The SOM Institute's Notes on Survey Methodology – 2023:5

The implications of offering questionnaires in multiple languages

Sebastian Lundmark
sebastian.lundmark@som.gu.se
The SOM Institute, University of Gothenburg

Frida Sandelin
frida.sandelin@som.gu.se
The SOM Institute, University of Gothenburg

Björn Rönnerstrand
bjorn.ronnerstrand@som.gu.se
The SOM Institute, University of Gothenburg

ABSTRACT

Lowering the thresholds to complete a questionnaire is thought to increase sample persons' willingness to complete questionnaires. However, this note presents results of an experiment indicating that offering multiple languages in which to complete a questionnaire did not increase response rates. In a self-administered mailed survey, 6,908 sample persons where randomly assigned to be offered to complete the questionnaire in Swedish, English, Arabic, Bosnian/Serbian/Croatian, Dari, Kurmanji, Persian, Polish, Somali, Sorani, Turkish, and Tigrinya, whereas 2,092 respondents were only offered to complete it in Swedish. The results showed that response rates were not improved by offering the questionnaire in multiple languages, and offering sample persons to complete the questionnaire in the majority language spoken in their birth country did not increase their willingness to do so. Offering questionnaires in multiple languages might be preferred on grounds other than response rates, for example, as a symbol of inclusion. Yet, as the results of this note indicate, this may come at the cost of reducing the willingness to participate among some groups.

Reviewed by Ph.D. Patrik Öhberg

INTRODUCTION

The risk of nonresponse bias increases rapidly when a survey request or questionnaire does not facilitate ease of access for eligible respondents to complete the interview. A survey request or questionnaire administered only in languages an individual is not proficient in, should likely minimize the likelihood of that individual participating. If groups of sample persons eligible to be interviewed cannot understand the questions being asked or even what task they are being asked to complete (i.e., participating in a study by filling out a questionnaire), point estimates and statistical inferences will likely be skewed and biased (Groves & Lyberg, 2010).

Offering multiple languages in which to complete a questionnaire may, therefore, be a potential way to facilitate questionnaire completion and increase response propensities, especially among such already hard-to-survey groups as those not born in the country being surveyed. This might be especially important for questionnaires administered in Sweden, since response rates among people born outside the Nordics have decreased more rapidly compared to among other groups, leading to the likely growth of non-response bias (Lundmark and Backström, 2023). This becomes potentially even more worrisome given that the proportion of foreign-born individuals almost doubled between 2000 and 2021 (from 12% to 20%, Statistics Sweden, 2022).

This note describes the results of an experiment where a sample of 9,000 individuals in Gothenburg, Sweden, were randomly assigned to be able to complete a self-administered push-to-web questionnaire (online or by paper-and-pencil) in Swedish only or in Swedish and eleven other languages.

HYPOTHESIS

Three hypotheses were assessed:

RESPONSE RATES

H1: Sample persons who are offered to complete the questionnaire in several languages may be more likely to complete the questionnaire than sample persons who are offered to complete the questionnaire only in Swedish.

NON-RESPONSE BIAS

H2: Sample persons who are offered to complete the questionnaire in several languages may show less nonresponse bias than sample persons who are offered to complete the questionnaire only in Swedish.

DATA QUALITY

Completing a questionnaire in a language in which the respondent is proficient may lead to more precise and reliable responses and reduce respondent burden. Therefore, individuals offered to complete a questionnaire in more than Swedish may increase the quality of the data obtained compared to when restricting all individuals to complete it in Swedish.

H3. Sample persons who are offered to complete the questionnaire in several languages may be more likely to produce better data quality than sample persons who are offered to complete the questionnaire only in Swedish.

EXPLORING THE LANGUAGES CHOSEN

In addition to the hypotheses, individuals born in a country where one of the offered languages is a majority language may be more likely to complete questionnaires in that language than in Swedish. The present note, therefore, also presents descriptive statistics of whether individuals from specific countries were more likely to choose to complete the questionnaire in the majority language of their country of birth.

