewer ethnicity, we employ an interpenetrated (Mahalanobis 1946) experimental design, with respondents randomly assigned to either a non-conational or conational interviewer. We also leverage an innovative nonparametric estimation technique known as coarsened exact matching (Iacus et al. 2012). Our approach affords a stronger basis for inferring causality over extant, observational studies of ethnicity-of-interviewer effects in the Middle East and Africa.

Background and Literature Review

Effects of interviewers on the measurement of variables have been recognized since the early days of survey research (Lyberg and Stukel 2017). Study of the impact of interviewers' racial and ethnic identities dates to the 1950s (Hyman et al. 1954) and is voluminous. Indeed, a recent review of interviewer effects studies by West and Blom (2017) reveals that race/ethnicity is the single most investigated interviewer characteristic—more than gender, age, or experience—with at least 30 studies published (p. 187). Nonetheless, existing work on ethnicity-of-interviewer effects remains limited in at least three important respects, which the present study helps to address.

First, it has focused on a single source of survey error. The total survey error framework (Groves et al. 2009) summarizes four types of errors potentially introduced by interviewers: (1) coverage error while generating a sampling frame; (2)

unit nonresponse error while contacting and gaining the cooperation of survey respondents; (3) measurement error/item nonresponse error while asking survey questions and conducting measurements; and (4) processing error while recording answers and measurements (West and Blom 2017, p. 178). Remarkably, all 30 ethnicity-of-interviewer effects studies identified by West and Blom fall into the third category, with 25 reporting more socially desirable (less offensive) responses given to non-coethnic interviewers. One might explain the neglect of coverage and processing error by a lack of theoretical motivation. But the fact that no study has examined how the interviewer's ethnic identity may affect success in recruiting survey respondents—i.e., unit nonresponse—represents a significant knowledge gap.

Second, the literature on ethnicity-of-interviewer effects has been advanced based almost exclusively on findings from developed Western countries, especially the United States. As a result, the range of ethnic categories of interviewers and respondents considered in previous research has been extremely limited, rooted in the American experience of inter-race relations and, to a lesser extent, relations between Hispanics and non-Hispanics. Work has only recently begun to expand to other contexts where survey enumerators' outwardly observable ethnic characteristics are salient due to a history of conflict or competition between descent-based groupings. In sub-Saharan Africa in particular, interviewer non-coethnicity has been associated with item nonresponse (Samii 2013) and measurement error due to socially desirable reporting across a range of countries and question types (Dionne 2014; Adida et al. 2016).

To date, however, almost no research has investigated how interviewers' observable ethnic affiliations may impact survey data collected in MENA countries. Surveys undertaken in the Arab world rarely report,³ and even more rarely examine the effects of, interviewers' ethnic traits. It is notable, for instance, that the recently completed fifth wave of the widely-used Arab Barometer survey (Tessler et al. 2016), which has carried out 50 polls across 15 Arab countries since 2006, contains an extensive enumerator questionnaire that records the gender, age, attire, and even facial hair of the interviewer—but no ethnic data. The few studies that have considered the influence of interviewer ethnicity in MENA surveys, meanwhile, identify consequential effects. A 2009 survey of Bahrain, for instance, showed that citizens interviewed by a member of the rival confessional group edited answers to sensitive questions so as to conform to the presumed views of the enumerator (Gengler 2015). Similarly, Gordoni and Schmidt (2010) found that Arab Israelis reported reduced intention to participate in surveys administered by a non-Arab enumerator.

A final limitation of scholarship on the effects of interviewer ethnicity is that it has ignored one increasingly salient ascriptive distinction: nationality. Very little is known about how the national identity, or perceived citizenship, of the enumerator shapes respondent-interviewer interactions during surveys, whether in the Arab world or in other settings where it constitutes a socially or politically salient distinction among the population. To our knowledge, only a single study of 79

³ An exception is Corstange 2012, which reports (p. 492) that respondents were matched with co-sectarian interviewers.

undergraduates by Hue and Sager (1975) has examined nationality-of-interviewer effects. More recently, authors of a United Nations-funded survey of Syrian refugees residing in Lebanon conclude that Syrians tend to underreport feelings of insecurity when interviewed by a Lebanese versus Syrian enumerator, but this finding is only noted in passing (Alsharabati and Nammour 2015, p. 12). Other recent academic surveys of Syrian refugees in Lebanon (e.g., Corstange 2019; Corstange and York 2018) do not report interviewer nationality/ies.

Group Conflict and Nationality-of-Interviewer Effects in the Arab World

The lack of attention to interviewer ethnicity in surveys of Middle East populations is notable given the numerous race- and ethnicity-of-interviewer effects studies conducted elsewhere, including in other developing settings. But it is also conspicuous in light of the heightened salience of group identity amid ongoing regional turmoil. Since the Arab uprisings begun in 2010, the MENA region has witnessed vicious sectarian conflicts and massive flows of displaced persons within and across national borders. Ascriptive identities have thus increased in social and political relevance (Potter 2014; Hashemi and Postel 2017), while more Arab men and women than perhaps ever before interact with members of other national and ethnic groupings. In fact, it is often these conflicts with ethnic dimensions that have motivated new survey-based studies in MENA countries, including in Iraq (Kao and Revkin 2018), Syria (Corstange 2019; Corstange and York 2018), and Yemen (Yemen Polling Center 2017).

The specific salience of nationality as an observable attribute of interviewers in the Arab context stems mainly from internal socioeconomic competition, rather than overt political or inter-state conflict. In the Middle East as elsewhere, citizenship endows rights and privileges not enjoyed by non-citizens. But a majority of Arab states are authoritarian, patronage-based regimes in which leaders cultivate political legitimacy by distributing material benefits to citizens, or some elite subset thereof, as private rather than public goods (Luciani 1987; Lust 2009). That benefits in Arab states accrue disproportionately to individuals rather than society as a whole engenders a clear and pervasive distinction between citizens and non-citizens in economic opportunity, social status, and political rights. Citizens and fiscally-conscious leaders will tend to oppose policies that would expand the pool of eligible state beneficiaries, and more generally may hold negative views of non-nationals as economic opportunists (Longva 2006). Non-citizens have the opposite preference for additional franchise and benefits, while being aware of the unpopularity of such a position, and perhaps themselves, among nationals (Okruhlik 2011).

