Technical report – online survey within the project
Studying Opinions and Populations in Online Text Data
Version 1 (2021-01-08)
Structure according to Schaurer, Kunz, & Heycke (2020)

1. General information

Commented [DB1]: Blue highlights: potential aspects to document (Schaurer et al. 2020)
Yellow highlights: to be added

Study topic, objectives, and description of research

The survey was part of the larger project *Studying Opinions and Populations in Online Text Data* (https://www.gu.se/en/linguistic-explorations-of-societies/research/research-projects). The survey's prime objective was to use open-ended survey questions to explore the understanding of and satisfaction with democracy in 10 countries (incl. 11 language versions).

_ Is the study part of a larger project?
_ Which overarching research questions guided the design of the study?
_ Which topics were included in the questionnaire?
_ Does the study have a specific title?

Researchers and their affiliation (Principal investigators, Project team members)

Stefan Dahlberg, professor at the Department of Humanities and Social Sciences at Miduniversitetet (Mid Sweden University) is the principal investigator of the project. Stefan Dahlberg and Dorothée Behr (GESIS – Leibniz Institute for the Social Sciences) were both involved in setting up this survey.

_ Who are the principle investigator(s) of the project?
_ Which project team members were actively involved?
_ Who is the contact person?

Study funding

The project is funded by the Swedish Research Council.

_ How was the study funded? Is there a contracting authority for the study?

Survey organization

The survey was contracted to Mantap Global AB (https://www.mantapglobal.com/?lang=en) following a tendering procedure. The company was responsible for programming the survey in the different languages, for correcting issues in the programming (following a technical pre-test in the different languages), for all aspects of fieldwork, and for delivering the final data set.

_ Who was responsible for the implementation of the survey? Did the project team conduct the survey by themselves or was an external agency contracted? If applicable, provide the name of the agency.
_ Were there other subcontractors? Provide the names and the services they provided.

Fieldwork time

The survey started xx.xx.2020 and was closed on xx.xx.2020. The following particularities occurred:

1) The original education variable differentiated between high and low education whereby high education was defined as qualifications allowing access to tertiary education (equivalent to Abitur, A-level etc.), and low as everything below. Notably in Estonia, Russia, Hungary, Poland, and Italy, the lower education category proved problematic, however. Jointly, with the panel provider, we agreed on the following solution: change of quota definition, whereby high education was subsequently defined as university studies or not (qualification of tertiary education). This equally meant an increase in respondents beyond n = 2004 per language version (every observation beyond the agreed 2004 respondents were charged, but with a 50% reduction). More information on the education variable can be found in the Excel "Education variable."
2) For Estonia (Russian language), it was not possible to reach the expected quotas, neither according to the first education definition not according to the second education definition. So we eventually stopped data collection.

_ Start and end date of the survey/ total duration in days or weeks.
_ If there were several waves, what was the time interval between the waves?
_ Are there any peculiarities that are additionallyworth mentioning, e.g., so_launch of the survey or
longer breaks?

Target population

The target population was defined as persons aged 18-65 with Internet access in Estonia (both Estonian- and Russian-speaking), France, Germany, Hungary, Italy, Poland, Russia, Spain, USA and Great Britain (10 countries, 11 language versions).
_ Whatwas the target population (e.g., Germanadult population, students at aparticular university)?

Sampling design, sample source, and provider

The survey was a quota survey, based on the quota scheme shown below (Table 1). For more information on the education variable underlying this scheme, please consult the Fieldwork section in this report. Mantap Global AB and international partners sampled the respondents for this study using non-probability access panels.

Quota	Sex	Men	Women	Men	Women
	Education	high	high	low	low
Age	18-30	167	167	167	167
	31-50	167	167	167	167
	51-65	167	167	167	167

Table 1: Expected sample per language version

_ How were the respondents sampled? Were they reached by a non-probability approach (e.g., quota sampling, snowballing) or a probability-based approach (e.g.,

random sample based on a population or employees register)? Or was it a Census survey? For further examples of online panel
sampling please see Baker et al., 2010, p. 719_.
_ How was the sampling frame defined? This could be, for instance, all members of a specific online panel that use a mobile device and are between 18 and 60 years old, or all students of a certain university that are reachable via email address.
_Who provided the sample? Was it an online (access) panel provider, were the addresses collected
at first hand, or did the sample come from another source?
_ In case of online (access) panel: Add some key information on the panel size, panel structure, panel
recruitment, panel maintenance, panel usage, and panel management if available. Also mention
the name of the panel (provider).