METHODS AND MATERIALS

PREREGISTRATION

The hypotheses, procedure, exclusion criteria, and analysis plan were preregistered prior to data collection had completed and prior to analyses. The preregistration can be found at https://osf.io/p6vky.

SAMPLE

A sample of 9,000 randomly selected individuals registered by the Swedish Tax Authority as residing in the city of Gothenburg was drawn on August 9, 2022. Only individuals 16-90 years old were invited to complete the questionnaire. The questionnaire was administrated by the SOM Institute at the University of Gothenburg.

PROCEDURE

SAMPLING THE LANGUAGES

The chosen languages were in part based on the languages spoken in the most common countries of birth among the citizens in Gothenburg, and in part based on an evaluation of historic response rates of people originating from these countries. Hence, although some countries were common places of birth for relatively large shares of the Gothenburg population, languages spoken in those countries were not necessarily chosen. The response rate among sample persons born in these countries had to be low enough to indicate the potential to improve response propensities. For example, previous data collections have shown that sample persons born in Germany were just as likely to complete questionnaires as those born in Sweden. Moreover, individuals born in India and China who live in Gothenburg tended to have a high level of education. The assumption was made that such individuals were likely to be fluent in English, so the primary languages spoken in these countries were not selected to be included. In addition, consultations were made with organizations and researchers with expertise in languages before the final decisions on which languages to include. In the end, the questionnaire was offered to be completed in Swedish, English, Arabic, Bosnian/Serbian/Croatian, Dari, Kurmanji, Persian, Polish, Somali, Sorani, Turkish, and Tigrinya.

QUESTIONNAIRE TRANSLATION

First, a translation company was instructed to translate the questionnaire into English. This translation was reviewed and edited by researchers and analysts at the SOM Institute. Thereafter, the translators at the translation company were given both the Swedish and the English versions of the questionnaire together with some additional information to be translated.[1] The translated questionnaires and information were then reviewed and edited by external reviewers proficient in the translated language and who had expertise in social science and/or survey methodology. The translation process and external review of the translations were intended to ensure that the quality of the questionnaire was just as high in Swedish as in the other offered languages.

[1] That is, a letter with brief information in all 11 non-Swedish languages included in the invitation and all postal reminders and frequently asked questions (FAQ) published on the web, referenced in the letter of information.

EXPERIMENTAL DESIGN AND PROCEDURE

Prior to being invited to complete the questionnaire, each sample person was randomly assigned to one of two groups using random numbers extracted from random.org. One group (called *the language group*) (n = 6,908) was offered to complete the questionnaire in the 12 different languages using the online questionnaire mode. The other group (called the *Swedish-only group*) (n = 2,092) was offered to complete the questionnaire only in Swedish. In both experimental groups, the primary language of information given was Swedish. To enable analyses between the chosen languages, the probability of being assigned to the language group was higher (seven in nine) than to the Swedish-only group (two in nine). The total number of survey questions was 61.

All sample persons were sent a pre-notification by postcard one week prior to being invited to complete the questionnaire. The sample persons assigned to the language group received information, in Swedish, on the card about their choice to complete the questionnaire in 12 different languages. In the first invitation, all sample persons were offered to complete the questionnaire through an online self-administered survey (a so-called push-to-web approach). In the Swedish-only group, a separate A4 paper informed sample persons in nine languages of what the questionnaire invitation would entail but they were not given the option to complete the questionnaire in other than Swedish.[2] For the language group, a separate A4 paper informed sample persons that they could complete the questionnaire in eleven additional non-Swedish languages. This A4 single page contained short general information about the study and that the questionnaire could be completed in one of the 12 languages. It also included a link to a webpage at the SOM Institute with a FAQ (frequently asked questions) in each non-Swedish language. The same FAQ in Swedish was included in the invitation to all sample persons of both groups. Each FAQ webpage included a link to the online questionnaire. Individuals in the language group were sent this separate A4 page in each mailed reminder. Moreover, the invitation and the following reminders informed the sample persons in the language group that they could, at any time, request a paper-and-pencil version of the questionnaire in one of the 12 languages.