Nowhere are the politics of citizenship on starker display than in the Arab states of the Gulf. The resource-exporting Gulf monarchies—Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates (UAE)—are home to millions of expatriate workers who make up between one-third (in Saudi Arabia) to more than 85% (in Qatar and the UAE) of the population. Needed to fill skilled and unskilled positions in the oil-based economy, foreigners are employed under

a temporary labor system that offers no political rights, reduced welfare benefits compared to nationals, and no path to citizenship (Okruhlik 2011). Qatari law, for instance, caps naturalization at 50 cases per year (Babar 2014). Ever-expanding non-citizen populations are a source of public discontent in most Gulf countries, as citizens must compete with white-collar expatriates for high-salary professional positions and view expatriates as dissipating government resources at the expense of nationals (Mitchell and Gengler 2019; Al Muftah 2016). This makes nationality a highly impermeable and highly salient ethnic category.

The question of possible interviewer effects arises because, due to their privileged position in society, nationals are rarely employed as enumerators in surveys conducted in the Arab Gulf states. Remuneration is highly uncompetitive compared to a public sector salary, and social expectations surrounding the type of employment fitting for nationals militate strongly against it. Surveys of nationals throughout the Gulf therefore almost always entail an interaction between a citizen respondent and non-citizen interviewer, raising the possibility of widespread bias if the interviewer's non-national identity influences the survey process. In the Arab world, nationality is readily observable on the basis of outward cues, the simplest of which is Arabic pronunciation and dialect. In the Gulf setting it is made even easier by the existence of country-specific 'national dress,' which citizens are compelled by law and social convention to wear in formal situations, including at work. Whether in telephone or face-to-face interviews, nationality is easily inferred.

Hypotheses

We expect interviewer nationality to be a source of survey error in Qatar and in other settings where citizenship status is a salient and outwardly observable trait. Moreover, we hypothesize that non-nationality will be associated not only with more socially desirable reporting and increased item nonresponse, as found in previous ethnicity-of-interviewer studies, but also unit nonresponse. Studies have shown that sociodemographic likenesses between the interviewer and respondent, such as gender and education, tend to increase cooperation rates (West and Blom 2017, pp. 181, 182), and we theorize that shared national or other group identity should function similarly.

In our telephone survey, we predict that Qatari respondents will alter their response and nonresponse behavior based on generalizations from observable cues during the survey process (Tajfel and Turner 1979)—namely, the interviewer's Arabic dialect and pronunciation. First, when the respondent detects a conational Qatari interviewer on the phone, as signaled by a Qatari dialect, she may be more likely to agree to and finish the interview. Conversely, when the respondent hears a non-Qatari Arabic dialect and perceives a non-Qatari interviewer, then he may be less inclined to cooperate and, if he does, more likely to terminate the interview early. These expectations about the effects of interviewer nationality on nonresponse are captured in Hypotheses 1a and 1b.

H1 Conational Qatari interviewers increase the likelihood that the respondent (a) acquiesces to be interviewed; and (b) completes the entire interview schedule.

In line with previous results from Western settings as well as North and sub-Saharan Africa, we also expect non-conational interviewers to be associated with greater item nonresponse. Rather than editing their responses, respondents may decline to answer sensitive items altogether in order to avoid conforming to what they perceive as the likely views of the interviewer (Benstead 2014a) as deduced from their nationality. We thus expect a lower incidence of “Don’t Know” or Refused responses when the respondent detects a Qatari interviewer. This is Hypothesis 2.

H2 Conational Qatari interviewers decrease the incidence of item nonresponse.

Finally, and also in accordance with previous results, we hypothesize that interviewer nationality will influence answers to individual survey items. We expect respondents to report less offensive views on sensitive questions about non-nationals when asked by an interviewer believed to be a non-Qatari, as inferred from speech. Qataris interviewed by a non-Qatari may also give more socially desirable responses to questions that are sensitive but unrelated to non-nationals or citizen-noncitizen relations. These predictions comprise Hypotheses 3a and 3b.

H3 Conational Qatari interviewers are associated with (a) more offensive responses to sensitive items concerning non-nationals; and (b) less socially desirable responses to questions that are sensitive but unrelated to non-nationals.

Data and Experimental Design

We measure the effects of interviewer nationality using data from an original telephone survey implemented in Qatar in June 2014.⁴ The survey was carried out by the Social and Economic Survey Research Institute (SESRI) at Qatar University. The survey sample was drawn from a nationally-representative cell phone frame obtained from the largest telecommunications provider in Qatar, with approximately 95% coverage of adult citizens. We used a split ballot technique to divide the sample into two groups: one assigned interviewers who are Qatari nationals, the other assigned non-Qatari interviewers.⁵ Respondents received random assignment to the Qatari or non-Qatari interviewer group using the Halton sequence (Le et al. 2018), and cases remained within their assigned group in the event of callback. A total of 1587 respondents were successfully recruited for participation, with 1288 finishing the complete interview schedule and 299 (19%) exiting the interview before

⁴ Data and replication code are available via the Harvard Dataverse at <https://doi.org/10.7910/DVN/7JW4EL>.

⁵ Non-Qatari interviewers included nationals of Jordan, Oman, Sudan, Syria, and Yemen.

completion. The overall survey response rate, following the AAPOR definition RR3, was 34.8%, with a sampling error of 4.1%.

Enumerators were all female Qatar University students aged between 20 and 30 years, were of similar experience level, and underwent the same pre-survey training. All interviewers were instructed to converse with respondents using their native Arabic accent and pronunciation. Beyond Arabic dialect, no explicit prompting identified interviewers as Qataris (29 total) or non-Qataris (13 total), making for a conservative test of the effects of interviewer nationality.

Introduced to participants generically as a study of “important social and cultural trends in Qatari society,” the survey instrument contained a mix of non-sensitive questions alongside items designed to touch on local sensitivities in the economic, social, and political domains. The social category included questions that asked explicitly about relations between nationals and non-nationals, including trust in various nationality groups, perceptions toward foreigners, and views about immigration, naturalization, and other policies related to non-nationals. The social category also included questions about intra-Qatari issues, such as changes in social and religious norms. The economic category comprised subjective assessments of the country’s and the respondent’s economy. The political category included, among others, questions on voting behavior, political interest, and concerns over government surveillance—topics unrelated to non-nationals.⁶ The survey concluded with an enumerator questionnaire used to verify the experimental treatment.

Methods

We estimate the effects of interviewer nationality on unit nonresponse, early termination of the interview (drop-off), item nonresponse, and answers to individual survey items. We carry out these comparisons using both *t*-tests and coarsened exact matching (CEM) (Iacus et al. 2012). Although the study’s experimental design means that simple *t*-tests can be used to assess overall differences in variable means based on the treatment, such comparisons cannot account for potential differences in respondent types between the Qatari and non-Qatari interviewer groups. As a result, difference-of-means testing leaves open the question of whether observed differences stem from variation in participant-interviewer interaction during the interview (measurement error), or instead reflect differences in the *types* of survey participants—e.g., younger or less educated—that tend to be recruited and/or retained by an interviewer group (nonresponse error). Equally, a lack of difference in overall means may result from competing effects on measurement and nonresponse that negate each other, giving the false appearance of no interviewer effect.