Sample size and response metrics

We aimed for n = 2004 respondents in each language version. The completion looks as follows:
Estonia, Estonian language
Estonia, Russian language
France,
Germany,
Hungary,
Italy,
Poland,
Russia,
Spain,
USA
Great Britain

This section should provide an overview of the key numbers that describe the characteristics and the
success of the survey. Not all of the numbers are available for all kinds of settings. We propose to report as much information as possible and anticipate before data collection which information is central and list them in a potential contract with the panel provider. The following information is central to judge the quality of a study:
_ Number of invited participants
_ Number of started surveys
_ Number of break-o_ (if applicable: by relevant subgroups)
_ Number of completed surveys
_ The numbers above are the basis for calculating response metrics that inform about the study's success.
For samples that are based on probabilistic sampling methods, the calculation of response

rates based on the AAPOR standards is possible (American Association for Public Opinion Research,
2016; DiSogra & Callegaro, 2016). For samples that are not based on a probability sample, at least
the completion rate can be reported (Callegaro & DiSogra, 2008). The completion rate is calculated
as the share of those who completed the online survey among all the eligible panel members who
were invited to the survey.
_ If a sample from an online access panel provider has been used, it should be ensured in advance
that all numbers required to calculate the response metrics are provided by the panel provider.
Especially the information about the number of invited panelists and the number and dates of reminder
e-mails is o_en not provided by default (see also section 3).

Survey mode

The survey was exclusively conducted online.
_ Was online the onlymodeorwas it a mixed-mode study? If yes, the specifics of a mixed-mode study
should be considered and reported (e.g., share of the di_erent modes, mode-specific response
rates).

Survey design

The survey was designed as a one-time survey.
_ Was the survey designed as a cross-sectional survey or as a longitudinal survey (trend or panel)?
_ Specify the number of waves, if more than one was conducted.

Sample characteristics

_ In addition to the response metrics, some basic information about the characteristics of the
final sample can be provided. This could be information about the distribution of basic sociodemographics, the proportion of smartphone respondents, and key variables of interest.
_ Are statements about the representation of the target population in terms of undercoverage and
overcoverage possible? If possible, report bias measures (see for example Biemer, 2010).

Data access for project participants and third parties for replication purposes

_ Are the data accessible for third parties, e.g., in a data archive, downloadable on a website, or will

they be provided on request?

2. Preparation

Questionnaire characteristics

The source questionnaire contains 26 questions for respondents who go through all filter conditions. The survey starts with 3 quota questions (sex, age, highest level of education). The subsequent main section contains 9 questions on politics and society, amongst which two open-ended questions related to the understanding of and the satisfaction with democracy. Furthermore, upon the request of the panel provider, this section also contains a test question to identify and filter out inattentive respondents. The remaining 14 questions refer to the background of the respondents (e.g., voting, partnership, citizenship) and the survey in general (2 feedback questions). In some countries, the maximum number of questions amounts to 28 questions, given that education or voting was covered with more than one question. We anticipated a 5-minute completion time. The actual completion time looks as follows:

Estonia, Estonian language
Estonia, Russian language
France,
Germany,
Hungary,
Italy,
Poland,
Russia,
Spain,
USA
Great Britain

The survey was fielded in the following languages: English (GB and USA), Estonian (Estonia), French (France), German (Germany), Hungarian (Hungary), Italian (Italy), Polish (Poland), Spanish (Spain), Russian (Estonia and Russia).