All sample persons were sent a postcard with an envelope nine days after the first invitation. The postcard thanked respondents and reminded non-respondents to

[2] This A4 single page has been included in all initial survey requests mailed by the SOM surveys since 2019. The languages are English, Arabic, Bosnian/Serbian/Croatian, Dari, Persian, Somali, Sorani, and Turkish.

complete the questionnaire. Regardless of whether participants had completed the survey or not, everyone was both thanked and reminded. The postcard contained information on how to respond to the online survey. The postcard informed the individuals in the language group, in Swedish, that they had the option to fill out the questionnaire in 12 different languages.

Seventeen days after the initial invitation, all sample persons' who had not returned their questionnaire received a reminder in the mail. This reminder included both a paper-and-pencil version of the questionnaire, in Swedish, and an envelope for its return, along with instructions on how to complete the survey online.

Twenty-eight days after the initial invitation, sample persons who had not completed their questionnaire received a text reminder in Swedish on their mobile phones. This message included a link to the online survey. Sample persons in the language group were notified, in Swedish, that they could choose from 12 different languages to complete the questionnaire. Sample persons' who had not responded and had not declined participation were sent four mailed reminders and four text reminders on their mobile phones with a link to the online survey, making it a total of nine reminders. The field period for the survey ended on day 109.

UNPLANNED DEVIATION FROM PREREGISTRATION

An unplanned deviation from the preregistered procedure occurred during the first 24 hours of questionnaire administration when the sub-contractor (responsible for programming the web questionnaire, printing, and mailing the physical questionnaires and letters of information) erroneously offered all foreign-born sample persons to complete the questionnaire in all 12 languages. Before the deviation was detected, 219 individuals entered the questionnaire. As a result, these individuals were removed from the analysis. The available sample persons for analyses ended up at 6,754 in the language group and 2,027 in the Swedish-only group.

ANALYSIS PLAN

RESPONSE RATE

To compare response rates between the groups, Response Rate 1 (RR1) was estimated according to the guidelines of the American Association for Public Opinion Research (AAPOR, 2016). Only sample persons who answered 80% or more of the eligible questions were considered a response. The parameters of an

OLS regression equation projected the difference in RR1 between the treatment and control groups.

NON-RESPONSE BIAS

Non-response bias was assessed through R-indicators. The R-indicator is an expression of the standard deviation of probabilities of responses of units. The R-indicators were estimated using the R-code provided on the web page of the Representativity Indicators for Survey Quality based on registered sample data. This data considered sample persons' sex, age cohort (18-29, 30-39, 40-49, 50-64, 65-74, and 75-90), marital status (married, not married), and birth country (born in the Nordics, not born in the Nordics but born in Europe, and born outside Europe). The higher the value of the R-indicator, the better representativity among respondents.

DATA QUALITY

Data quality was assessed by the proportion of questions a respondent chose to skip. Sample persons who answered at least one question were included in the estimations.

RESULTS

RESPONSE RATES

In contrast to the expected, offering to complete the questionnaire in 12 different languages did not increase response propensities (see Table 1).

Table 1. *Response rate (RR1) and the proportion of breakoffs, partials, and completes.*

	Response rate	*n*	Standard error	Breakoffs	Partials	Completes
Languages	42.02%	6,754	.01	4.17%	3.98%	91.84%
Swedish-only	43.36%	2,027	.01	4.86%	4.34%	90.81%
Difference	-1.34		.01	-0.69	-0.36	1.03

Note. Response rates were calculated in accordance with AAPOR 2016 standard (RR1). The share of responses was based on all questionnaires with at least one question answered.

Among sample persons offered to complete the questionnaire in multiple languages, 42 percent completed the questionnaire, which was not statistically significantly different from the 43 percent who completed it when offered to

complete it only in Swedish (Δ = -1.34%, $b_{languages}$ = -0.01, *SE* = 0.01, *p* = .28) (see Table 2).