In order to differentiate and examine separately these two types of effects, we use matching to evaluate between-group differences while controlling for discrepancies

⁶ The full schedule of questions utilized in the analysis is provided in the online appendix. It should be noted that questions were not arranged in the questionnaire thematically, so that priming occurred at the level of an individual item rather than in the context of an entire survey section, creating a harder test.

in respondent attributes. We employ coarsened exact matching, or CEM, which has several attractive properties compared to other matching and non-matching approaches to assessing interviewer effects. First, CEM is designed precisely to account for the confounding influence of pretreatment control variables. CEM also does not rely on parametric modeling and thus offers reduced standard errors compared to mixed-effects regression (Heckman et al. 1998; Rubin and Thomas 2000; Rubin 2001), which has often been used to study interviewer effects.⁷ Finally, the relatively high ratio of interviewers to respondents in our data, and variation in the number of interviews carried out by individual enumerators, makes multi-level modeling less suitable.

Coarsened Exact Matching

Any study aims to estimate the effect of a treatment by comparing outcomes. The treatment effect for an individual is the comparison between the outcome if the individual receives the treatment and the outcome if the individual does not (Rubin 1974). The problem in this estimation is that we observe only one outcome per individual, as each individual receives either the treatment or the control, but not both (Holland 1986). In our case, the treatment is the interviewer's nationality, and outcomes of interest are respondents' answers and nonresponses. Each individual is assigned a Qatari interviewer or not, so we observe each individual in one condition only. It is not possible, therefore, to estimate the treatment (interviewer) effect for each individual, and we must instead estimate a treatment effect averaged across the population.

The standard estimation framework is based on a Rubin (1974) model. In this model, the outcome if individual i receives the treatment effect is $Y_i(1)$ and the outcome if individual i receives the control is $Y_i(0)$. The treatment effect for an individual i can be written as $Y_i(1) - Y_i(0)$. This individual effect cannot be estimated; instead we estimate the average treatment effect:

$$\tau = E[Y(1)] - E[Y(0)],$$

where $E[Y(1)]$ is the expected outcome of the treated group and $E[Y(0)]$ is the expected outcome of the control group. If this treatment effect is calculated for all individuals, then it is called the average treatment effect (ATE); but if it is calculated for individuals in the treated group only, then it is called the average treatment effect on the treated (ATT). Here, as in most applied studies based on this model, ATT is utilized to calculate the effect of interviewer conationality.

We use matching to estimate the treatment effect. Matching proceeds by identifying a set of individuals in the control group who are similar to the individuals in the treated group across all relevant characteristics. Then, the difference in outcome

⁷ For comparison, we report the corresponding mixed-effects estimate as part of the diagnostic analysis of Table 2 (see Model 6). The two methods do not produce different substantive conclusions about the effect of interviewer conationality. Interviewer random effects also did not improve model fit for responses to individual survey items.

between this well-selected control group and the treated group will reflect the treatment effect (Rubin 1974; Heckman et al. 1998). Many matching methods exist, the most common of which are exact matching, Mahalanobis distance matching, propensity score matching, and more recently coarsened exact matching. A growing body of literature suggests that CEM is easier to use and understand, requires fewer assumptions, and has more attractive statistical properties than other popular matching strategies (Iacus et al. 2011, 2012; King and Nielsen 2016).

The motivation for CEM is that while exact matching between treated and control groups offers perfect balance in individual characteristics, in practice it usually produces relatively few matches due to the curse of dimensionality (Stuart 2010). The idea behind CEM is to coarsen each characteristic into meaningful categories. For example, education can be categorized into primary, secondary, and tertiary levels. Then, exact matching can be performed on this coarsened attribute. Our implementation of CEM closely follows the procedure of Iacus et al. (2012). First, individual characteristics are coarsened into categories. Next, a set of strata are created from the combination of these categories.⁸ Only strata with at least one individual from both control and treated groups are retained for matching. This matched sample becomes the basis for subsequent analysis.

In each stratum, the number of individuals from the control group will most often differ from the number of individuals from the treated group, so in the second stage of CEM adjustment weights are needed to account for this difference:

$$w_i = \begin{cases} 1 & \text{if individual } i \text{ is in the treated group} \\ \frac{m_C m_T^s}{m_T m_C^s} & \text{if individual } i \text{ is in the control group} \end{cases}$$

where m_C is the total number of individuals in the control group, m_T is the total number of individuals in the treated group, m_T^s is the number of individuals in the treated group in stratum s , and m_C^s is the number of individuals in the control group in stratum s . Matching quality is then assessed by comparing individual characteristics between the treated group and the control group, taking into account the adjustment weights. A good matching will find no significant between-group difference in characteristics. Here we utilize the multivariate imbalance measure of Iacus et al. (2011, 2012) to make this assessment.

Finally, a simple weighted regression is used to estimate the treatment effect. The sole explanatory variable is the treatment (1 for treated and 0 for control), and the coefficient on this variable represents the treatment effect. The dependent variable is the survey outcome of interest. Different types of regression can be used at this stage, depending on the outcome variable type. In this way, CEM affords considerable flexibility and ease of use.

⁸ Consider the case of matching on two respondent characteristics: education and income. If education is coarsened into three categories (primary or below, secondary, high school or above) and income into five categories (very low, low, medium, high, very high), then a set of 15 (3×5) strata results from the combination of these two characteristics.

Estimating Interviewer Effects

Our explanatory variable, interviewer conationality, is a dichotomous indicator coded 1 for Qatari interviewers and 0 for non-Qatari interviewers.⁹ We also include in our baseline model three coarsened variables to control for respondent demographic attributes that may affect response and nonresponse: gender, age, and education. Gender is a binary variable; age is divided into four categories according to sample quartiles; and education is also coded into four categories (primary and below, secondary, post-secondary, Bachelor's and above). Matching is achieved for all but 7 (less than 1% of) treated cases, with no evidence of between-group imbalance in the strata.

We begin our analysis by assessing the impact of interviewer conationality on unit nonresponse. We then estimate its effect on the likelihood of early survey termination. This outcome variable is binary—coded 0 for a completed interview and 1 for a break-off—with a mean value of 0.18 and standard deviation of 0.39.¹⁰ We thereafter consider the impact of nationality on item nonresponse. Past research has focused on nonresponse rates on sensitive questions, but here we take a wider view of item nonresponse by estimating the effect of interviewer conationality on the total number of “Don't Know” and Refused events across the full substantive interview schedule. The resulting variable has a mean of 1.3 and standard deviation of 2.8.