These language versions were produced as follows:

- A source version was produced in the English language; most items were taken as is or with slight changes (due to the online mode) from the source questionnaires of the European Social Survey (ESS) or the International Social Survey Programme (ISSP). Other items (open-ended questions) or texts (e.g. introduction, bridges) were newly produced for this survey.
- For the translation, this meant that existing translations from the ESS and the ISSP were adopted (and adapted in case of mode modifications or errors identified) whenever possible.
- Translators received an Excel with existing translations pre-inserted. Blue highlighted text in the source text column required a translation from scratch. This was new survey text produced by the project team. Yellow highlighted text required a modification (compared to the existing translation) following a

translation and adaptation guideline provided in a separate column (See Figure 1). Besides translating or modifying text, the translators were also instructed to review the existing (ESS/ISSP) translations and comment on these in case of gross errors and required corrections. The main goal, though, was to stay true to existing ESS/ISSP translations to allow a comparison of results with these studies (if needed).

- **Translation procedure:** While best practice calls for double translation and team reconciliation (Harkness, 2003), we carried out a simplified approach due to the fact that hardly any items required a new translation; after all, most of the items did already exist in a translated version. Against this backdrop, an approach employing three persons per language was deemed as more than needed and, given the few changed, not feasible. For each language, we recruited one professional translator, in most instances using the translator database of the BDÜ (https://suche.bdue.de/). We looked out for native speakers of the respective languages and a specialization in social sciences, market research and/or politics. Translators ideally had a background in survey translation, too, which could not always be assured. Before contacting potential translators, we also checked their websites to be able to identify suitability. The chosen translators received information on the study as well as item-specific translation and adaptation notes (see Figure 1). Upon delivery, the project team (Dorothée Behr) checked their translations (to the extent possible) and their comments. If needed, several iterations were used to complete the questionnaire. In particular, in some languages there were comments on the existing translations, which we then partly changed following these comments. Here, certainly, a second native speaker would have been helpful to better assess the need of corrections of existing ESS/ISSP translations. The project team passed on the final translation to the panel provider.
- **Testing procedure:** Once the panel provider had programmed the surveys in the different languages, the translators were asked to 1) modify and add to the translations (in case of source text changes) and 2) test the questionnaire for completeness and appropriate layout (e.g. appropriate line breaks) in the survey software. Since the survey provider had not yet programmed the routing at the testing stage, this could not be bested nor could questionnaire flow in these cases. Any feedback from the translators was transmitted to the panel provider via the project team, if needed, with several iterations. Even though this process in general worked out, not all feedback was correctly implemented by the provider. This suggests that this process needs to be optimized in future studies.

	A	B	C	F	G	H	
1	Text type	#	Source text	Translation /Adaptation guideline		France	Commer
44	RC		Country-specific categories				
45							
46	Intro		We now ask a few questions on politics and society.			Maintenant, quelques questions de politique et société.	
47							
48	Question	4	On the whole, how satisfied are you with the way democracy works in [country]?	* "And" at the beginning of the question should be deleted. * "Still use this card" should be deleted		Dans l'ensemble, êtes-vous satisfait(e) ou pas satisfait(e) de la manière dont la démocratie fonctionne en France?	
49	RC		Extremely dissatisfied			Pas du tout satisfait	
50	RC		Extremely satisfied			Tout à fait satisfait	
51	RC		*Don't know*			*Ne sait pas*	
52							
53	Question	5	Could you please motivate your answer? Please write down a few lines. The question was: "On the whole, how satisfied are you with the way democracy works in [country]?" Your answer was "[placeholder]" on a scale from 1 (extremely dissatisfied) to 11 (extremely satisfied)"	""motivate" - in the sense of: explain the reason(s) * Keep [placeholder] in the English language. * Translation from question 4 needs to be inserted here.		Pourriez-vous expliquer votre réponse en quelques lignes? La question était "Dans l'ensemble, êtes-vous satisfait(e) de la manière dont la démocratie fonctionne en [placeholder]? Vous avez répondu "[placeholder] sur une échelle de 1 (pas du tout	
54	RC						
55							
56	Intro		There are different opinions about immigrants from other countries living in [country]. By "immigrants" we mean people who come to settle in [country]. How much do you agree or disagree with each of the			Il existe différentes opinions concernant les immigrés venus d'autres pays pour vivre en France. Êtes-vous d'accord ou pas d'accord avec chacune des	
57	Question	6	Immigrants are generally good for [country's] economy.			Les immigrés sont généralement une bonne chose pour l'économie	
58	Question	7	[Country's] culture is generally undermined by immigrants.			En général, la culture française est menacée par les immigrés.	
59	Question	8	Immigrants increase crime rates.			Les immigrés font augmenter le taux de criminalité.	
60	RC		Agree strongly			Tout à fait d'accord	
61	RC		Agree			Plutôt d'accord	
62	RC		Neither agree nor disagree			Ni d'accord, ni pas d'accord	
63	RC		Disagree			Plutôt pas d'accord	
64	RC		Disagree strongly			Pas du tout d'accord	
65	RC		*Don't know*			*Ne sait pas*	

