

Appendix 1. Arabic versions of question wording and their English translations

1) Trust (generalized) AB III, AB V:

بشكل عام، هل يمكن الثقة في أغلبية الناس أم أنه لا يمكن
الثقة بهم؟ اقرأ

In general, can the majority of people be trusted or not. Can you trust them? (Read)

WVS 6, WVS 7:

بشكل عام، هل ترى أن أغلب الناس يمكن الثقة بهم، أم أنك تحتاج للحذر الشديد في التعامل مع الناس؟ اقرأ

In general, do you think that most people can be trusted, or do you need extreme caution in dealing with people? (Read)

In both questions the same word *ath-thiqa* (something in between confidence and trust in English) is used in both)

2) Trust to police AB III, AB V:

سوف أقوم بتسمية مجموعة من المؤسسات، وأود أن تخبرني إلى أي درجة تثق في كل واحدة منها: الأمن العام الشرطة لبنان: قوى الأمن الداخلي الشرطة

I will name a group of institutions, and I would like you to tell me how much you trust each one: Public Security the police. Lebanon: The Internal Security Forces the police WVS 6, WVS 7:

لى أي مدى تثق بكل من المؤسسات التالية؟ ثقة كبيرة جداً، أم ثقة كبيرة، أم ثقة قليلة، أم أنك لا تثق بها أبداً؟ للباحث: تسأل البنود المدرجة بالأسلوب نفسه الشرطة

To what extent do you trust each of the following institutions? Too much trust, too much confidence, or too little trust, or you never trust it, to what extent do you trust the following institutions? To the researcher: You ask the listed items in the same way: the police Again, the same keywords, *ath-thiqa* (trust or confidence) and *asshurta* (police) are used in both questions

3) University education is more important for boys:

AB III, AB V:

التعليم الجامعي للأولاد أهم من التعليم الجامعي للبنات

University education for boys is more important than university education for girls

WVS 6, WVS 7:

التعليم الجامعي مهم للولد أكثر من البنت

University education is more important for a boy than a girl

These two questions look almost identical in terms of wording and gist

4) Men are better political leaders

AB III, AB V:

بصفة عامة، للرجال أفضل، في تولي القيادة السياسية من النساء

In general, men are better at political leadership than women WVS 6, WVS 7:

بشكل عام، الرجل أفضل كقائد سياسي من المرأة

In general, a man is better as a political leader than a woman These two questions are also very similar in wording

Appendix 2. Detailed description of analytic strategy

First step. Analytically, first, we compare the univariate distributions of the four attitudes between the AB and the WVS (cf. Brown, Micklewright, Schnepf, and Waldmann, 2007; Hayford and Morgan, 2008; Kieffer, 2010; Vandemplas and Lipps, 2014). For each issue in each country, we assess the comparability of a) average support (non-statistically significant differences in mean values); and b) diversity in stances (non-statistically significant differences in standardized Simpson's D values) (Simpson, 1949). We then assess the magnitude of differences between the AB and the WVS data, using Duncan's Dissimilarity Index (DDI), which runs from 0, indicating the distributions are completely equal, to 1, indicating all respondents have to change their answers to achieve equality between the WVS and the AB (Duncan and Duncan, 1955; see also Ortmann and Schneider, 2016; Schneider, 2009). Next, our multivariate analyses address whether relations between demographics and the four attitudes differ between the AB and the WVS. Similarity in relations is pivotal, because public opinion scholars are often more interested in multivariate relations than in univariate descriptives. We focus on relations with demographics because demographics represent important societal divisions and are often included in public opinion scholars' models. Additionally, demographics probably suffer less from survey error, for instance due to less socially desirable answering and lower cognitive strain on respondents. Analytically, we estimate regression analyses using gender, age, educational attainment, marital status, and employment status as predictors, and the attitudes as the dependent variables. We include a moderation term with survey type (WVS, coded 0, or AB, coded 1) that shows whether a certain demographic has a significantly different effect on a particular attitude in the AB than in the WVS (see also Glas, Spierings, Lubbers, and Scheepers, 2019; Kahn, Kalsbeek, and Hofferth, 1988; Swicegood, Morgan, and Rindfuss, 1984).