As a robustness check, we next perform a diagnostic analysis that introduces additional variables to our CEM model meant to capture latent respondent traits that may impact survey behavior: overall comfort level as evaluated by the interviewer; and psychological susceptibility to engaging in impression management.¹¹ The former is coded 1 if the enumerator reports that “The respondent was comfortable answering almost all questions” (82% of cases), and 0 otherwise. The latter is measured using what is, to our knowledge, the first Arabic-language implementation of a shortened form of the BIDR Impression Management subscale (Winkler et al. 2006), a well-established measure of susceptibility to socially desirable responding (Paulhus 1984). We combine items from the subscale to create a straightforward additive index, which we coarsen into terciles. Inclusion of the respondent comfort and impression management controls increases the proportion of unmatched treated cases only slightly, to 3% and 4%, respectively.

Finally, we examine the influence of interviewer conationality on the several categories of survey items described already. As explained, we both assess the overall effect of nationality, including via nonresponse error, and isolate its direct effect on measurement error after controlling for differences in respondent characteristics between the treated and control groups.

⁹ To avoid disparities in interviewer experience, we exclude from the matching analysis seven Qatari interviewers who conducted fewer than 10 interviews each (i.e., ever), totaling 24 cases.

¹⁰ Break-offs due to technical fault (11 cases) are excluded from the analysis.

¹¹ Note that, because these questions appear at the end of the survey, the comfort and image management controls cannot be included in diagnostic estimations of early termination/drop-out.

Table 1 Effect of interviewer conationality on early termination/drop-off and item nonresponse

Variable	<i>t</i> -test	CEM
Early termination/ drop-off (0 completed, 1 dropped)	-0.0402*** (0.0196)	-0.0260 (0.0201)
Total item nonre- sponse (continuous with range 0–31)	-0.679*** (0.158)	-0.870*** (0.192)

Standard errors in parentheses

CEM column reports post-regression estimated conditional marginal effect with standard error

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Estimations are carried out via CEM-weighted logistic, ordered logistic, negative binomial, and OLS regression, as appropriate, using the CEM Stata package of Iacus et al. (2009).

Results

Interviewer Nationality and Unit Nonresponse, Early Termination, and Item Nonresponse

The data indicate significant differences in the number and characteristics of respondents recruited for survey participation by Qatari versus non-Qatari interviewers. Given random assignment of interviewers,¹² if interviewer nationality had no effect on nonresponse, one should observe each interviewer group accounting for a statistically indistinguishable number of recruited participants. This was not the case. Instead, non-Qatari enumerators commenced interviews with a total of 755 respondents, compared to 832 respondents among Qatari interviewers. The conational interviewer group thus accounted for 52.4% of all interviews, with an associated 95% confidence interval of between 50.0 and 54.9%. Conational interviewers thus appear to bolster the response rate, in line with our Hypothesis 1a.

Qatari interviewers also tended to recruit participants who are less educated (one-tailed $p = 0.013$) as well as potentially older ($p = 0.114$) and disproportionately male ($p = 0.096$) compared to those recruited by non-Qatari interviewers. These effects suggest that such categories of citizens perceive greater social distance between themselves and non-nationals, and so are more likely to decline to participate in the survey when they detect a non-Qatari interviewer. It is also plausible that less educated and older respondents face more difficulty comprehending and conversing with enumerators speaking in foreign Arabic accents, and thus participate less often with non-Qatari interviewers. However, the latter interpretation cannot account for

¹² Interviewer group assignment achieved near-perfect balance across the survey sample, at 49.94% versus 50.06%.

the seeming discrepancy in participant gender. Whatever the case, the two respondent subsamples differ across one or more key demographic categories that are potentially correlated with survey response and nonresponse.

The effects of interviewer conationality on early survey termination (drop-off) and item nonresponse are given in Table 1. The first column reports the outcome of a *t*-test of the difference in means between the treated and untreated cases. The CEM column reports the estimated effect of Qatari interviewer nationality after matching on respondent attributes. The results indicate that conationality does reduce the likelihood of respondent drop-off, and that this effect stems from differences in the types of respondents recruited by Qataris versus non-Qataris. In particular, early termination is more likely among less educated ($p < 0.000$) and older ($p = 0.007$) respondents, two groups that, as shown already, are recruited in higher proportions by Qatari interviewers.

Regarding item nonresponse, by contrast, CEM results show that interviewer conationality introduces bias via measurement error rather than nonresponse error. That is, its impact does not stem from a difference in respondent *types* recruited by the two groups of interviewers, but rather from differences in respondent-interviewer interaction. As demonstrated in Table 2, this finding is robust to a wide range of alternative CEM specifications, including models that control for latent respondent psychological variables such as nervousness and propensity to engage in impression management. Overall, the data give strong support to Hypotheses 1a, 1b, and 2: Qataris are more likely to begin and finish a survey, and respond to questions, when interviewed by fellow nationals.¹³

Practically, the impacts of interviewer conationality on drop-off and item nonresponse are significant. The rate of early termination is 20.4% among respondents assigned to a non-Qatari interviewer and only 16.4% among respondents with a conational interviewer, a relative difference of 20%. Although this effect is an artifact of higher incidence of break-off among older and less educated individuals, still the bias it introduces is substantial. In the case of item nonresponse, the impact of conationality is tempered only somewhat by controlling for respondent characteristics. Even after accounting for the effects of conationality via respondent recruitment, Qataris interviewed by Qataris are estimated to have 1.1 nonresponse events across the interview schedule, compared to 2.0 for respondents with non-Qatari interviewers, a relative increase of 82%. There is no difference in this effect based on respondent gender.

Another pertinent question is whether these observed treatment effects are attributable to conational favoritism, or instead to non-conational discrimination. To help make this distinction, we test for differential effects on drop-out and nonresponse based on the particular nationality of the non-Qatari interviewer. Specifically, we

¹³ A manipulation check included in the enumerator questionnaire confirms the respondent's perception of interviewer nationality: 76% of Qatari interviewers indicated that the respondent signaled directly (by asking or stating) or indirectly (e.g., by switching to Qatari dialect) that s/he recognized the interviewer as a Qatari. Meanwhile, only 5% of non-Qatari interviewers reported a direct or indirect indication that the respondent recognized them as a Qatari.