Table 2. *Effect of offering multiple languages on response propensities (RR1).*

	Completed the questionnaire				
	Base model	**Sex**	**Age**	**Marital status**	**Immigrant status**
Languages (reference: Swedish)	-.01 (.01)	-.01 (.02)	-.00 (.03)	-.02 (.02)	-.02 (.02)
Demographic		.06 (.02)	.01 (.00)	.11 (.02)	-.16 [1] (.04) -.33 [2] (.03)
Languages * Female		-.01 (.03)			
Languages * Age (16-85)			-.00 (.00)		
Languages * Married (reference: not married)				-.00 (.03)	
Languages * Born inside Europe but outside the Nordics (reference: Born in the Nordics)					-.02 (.04)
Languages * Born outside Europe (reference: Born in the Nordics)					.04 (.03)
Constant	.43*** (.01)	.40*** (.02)	.19*** (.03)	.35*** (.01)	.51*** (.02)
Observations	8,781	8,781	8,781	8,781	8,781
R^2	.00	.00	.03	.01	.09

Note. [1] Refers to being born in Europe. [2] Refers to being born outside Europe. Unstandardized OLS regression coefficients with standard errors in parentheses.

*** *p* < .001.

Furthermore, the sex ($b_{languages * female}$ = -0.01, *SE* = 0.03, *p* = .59), age ($b_{languages * age\ continuous}$ = -0.00, *SE* = 0.00, *p* = .76), marital status ($b_{languages * married}$ = -0.00, *SE* = 0.03, *p* = .89), or

immigration status ($b_{languages * born\ in\ Europe}$= -0.02, SE = 0.04, p = .66; $b_{languages * born\ outside\ Europe}$ = 0.04, SE = 0.03, p = .17) of the sample persons did not affect the response propensity among those offered to complete the questionnaire in multiple languages compared to being offered to complete the questionnaire only in Swedish.

EVALUATION OF LANGUAGES

Among the 6,754 individuals in the language group, 2,838 completed the questionnaire, and 132 responded in a language other than Swedish (2 percent of the sample persons in the language group). The most common non-Swedish language utilized was English with 87 responses, and thereafter Arabic with 23 responses. Only a few respondents opted to complete the questionnaire in Persian, Polish, Bosnian/Serbian/Croatian, and Turkish. The questionnaire was not completed in any of the languages Dari, Kurmanji, Somali, Sorani, and Tigrinja. The countries of birth among the individuals who completed the questionnaire in English consisted of a total of 34 countries, where 4 respondents born in Sweden choosing to complete it in English. See Table 3 for an overview of the languages chosen by respondents and their country of birth.

Table 3. *Responses (RR1) in non-Swedish languages, observations, and country of birth.*

Language	Responses	Country of birth
English	87	India (11), United Kingdom (7), USA (7), China (6), Brazil (5), France (5), Germany (4), Syria (4), Sweden (4), Italy (3), Japan (3), Bangladesh (2), Greece (2), Mexico (2), Russia (2), Switzerland (2), Spain (2), Bolivia (1), Bulgaria (1), Canada (1), Ghana (1), Iran (1), Kosovo (1), Nigeria (1), North Macedonia (1), Norway (1), Poland (1), Romania (1), Somalia (1), South Africa (1), South Korea (1), Czech Republic (1), Turkey (1), Ukraine (1)
Arabic	23	Syria (8), Iraq (4), Lebanon (4), Egypt (2), Morocco (1), United Arab Emirates (2), Kuwait (1), Tunisia (1)
Persian	7	Iran (7)
Polish	7	Poland (7)
Bosnian/Serbian/Croatian	5	Serbia (2), Bosnia-Hercegovina (1), Jugoslavia (1), Macedonia (1)
Turkish	3	Türkiye (3)
Total	**132**	