Figure 1: Translation Excel for French after translation step

- Special case: education variable (see Excel: Education variable)

One should give a comprehensive overview of all relevant aspects of the content and structure of the
questionnaire. Potential aspects are:
_ (Various) topics of the questionnaire
_ Sections/ modules of the questionnaire
_ The sequence of the topics
_ Overall number of questions
_ Routing information and information on major branches
_ The anticipated completion time and the actual average completion time
_ Language versions, translation
_ Number of questionnaire versions and the di_erences between the versions
_ Multimedia elements that were included

Questionnaire implementation

The survey provider Mantap Global was responsible for programming the survey. The project team reviewed the initial programming and provided feedback and/or answered open questions. Major decisions related to the addition of text (“mandatory question”) to the three quota questions on age, sex and education, and to the non-mandatory nature of all other questions.

The design of the open-ended questions was unfortunately restricted by the survey software used by the panel provider. This affected the size of the text box and the position of the category-selection probe (satisfaction with democracy probe). The size of the open-ended questions was as shown in the screenshots below – the size of the box itself could not be changed, even though we would have preferred a larger box, acknowledging that the text box size indicates to respondents the kind of answer (length) that is expected.

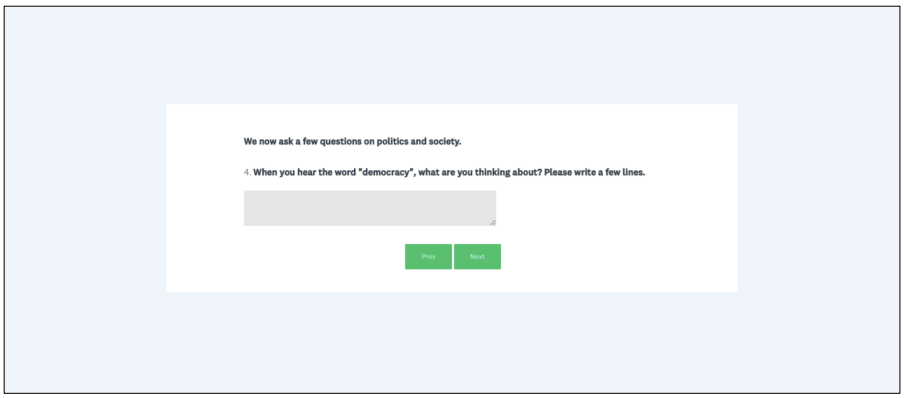

Figure 2: Open-ended question on understanding of democracy

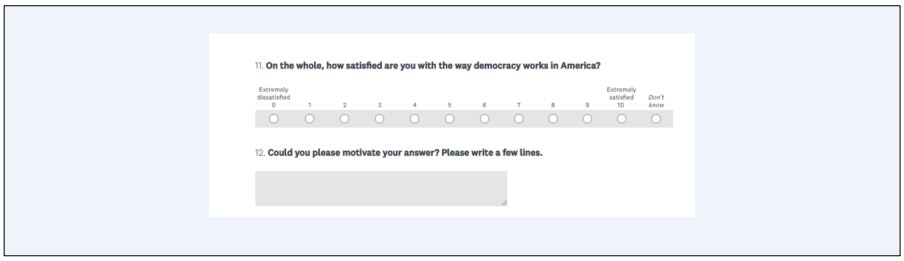

Figure 3: Category-selection probe related to satisfaction with democracy

For the category-selection probe related to the satisfaction with democracy item, we would have preferred a probe that is only shown once the closed item is answered (e.g. on a separate screen), with the probe design repeating the previous item and the chosen answer category. The panel provider could not program this in their software. Thus, we eventually had a one-screen design as shown in Figure 3.