Table 2 Effect of interviewer conationality on item nonresponse, by CEM specification

	(1) Baseline	(2) + Comfort	(3) + IM	(4) + Comfort, IM	(5) + Workload	(6) Mixed-effects
Qatari interviewer	-0.586 ^{***} (0.000)	-0.632 ^{***} (0.000)	-0.445 ^{***} (0.000)	-0.525 ^{***} (0.000)	-0.490 ^{***} (0.000)	-0.538 [*] (0.014)
Constant	0.673 ^{***} (0.000)	0.696 ^{***} (0.000)	0.510 ^{***} (0.000)	0.400 ^{***} (0.000)	0.579 ^{***} (0.000)	-0.175 (0.459)
ln(alpha)	0.995 ^{***} (0.000)	0.957 ^{***} (0.000)	1.045 ^{***} (0.000)	1.037 ^{***} (0.000)	0.946 ^{***} (0.000)	0.802 ^{***} (0.000)
N	1260	1241	1225	1116	1249	1253
N unmatched	7	24	34	108	18	-

Model 1 is the baseline CEM model whose predicted marginal effect is reported in Table 1, and that matches only on respondent gender, age, and education. Model 2 adds into the matching specification a control for respondent comfort level, as subjectively assessed by the interviewer. Model 3 matches on latent propensity for socially desirable reporting, as measured by the IM scale index. Model 4 includes both the comfort and IM controls. Model 5 matches on interviewer workload as measured by total number of interviews, coarsened into terciles. Model 6 estimates a mixed-effects negative binomial regression with interviewer random effects and the same controls for respondent age, education, and gender

p-values in parentheses

^{*}*p* < 0.10; ^{**}*p* < 0.05; ^{***}*p* < 0.01; ^{****}*p* < 0.001

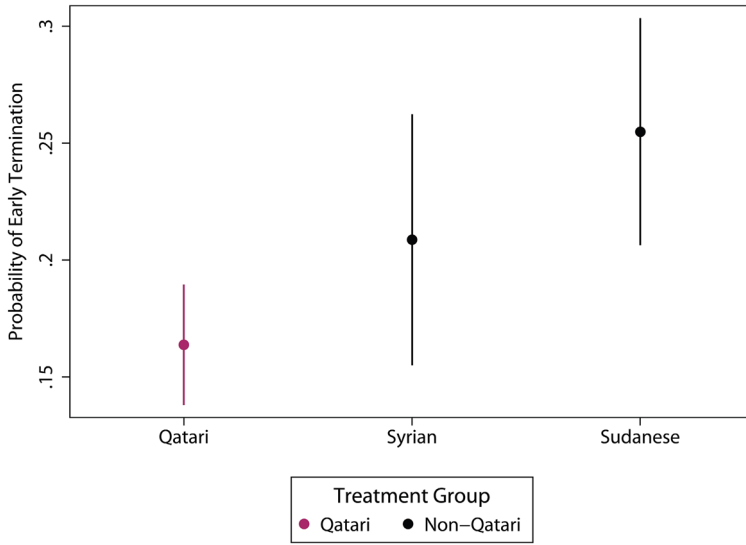


Fig. 1 Effect of interviewer nationality on likelihood of early termination/drop-off

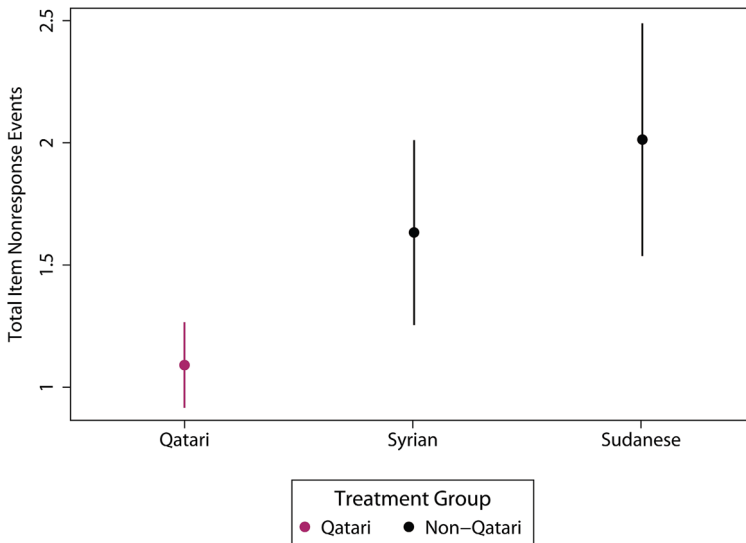


Fig. 2 Effect of interviewer nationality on total item nonresponse

compare effects between the two largest non-Qatari interviewer nationalities, Sudanese and Syrian, which comprise 84% of non-Qatari cases. Figures 1 and 2 show that there is no difference in effect based on non-Qatari interviewer nationality, either on early termination ($p=0.218$) or item nonresponse ($p=0.215$). However,

Table 3 Effect of interviewer conationality on sensitive items: economics

Survey item	<i>t</i> -test	CEM	Percent change Y (%)
Subjective evaluation of			
Qatar's economy (1 very weak, 6 excellent)	0.0116 (0.0360)	0.0270 (0.118)	–
Respondent's household economy (1 very weak, 6 excellent)	0.155** (0.0540)	0.260** (0.0984)	+3.1
Overall satisfaction with			
Respondent's household economy (0 lowest, 10 highest)	0.440*** (0.121)	0.417** (0.126)	+5.8

CEM column reports weighted regression coefficients with standard errors in parentheses; Final column reports relative change in predicted value of dependent variable due to interviewer conationality, based on CEM results; See the Online Appendix for individual question wordings

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

in both instances the predicted difference in outcome is greater between Qatari and Sudanese interviewers as compared to Qatari versus Syrian interviewers. On balance, the results lend more support for an explanation of coethnic favoritism rather than aversion toward specific non-coethnic groups as the driver of nationality-of-interviewer effects in Qatar, but the evidence is certainly not conclusive.

Interviewer Nationality and Answers to Sensitive Questions

We next test the impact of interviewer conationality on four different categories of survey items: economics, inter-communal social relations, intra-Qatari social relations, and politics. All are sensitive in the local context, but only one category—inter-communal social relations—is sensitive because it asks about non-nationals or relations between citizens and non-citizens.

Table 3 summarizes the effect of interviewer conationality on survey items related to economics. Treatment effects are reported first as simple differences of means without accounting for respondent demographics, and then as obtained by CEM. The final column quantifies the substantive impact of an effect, computed as the relative percent change in predicted value of the dependent variable due to conationality, based on the CEM estimation. Even after controlling for between-group disparities in respondent demographics, Qatari interviewers are associated with a 3–6% increase in the predicted mean of survey items that entail financial self-assessments. Notably, there is no such effect when the question asks a respondent to rate the country's economic situation rather than their personal circumstances, suggesting that the observed differences stem from measurement error due to social desirability bias rather than some other process.