A not preregistered exploratory analysis showed that among respondents born in a country where English was the primary language, 48 percent chose to complete the questionnaire in English. In contrast, only 27 percent of those born in a country with Arabic as the primary language chose to complete the questionnaire in Arabic, meaning that a large majority of those who completed the survey chose Swedish despite not being born in Sweden. Respondents born in the other countries with a corresponding primary language followed a similar trend as those born in Arabic-speaking countries, where most opted to complete the questionnaire in Swedish despite being offered another language (see Table 4). Another not preregistered exploratory analysis showed that being born in a country where English was the primary language increased the probability of completing the questionnaire in any of the non-Swedish languages compared to being born in a country where any of the other languages was a primary language ($b_{languages}$ = -0.25, SE = 0.06, p < .001) (see Appendix, Table A1).

Table 4. *Responses (RR1) in non-Swedish languages among sample persons' whose birth country has the language as a major language.*

A primary language	The share of responses in a language of those whose birth country has that language as a primary language
English [1]	48% (16 out of 33)
Arabic [2]	27% (21 out of 78)
Persian [3]	12% (7 out of 57)
Polish [4]	30% (7 out of 23)
Bosnian/Serbian/Croatian [5]	5% (3 out of 58)
Turkish [6]	20% (3 out of 15)

Note. [1] A major language in Australia, Ireland, Canada, New Zealand, United Kingdom, South Africa, USA. [2] A major language in Algeria, Bahrain, Chad, Djibouti, Egypt, Jordan, United Arab Emirates, Iraq, Yemen, Comoros, Lebanon, Libya, Mauritania, Morocco, Oman, Qatar, Saudi Arabia, Syria, Tunisia, State of Palestine. [3] A major language in Iran, Afghanistan, Pakistan, and Tajikistan. [4] A major language in Poland. [5] A major language in Bosnia and Herzegovina, North Macedonia, Montenegro, and Serbia. [6] A major language in Türkiye.

NON-RESPONSE BIAS

Despite not increasing response propensities, offering sample persons to respond in multiple languages decreased nonresponse bias. Respondents in the language group were more alike nonrespondents (R-indicator = .69, 95% CI[0.68, 0.71]) than respondents were in the Swedish-only group (R-indicator = .67, 95% CI[0.64, 0.68]) (the confidence intervals of the two R-indicators did not overlap) (see Table 5).

Table 5. *R-indicators and 95% CI for the language group versus the control group.*

	Languages	Swedish-only
R-indicator	0.69 (0.01)	0.67 (0.02)
Confidence interval	[0.68, 0.71]	[0.64, 0.68]

Note. Standard errors in parenthesis.

However, the R-indicators statistically significantly differed due to unwanted reasons, sample persons older than 29 years became less likely to complete the questionnaire when being offered multiple languages leading to slightly lower response rates overall than greater response rates.

DATA QUALITY

The data quality was not affected by offering to complete the questionnaire in multiple languages. Sample persons who were offered to complete the questionnaire in multiple languages were not less likely to skip questions than sample persons offered to complete it only in Swedish ($b_{languages}$ = 0.01, SE = 0.01, p = .36). Restricting the analyses to only sample persons born in a country where one of the offered languages was a primary language did not alter this result ($b_{languages}$ = 0.01, SE = 0.01, p = .69).

Table 6. *Effects of item non-response (OLS regression coefficients).*

	Full sample	Language of birth country sample [1]
Languages (reference: Swedish-only)	.01 (.01)	.01 (.02)
Constant	.93*** (.01)	.86*** (.02)
Observations	4,058	453
R^2	.00	.00

Note. [1] The sample includes only persons born in a country where one of the offered languages is a primary language.

*** p < .001.

CONCLUSION

The current pre-registered experimental study was designed to test the effect of offering a survey questionnaire in multiple languages on response rates, non-response bias, and data quality. A descriptive aim of the study was to explore whether sample persons from certain countries had a higher tendency to respond in languages other than Swedish.

The most striking finding was that translating the Swedish questionnaires into eleven additional languages did not increase the overall response rate. Thus, while the response rate has tended to be lower among people born outside the Nordics, reducing the language barrier by offering the survey request and the questionnaire in multiple languages did not, in itself, increase the response propensities of these groups. It appears as if more languages are not a silver bullet when it comes to countering declining response rates or the lower response rates among people born outside the Nordics.