Second step. In the second step, we re-estimate the DDIs and regression models on AB and WVS samples that have been made equal by weighing them by each other. We weighted the AB data by the WVS demographic distributions (although for our purposes it in fact does not matter which survey is weighted by which one, because we are not trying to approximate the “most like the population” sample, but an intersurvey equal sample). We cannot stress that point enough; we are not using weights in the conventional sense to let our surveys approximate the “real” populations in Arab countries. Our weights do not show that at all. The reason we weigh our surveys is to let them approximate not the “actual” Arab populations, but each other. We weighted the AB surveys so that their combinations of gender * age * education * marital status * employment status (48 strata) are the same as the WVS’s combinations of gender * age * education * marital status * employment status (again, obviously, 48 strata). Gender here distinguishes between two categories: men and women. Age is categorized in three categories: under 30, 30 to 49, and 50 and over. Education is dichotomized: secondary or lower versus tertiary. Marital status is also dichotomized: married versus single or other. Employment status refers to employed versus nonemployed.

Democracy index. The extent to which a particular Arab country in a particular year is democratic was measured by Freedom House’s Democracy Index. The values of this index for our sample are: Algeria ’13-’14 5,5; Egypt ’13 5; Egypt ’18 6; Iraq ’13 6; Iraq ’18 5,5; Jordan ’13-’14 5,5; Jordan ’18 5; Kuwait ’14 5; Lebanon ’13 4,5; Lebanon ’18 5; Libya ’14 4,5; Morocco ’11-’13 4,5; Palestine ’12-’13 5,75; Tunisia ’13 3,5; Tunisia ’18-’19 2,5; Yemen ’13-’14 6. Because in this index, higher scores indicate less democracy, we reversed them to ease intuitive interpretations of our analyses.

References

- Brown, G., Micklewright, J., Schnepf, S. V., & Waldmann, R. (2007). Cross-national surveys of learning achievement: How robust are the findings? *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, *170*(3), 623–646. doi:10.1111/j.1467-985X.2006.00439.x
- Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, *20*(2), 210–217. doi:10.2307/2088328
- Glas, S., Spierings, N., Lubbers, M., & Scheepers, P. (2019). How politics shape support for gender equality and religiosity’s impact in arab countries. *European Sociological Review*, *35*(3), 299–315. doi:10.1093/esr/jcz004
- Hayford, S. R., & Morgan, S. P. (2008). The quality of retrospective data on cohabitation. *Demography*, *45*(1), 129–141. doi:10.1353/dem.2008.0005
- Kahn, J. R., Kalsbeek, W. D., & Hofferth, S. L. (1988). National estimates of teenage sexual activity: Evaluating the comparability of three national surveys. *Demography*, *25*(2), 189–204. doi:10.2307/2061288
- Kieffer, A. (2010). Measuring and comparing levels of education: Methodological problems in the classification of educational levels in the european social surveys and the french labor force surveys. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *107*(1), 49–73. doi:10.1177/0759106310369974
- Ortmanns, V., & Schneider, S. L. (2016). Harmonization still failing? inconsistency of education variables in cross-national public opinion surveys. *International Journal of Public Opinion Research*, *28*(4), 562–582. doi:10.1093/ijpor/edv025
- Schneider, S. (2009). *Confusing credentials: The cross-nationally comparable measurement of educational attainment* (Doctoral dissertation). doi:10.13140/RG.2.2.35997.51683
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688–688. doi:10.1038/163688a0
- Swicegood, C. G., Morgan, S. P., & Rindfuss, R. R. (1984). Measurement and replication: Evaluating the consistency of eight u. s. fertility surveys. *Demography*, *21*(1), 19–33. doi:10.2307/2061024
- Vandenplas, C., & Lipps, O. (2014). Robustness of items within and across surveys. *FORS working paper series*.