The corresponding findings in the domain of inter-communal social relations are reported in Table 4. Interviewer effects remain consistent but in the opposite substantive direction: Qataris receiving the conational interviewer treatment are

Table 4 Effect of interviewer conationality on sensitive items: inter-communal social relations

Survey item	<i>t</i> -test	CEM	Percent change Y (%)
Views toward			
Expats' right to same salary increase as Qataris (1 deserve nothing, 3 deserve the same)	-0.115*** (0.0255)	-0.611*** (0.138)	-6.0
Favoritism toward expats in hiring (1 expat more likely to be hired, 3 Qatari more likely)	-0.0297 (0.0525)	-0.162 (0.118)	-
Opening new places of worship for non-Muslims (1 would be very bad, 4 would be very good)	-0.118* (0.0557)	-0.217* (0.105)	-5.0
Allowing citizenship for residents of 25+ years (1 would be very bad, 4 would be very good)	-0.0247 (0.0544)	-0.116 (0.107)	-
Size of foreign population in Qatar (1 should be reduced a lot, 5 should be increased a lot)	-0.171** (0.0669)	-0.310** (0.110)	-7.8
Trust in			
Arab expats (1 least, 4 most)	-0.116** (0.0432)	-0.317** (0.114)	-4.2
Western expats (1 least, 4 most)	-0.110* (0.0472)	-0.278* (0.109)	-4.8
Expats from the Indian subcontinent (1 least, 4 most)	-0.0568 (0.0469)	-0.194 ⁺ (0.108)	-3.4
Agreement that			
Newly naturalized citizens represent Qatar (1 strongly disagree, 4 strongly agree)	-0.0475 (0.0542)	-0.189 ⁺ (0.110)	-3.5
People of foreign ancestry can never truly be Qatari (1 strongly agree, 4 strongly disagree)	-0.0199 (0.0580)	-0.0285 (0.110)	-
Qatar's naturalization laws should be relaxed (1 strongly disagree, 4 strongly agree)	-0.113* (0.0580)	-0.203 ⁺ (0.111)	-3.2
Combined inter-communal relations (sum of all non-missing items)	-0.0902*** (0.0235)	-0.109*** (0.0258)	-4.4

CEM column reports weighted regression coefficients with standard errors in parentheses

Final column reports relative percent change in predicted value of dependent variable due to interviewer conationality, based on CEM results; See the Online Appendix for individual question wordings

⁺ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

associated with values of the dependent variable that are between 3 and 8% *lower* than those interviewed by non-Qataris, where higher values represent more positive attitudes regarding non-nationals. Overall, among 11 questions in the inter-communal social domain included in the survey, 8 (or 73%) show a difference in response based on conationality, and all estimated coefficients are negative. A test of the combined effect of Qatari interviewer nationality on responses to all 11 questions in this category indicates an average impact of -4.4% , with an associated significance level of $p < 0.000$. Finally, comparison of the t -test and CEM results shows that these effects are not due to different respondent types recruited by conational interviewers, but to respondent-interviewer interaction. Meanwhile, nonresponse bias is shown in two instances to *mask* interviewer effects that are not apparent from the t -test results, being only revealed when confounding respondent attributes are controlled via CEM.

By contrast, questions that ask about societal issues internal to Qatari society, rather than relations between nationals and non-nationals, elicit far fewer treatment effects. (Results shown in Table A1 in the Online Appendix.) The magnitude and direction of the effects are also inconsistent. Only two of eight questions evidence a between-group difference after controlling for respondent characteristics via CEM. What is more, the substantive direction of an effect appears to vary by question type: for those that ask about personal cultural and religious orientations, responses to Qatari interviewers are less conservative than those reported to other nationalities; but for questions that ask about the appropriateness of behavior in outside contexts featuring foreigners—namely, a university campus or foreign country—Qataris report more conservative responses to conationals. This is consistent with the idea that questions touching on inter-communal relations will prime nationals to think about their relationship to the out-group interviewer and provoke answers that minimize the perceived social distance between them. Finally, we observe again disagreement between the t -test and CEM results. In two instances, interviewer effects are seen to be artifacts of respondent demographic differences resulting from the impact of conationality on recruitment and retention. Such cases reiterate the importance of understanding and estimating the impact of interviewer attributes not only on measurement of sensitive items but also nonresponse.

With respect to the last category of questions in the survey—items about sensitive political attitudes and behavior—responses are essentially unaffected by conationality. Of nine questions in this category, none features a between-group difference significant at the $p < 0.05$ level after respondent demographics are controlled, and only one at the $p < 0.10$ level (Results not shown.)

In sum, only two question domains are associated with consistent nationality-of-interviewer effects: economic self-evaluations; and views of foreigners and policies related to immigration, naturalization, and related themes. By contrast, societal issues internal to the Qatari citizenry, including personal religiosity and cultural norms, are not consistently affected by conationality. Finally, strictly political items that are sensitive but do not involve citizen-expatriate relations elicit no interviewer effect. These patterns lend robust support for Hypothesis 3a, but not 3b.

Discussion and Implications

This study finds strong evidence that the nationality of the interviewer has systematic effects on unit nonresponse, early termination of the interview, and item nonresponse, as well as answers given to individual survey items, in a representative survey of Qatari citizens. Qataris are less likely to agree to an interview, and to remain on the phone for the duration of the survey, if the interviewer is linguistically recognizable as a non-national. They are also more likely to forego responding to questions when talking to a foreigner. Finally, when answering questions that touch on evaluations of, relations with, and policies regarding non-nationals, Qataris engage in socially desirable reporting when speaking with non-national interviewers, reporting views that are less exclusionary and more positive toward out-group members than responses given to those assumed to be fellow nationals. In short, few measurable response and nonresponse behaviors of Qatari citizens are *not* mediated by the national identity of the interviewer.

Several considerations highlight the importance of these findings for survey data producers and consumers working in settings where nationality and other observable ethnic categories hold social or political relevance. Methodologically, the experimental design of our study represents an advance over extant work on ethnicity-of-interviewer effects in Africa and the Middle East, which relies on observational data. We also leveraged a relatively new matching technique, CEM, which offered advantages to estimating interviewer effects in our data, and represents a fruitful addition to the toolkit available to survey researchers. Separately, interviewer effects are known to be much smaller in telephone surveys compared to face-to-face surveys (Groves and Magilavy 1986), and so our results likely represent a conservative estimate of the magnitude of bias. Our estimates are likewise conservative insofar as interviewers did not directly state their nationality to respondents.