The study made use of a "push-to-web" approach. In the first invitation, respondents were offered the opportunity to participate via an online self-administered survey. Respondents who wanted to answer "non-digitally" in a language other than Swedish could request a paper survey to fill in. Yet not a single such request was made. The process of requesting a paper survey is likely a barrier for (older) people not proficient in Swedish but who would prefer to answer via paper and pencil.

Furthermore, offering the questionnaire in multiple languages did not affect data quality. There were no differences between the experimental groups in the tendency to skip questions. Yet, there were differences in terms of non-response biases, where the differences between the responding sample and the nonresponding sample were smaller in the language group. This was because individuals older than 29 were less likely to complete the survey if multiple languages were offered. Hence, respondents and nonrespondents became more similar because older (easy-to-survey) respondents seemingly shied away from completing the questionnaire when being offered multiple languages.

This points to a dilemma for survey researchers and practitioners. Offering questionnaires in multiple languages might be preferred on grounds other than response rates, for example, as a symbol of inclusion. Yet, this may come at the cost of reducing the willingness to participate among some groups.

Our analysis of language use among people from different countries revealed that most respondents with an origin from a country where one of the languages was spoken still preferred to answer in Swedish. To illustrate, only 3 out of 58 responses from people from the Balkans were submitted using the survey in Bosnian/Serbian/Croatian. Comparatively, forty-eight percent of all respondents born in a country where English was a major language completed the survey in English.

One could speculate that the reason for these results may be that people born in English-speaking countries may have less incentive to learn Swedish compared to those born where other languages dominate, simply because of the Swedish population's proficiency in speaking and writing in English.

REFERENCES

Bergquist, J., Falk, E. & Weissenbilder, M. (2023). *SOM-undersökningen i Göteborg 2022 – En metodöversikt.* SOM-rapport nr 2023:6. University of Gothenburg. The SOM Institute.

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly, 74*(5), 849-879.

Statistics Sweden. (2022). *Inrikes och utrikes födda efter region, ålder och kön. År 2000 - 2022.* Accessed on May 18, 2022 at https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101E/InrUtrFoddaRegAlKon/table/tableViewLayout1/

APPENDIX

Table A1. *The effect of being born in a birth country where a language is a primary language on the likelihood of responding in a non-Swedish language (OLS regression coefficients).*

	Response (RR1) in a non-Swedish language	
	Language countries binary	**Language countries multinominal**
Languages countries of birth[1] (ref: English)	-.25*** (.06)	
Language country: Arabic (ref: English)		-.15* (.07)
Language country: Persian (ref: English)		-.30*** (.07)
Language country: Polish (ref: English)		-.15* (.09)
Language country: Bosnian (ref: English)		-.36*** (.07)
Language country: Turkish (ref: English)		-.20** (.13)
Constant	.40*** (.06)	.40*** (.06)
Observations	365	365
R^2	.04	.04

Note. [1] Variable coded 1 for language country Arabic, Persian, Polish, Bosnian/Serbian/Croatian and Turkish, and 0 for English.

* $p < .05$, ** $p < .01$, *** $p < .001$.

The SOM Institute is an academic organization located at the University of Gothenburg, Sweden. Since 1986 the SOM Institute conducts annual cross-sectional surveys among the Swedish population with a focus on Society, Opinion, and Media, as well as administering the web panel called the Swedish Citizen Panel. The annual surveys and the web panel both function as infrastructures, enabling researchers and public organizations to effectively collect research and opinion data in collaboration with researchers at the SOM Institute.

In order to strengthen contemporary research on Swedish society, as well as to contribute to international methodological development, the SOM Institute frequently publishes notes on methodological research. These are made available for scholars and the public alike at https://www.gu.se/en/som-institute/publications/research-on-survey-methodology

For more information, please contact us at info@som.gu.se