This part focuses on all the aspects with regard to the technical implementation of the online survey
project.
_ Was the survey implemented browser-based or app-based?

_ Which survey so_ware solution was used? In case of commercial survey so_ware: Which so_ware
provider was used? In case of an in-house developed survey so_ware: Which programming language
was used?
_ Who was responsible for the survey programming and where was the questionnaire hosted?
_ Were any measures implemented to account for multiple devices and various display sizes? Was
a responsive layout implemented that adapts to the screen size? If so, what was the rationale for
di_erent layouts (e.g., type of device, screen size)?
_ Were there any technical requirements for respondents' devices (e.g., specific mobile phone type)
in order to participate?
_ Were any paradata about the survey data collection process gathered? If yes, by server-side or
client-side scripting?
_ Were any additional scripts for special purposes included (e.g., JavaScript)?
_ Was it allowed to skip answers or were forced answers implemented? Were any so_-prompts implemented
that inform about incomplete answering and that are ignorable?
_ Was a back-button included?
_ Was a progress indicator included?
_ Were any plausibility and consistency checks implemented?
_ Was there an option to suspend the survey and resume later?
_ Were any additional configurations applied that are worth mentioning (e.g., time out, automated
forwarding)?
_ Was any randomization implemented? This applies to the randomized order of answer categories,
items, or question blocks

Questionnaire documentation (metadata)

- Source questionnaire, incl. source of items: *Source questionnaire.xlsx*
- Export of all survey questionnaires are available.
- The two education classifications for the quota variable education are described in a separate document: *Education variable.xlsx*

Codebook?

All details that are needed to get a sound impression of the exact content and the look-and-feel of the
questionnaire should be included in the documentation. Potential aspects are:
_ Screenshots of the survey (if applicable, di_erent versions for di_erent devices)

_ A codebook export from the online survey so_ware which provides information about question
texts, answer categories, and filter information, as well as variable names, type, and labels of the
variables. If applicable, name respective survey from which questions were taken.
_ Export of the survey project, that enables others to import the questions in the respective online
survey so_ware.
_ If the survey includes interactive features, the provision of a short video with screen capture is an
additional option (see Heycke and Spitzer, 2019).

Pre-test

The multilingual versions underwent a technical check by the translators, as outlined in the section on Questionnaire characteristics. The feedback was integrated into the survey questionnaires by the panel provider. Since all other items (with the exceptions of the probes) were already part of major surveys, the items themselves were not further tested.

_ Was a pretest conducted beforehand?
_ Which pretesting techniquewas applied (e.g., usability testing, field pilot study, expert review, cognitive
interviewing, eyetracking)?
_ What was the aim of the pretest (e.g., testing all field procedures including questionnaire instrument,
filtering and branching, advance materials, respondent selection procedures, invitation of
participants; questionnaire development)?
_ At a minimum, information about field time and sample size of the pretest should be provided.
_ In addition, results of the pretest and possible consequences and/or changes for the main study
can be reported.

Experimental design

The survey did not include an experiment.
_ If applicable, what kind of experimental design was used (e.g., between-subjects vs. withinsubjects
design, quasi-experimental design, pretest vs. posttest design, counterbalancing, matched pairs design)?
_ How were participants assigned to the experimental groups?

3. Data collection

Contact strategy

Usual procedure of panel provider
_ How were potential respondents invited to the survey?
Depending on the project type (i.e., personalized vs. anonymous survey), one can either send the
survey link (w/o login data) by e-mail, text message, or postal mail to specific persons vs. publish

the survey link orQRcode, for instance, on awebsite, in a newsletter, via social media or in a pop-up
window.