Yet our study has limitations, of course. It focuses on only one of many ethnic attributes of interviewers that may be expected to alter response and nonresponse behavior in Arab opinion surveys. While nationality possesses particular salience in Qatar and the Arab Gulf states, future ethnicity-of-interviewer effects research in other MENA settings could examine the influence of Arab versus non-Arab descent, tribal background, or confessional religious affiliation, according to the character of local group cleavages. Extensions of our study could also investigate potential interaction effects between the interviewer's ethnic category and other observable characteristics such as gender or religiosity, which have been at the center of previous interviewer effects research in the MENA region. Finally, our use of a telephone survey to assess nationality-of-interviewer effects in Qatar was advantageous in several practical and methodological respects, and represents a harder test in comparison to a face-to-face interview. But future work could explore possible differences in the nature and magnitude of bias due to interviewer ethnicity when the interview is conducted in person rather than over the phone. One straightforward next step would be the collection of interviewer ethnicity data via the enumerator questionnaires already employed in regional survey projects such as the Arab Barometer.

Far from a theoretical question of survey methodology, nationality-of-interviewer effects in Qatar and the demographically-similar Arab Gulf states represent a very practical problem. The simplest approach to addressing the influence of interviewer nationality and other ethnic identities is to code and control for these characteristics in the manner of other interviewer attributes. But this solution depends on recruiting conationals or other coethnics to work as interviewers. In many Gulf settings, social norms and remuneration expectations may make this unduly expensive or, depending on the survey mode and topic, altogether impossible. Fully self-administered modes such as web-based surveys may eliminate the question of interviewer effects altogether, but they may also introduce new sources of survey error that could render the cure worse than the disease.

The problem of respondent-interviewer nationality mismatches is not only a Gulf concern. Competition between citizens and noncitizens features in a growing number of settings inside and outside the Middle East where poverty and conflict have fuelled migration across national borders and amplified the social and political salience of citizenship status. The Syrian civil war alone has prompted numerous survey-based studies of Syrian refugees residing in Lebanon (Alsharabati and Nammour 2015; Corstange and York 2018; Corstange 2019), Turkey (Fabbe et al. 2017a, b), and even Western Europe (The Syria Campaign 2015). Palestinians living in neighboring states have also been a frequent population of study (Sirhan 1975). Yet surveys of non-national groups still rarely offer details about enumerators' national identities, suggesting that local citizens were used or, at least, that the nationality/ies of interviewers was not deemed relevant to report.

More broadly, the same generally applies to other ethnic attributes of interviewers used in Arab opinion surveys. The Middle East and North Africa is an ethnically diverse region, and social and political cleavages often run along family, tribal, religious, and other ascriptive lines. As survey research continues to expand in the Arab world, and as researchers pay more attention to producing reliable and representative survey data, further study is needed into the ways that interviewers' ethnic identities affect behavior in public opinion surveys.

Acknowledgements Open Access funding provided by the Qatar National Library. The authors would like to thank the anonymous referees for their helpful feedback and suggestions that greatly contributed to improving the final version of this article. They would also like to thank the Editors for their generous support during the review process. Data collection for this study was supported by a Grant (NPRP 6-086-5-014) from the Qatar National Research Fund, a member of The Qatar Foundation.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adida, C. L., Ferree, K. L., Posner, D. N., & Robinson, A. L. (2016). Who's asking? Interviewer coethnicity effects in African survey data. *Comparative Political Studies*, 49, 1630–1660.

- Al Muftah, H. (2016). Demographic policies and human capital challenges. In M. E. Tok, L. Alkhatir, & L. A. Pal (Eds.), *Policy-making in a transformative state*. London: Palgrave.
- Alsharabati, C., & Nammour, J. (2015). Survey on perceptions of Syrian refugees in Lebanon. Beirut: Institut des Sciences Politiques. Université Saint-Joseph de Beyrouth. <https://data2.unhcr.org/en/documents/download/45083>.
- Babar, Z. R. (2014). The cost of belonging: Citizenship construction in the state of Qatar. *Middle East Journal*, 68, 403–420.
- Benstead, L. J. (2014a). Effects of interviewer-respondent gender interaction on attitudes toward women and politics: Findings from Morocco. *International Journal of Public Opinion Research*, 26, 369–383.
- Benstead, L. J. (2014b). Does interviewer religious dress affect survey responses? Evidence from Morocco. *Politics and Religion*, 7, 734–760.
- Benstead, L. J. (2018). Survey research in the Arab world: Challenges and opportunities. *Political Science and Politics*, 51, 535–542.
- Benstead, L. J., & Malouche, D. (2016). Interviewer Religiosity and Polling in Transitional Tunisia. Paper presented at the annual meeting of the southern political science association, San Juan, Puerto Rico.
- Blaydes, L., & Gillum, R. M. (2013). Religiosity-of-interviewer effects: Assessing the impact of veiled enumerators on survey response in Egypt. *Politics and Religion*, 6, 459–482.
- Corstange, D. (2012). Vote trafficking in Lebanon. *International Journal of Middle East Studies*, 44, 483–505.
- Corstange, D. (2014). Foreign-sponsorship effects in developing-world surveys: Evidence from a field experiment in Lebanon. *Public Opinion Quarterly*, 78, 474–484.
- Corstange, D. (2016). Anti-American behavior in the Middle East: Evidence from a field experiment in Lebanon. *Journal of Politics*, 78, 311–325.
- Corstange, D. (2019). The Syrian conflict and public opinion among Syrians in Lebanon. *British Journal of Middle Eastern Studies*, 46, 178–200.
- Corstange, D., & York, E. (2018). Sectarian framing in the Syrian civil war. *American Journal of Political Science*, 62, 441–455.
- Dionne, K. Y. (2014). The politics of local research production: surveying in a context of ethnic competition. *Politics, Groups, and Identities*, 2(3), 459–480.
- Diop, A., Le, K. T., & Traugott, M. (2015). Third-party presence effect with propensity score matching. *Journal of Survey Statistics and Methodology*, 3, 193–215.
- Fabbe, K., Hazlett, C., & Sinmazdemir, T. (2017a). Displaced loyalties: Attitudes among Syrian refugees in Turkey. Working paper 18-024. Harvard Business School.
- Fabbe, K., Hazlett, C., & Sinmazdemir, T. (2017b). What do Syrians want their future to be? *Foreign Affairs*.
- Gao, E. (2016). Tribal mobilization, fragmented groups, and public goods provision in Jordan. *Comparative Politics Studies*. <https://doi.org/10.1177/0010414015621075>.
- Gengler, J. (2015). *Group conflict and political mobilization in Bahrain and the Arab Gulf: Rethinking the rentier state*. Bloomington: Indiana University Press.
- Gengler, J., Tessler, M., Lucas, R., & Forney, J. (2019). ‘Why do you ask?’ The nature and impacts of attitudes toward public opinion surveys in the Arab world. *British Journal of Political Science*. <https://doi.org/10.1017/S0007123419000206>.
- Gordoni, G., & Schmidt, P. (2010). The decision to participate in social surveys: The case of the Arab minority in Israel—An application of the theory of reasoned action. *International Journal of Public Opinion Research*, 22, 364–391.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: Wiley.
- Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly*, 50, 251–266.
- Hashemi, N., & Postel, D. (2017). *Sectarianization: Mapping the new politics of the Middle East*. New York: Oxford University Press.
- Heckman, J. J., Hidehiko, I., & Todd, P. E. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65, 261–294.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960.
- Hue, P. T., & Sager, E. B. (1975). Interviewer’s nationality and outcome of the survey. *Perceptual and Motor Skills*, 40, 907–913.

- Hyman, H. H., Cobb, W. J., Feldman, J. J., Hart, C. W., & Stember, C. H. (1954). *Interviewing in social research*. Chicago: University of Chicago Press.
- Iacus, S. M., King, G., & Porro, G. (2009). CEM: Software for coarsened exact matching. *Journal of Statistical Software*, *30*, 1–27.
- Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, *106*, 345–361.
- Iacus, S., King, G., Porro, G., & Katz, J. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*, 1–24.
- Inglehart, R., Moaddel, M., & Tessler, M. (2006). Xenophobia and in-group solidarity in Iraq: A natural experiment on the impact of insecurity. *Perspectives on Politics*, *4*, 495–505.
- Kao, K. & Revkin, M. (2018). To punish or to pardon? Reintegrating rebel collaborators after conflict in Iraq. Working paper no. 17. University of Gothenburg, Sweden. <https://gld.gu.se/media/1503/gld-working-paper-17-final.pdf>.
- King, G., & Nielsen, R. (2016). Why propensity scores should not be used for matching. Working paper at <http://j.mp/1sexgVw>. Retrieved March 8, 2019.
- Le, K. T., McRoy, M., & Diop, A. (2018). Randomization in small sample surveys with the Halton sequence. *Survey Research Methods*, *12*, 247–257.
- Longva, A. (2006). Nationalism in pre-modern guise: *Hadhar* and *Badu* in Kuwait. *International Journal of Middle East Studies*, *38*, 171–187.
- Luciani, G. (1987). Allocation vs. production states. In H. Beblawi & G. Luciani (Eds.), *The rentier state*. London: Croom Helm.
- Lust, E. (2009). Competitive clientelism in the Middle East. *Journal of Democracy*, *20*, 122–135.
- Lyberg, L. E., & Stukel, D. M. (2017). The roots and evolution of the total survey error concept. In P. P. Biemer, et al. (Eds.), *Total survey error in practice*. New York: Wiley.
- Mahalanobis, P. C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, *109*, 325–378.
- Mitchell, J. S., & Gengler, J. (2019). What money can't buy: Wealth, inequality, and economic satisfaction in the rentier state. *Political Research Quarterly*, *72*, 75–89.
- Mneimneh, Z. N., De Jong, J., Cibelli, K. H., & Moaddel, M. (2018). Do I look and sound religious? Interviewer religious appearance and attitude effects on respondents' answers. *Journal of Survey Statistics and Methodology*. <https://doi.org/10.1093/jssam/smy020>.
- Mneimneh, Z. M., et al. (2015). Cultural variations in the effect of interview privacy and need for social conformity on reporting sensitive information. *Journal of Official Statistics*, *31*, 673–697.
- Okruhlik, G. (2011). Dependence, disdain, and distance: State, labor, and citizenship in the Arab Gulf States. In J. F. Seznec & M. Kirk (Eds.), *Industrialization in the Gulf: A socioeconomic revolution*. New York: Routledge.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598–609.
- Potter, L. G. (Ed.). (2014). *Sectarian politics in the Persian Gulf*. Oxford: Oxford University Press.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.
- Rubin, D. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169–188.
- Rubin, D., & Thomas, N. (2000). Combining propensity score matching with adjustments for prognostic covariates. *Journal of the American Statistical Association*, *95*, 573–585.
- Samii, C. (2013). Perils or promise of ethnic integration? Evidence from a hard case in burundi. *American Political Science Review*, *107*, 558–573.
- Sirhan, B. (1975). Palestinian refugee camp life in Lebanon. *Journal of Palestine Studies*, *4*, 91–107.
- Smith, T. (2019). Improving multinational, multiregional, and multicultural (3MC) comparability using the total survey error (TSE) paradigm. In T. P. Johnson, et al. (Eds.), *Advances in comparative survey methods*. New York: Wiley.
- Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science*, *25*, 1–21.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of social conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations*. Monterey, CA: Brooks/Cole.
- Tessler, M. A. (2011). *Public opinion in the Middle East: Survey research and the political orientations of ordinary citizens*. Bloomington: Indiana University Press.

- Tessler, M., Jamal, A., Bedaida, A., Abderebbi, M., Shikaki, K., Braizat, Gengler, J., & Robbins, M. (2016). Arab Barometer: Public opinion survey conducted in Algeria, Morocco, Jordan, Lebanon, Palestine, Yemen, and Bahrain 2006–2009. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2016-02-26.
- The Syria Campaign. (2015). Survey amongst Syrian refugees in Germany. <http://www.adoptrevolution.org/en/survey-amongst-syrian-refugees-in-germany-backgrounds>.
- West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, 5, 175–211.
- Winkler, N., Kroh, M., & Spiess, M. (2006). Entwicklung einer deutschen Kurzsкала zur zweidimensionalen Messung von sozialer Erwünschtheit. DIW discussion paper no. 579. Berlin: German Institute for Economic Research.
- Yemen Polling Center. (2017). *Perceptions of the yemeni public on living conditions and security-related issues: Survey findings*. Sanaa: Yemen Polling Center.
- Zogby, J., et al. (2012). Looking at Iran: How 20 Arab and Muslim nations view Iran and its policies. Washington, D.C.: Zogby Research Services, LLC. <https://www.wilsoncenter.org/sites/default/files/iranpollresultsreport.pdf>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Justin J. Gengler¹ · Kien T. Le¹ · Jill Wittrock²

Kien T. Le
kienle@qu.edu.qa

Jill Wittrock
jillwittrock@gmail.com

¹ Social and Economic Survey Research Institute, Qatar University, Doha, Qatar

² Harman Connected Services, Inc., Bellevue, WA, USA