Reminder strategy

Usual procedure of panel provider
_ What were the number of contacts and their scheduling?
_ Has one or more reminders been sent to encourage participation in the survey?
_ To whom (e.g., inactive respondents, dropouts, everybody) were reminders sent?
_ At what time (e.g., 10 days a_er initial invitation) were reminders sent?

Documentation of invitation and reminder letters

Usual procedure of panel provider
_ What was the wording of the invitation/ reminder letter? If possible, make documents available to
the public.

Incentives

Were any benefits o_ered to potential respondents to encourage participation in the survey? Incentives
can take di_erent forms:
_ Prepaid vs. post-paid incentives (i.e., provided before vs. a_er respondents completed the survey)
_ Monetary vs. non-monetary incentives (i.e., cash, gi_ cards, and coupons or "points" that can be
redeemed for cash or gi_ cards vs. small gi_s such as a pen, USB stick or notebook or things like a
brochure or a donation for charity)
_ Sweepstakes and free prize draws vs. individual incentives

Additional data sources

No further data was collected
_ Were additional data collected (e.g., social media data, sensor data, pictures)?
_ If yes, describe the implementation and provide sample material.

Informed consent, privacy information

Usual procedure of panel provider
_ How was consent to the collection and use of survey data obtained?
_ How were respondents informed about the use of the data and the essential elements of the research,
including the risks and benefits of participation? If possible, make documents available to
the public.
_ If any other types of data were gathered (e.g., paradata, passive data) or planned to be linked (administrative
data, geodata): how was consent obtained?

Fieldwork monitoring, fieldwork interventions

Usual procedure of panel provider
_ Was there a continuous evaluation of key performance indicators (KPIs) such as response rates,
breako_ rates, risk of nonresponse bias, contact attempts, etc. during fieldwork?

_ Have specific interventions been performed during fieldwork, for instance, additional reminders,
contact mode switching, survey mode switching?

Documentation of potential problems

The panel provider noticed during fieldwork that the male/female answer categories for the sex question were switched in the Estonian version. This let to sending out invitations to the wrong respondent group, when sending out further batches of invitations. Upon noticing this issue, the quota cells could be correctly interpreted and thus the right respondent group invited. The issue had been brought to the provider's attention during the technical pre-test (the translator had noticed this mistake), but it seems that the issue was not implemented following the pre-test feedback.

_ Were there any unforeseen problems during the data collection phase? If applicable, what problems
occurred (e.g., errors in filtering, incorrect quota definition)?
_ Howwere problems handled (e.g., correction of the filter and exclusion of the first k respondents)?

4. Data processing

For the sake of completeness, we additionally included aspects of data processing, which goes beyond the data collection and might not be necessary to replicate the data collection process.

Post-collection edits/ quality checks

_ What checks were carried out? Document how the controls were performed and how errors or
inconsistencies were handled (i.e., accepting, flagging, recoding variables, excluding cases).
It can be distinguished between:
• formal-logical errors (e.g., duplicates, filter errors, wild codes, multiple responses where only
one answer is possible) vs.
• inconsistent or implausible information (i.e., contradictions in the responses) vs.
• other data quality issues (e.g., breako_s, don't knows, speeding, straightlining)

Coding of open-ended responses, if applicable

The answers to the open-ended questions were machine-translated using the software MateCat. By choosing to ave the text data in private translation memories, the answers and their machine-translated answers were not saved in the universally accessible MyMemory translation memory.

> *MateCat already provides matches from a **public translation memory** with over 12 billion words and free **machine translation**. You can also create **private translation memories** for better reuse of your translations and for confidential projects (click on* Settings *to add your personal TM).*
>
> https://site.matecat.com/support/introducing-matecat/getting-started-matecat/

Also, even though MateCat uses Google translate, the texts entered are not sent to Google:

When using MateCat, are my translations used to train the default MyMemory MT engine or other MT engines used by MateCat?
No, we do not retrain the MT engine with your translations.
By default, the MT engine used by MateCat is Google Translate. No target segments are sent to Google Translate and, since we use the Google Translate API, Google does not have the right to keep the source segments either.

(https://site.matecat.com/faq/machine-translation/)

_ How were the answers to open-ended questions coded?
Provide information on coding scheme, coder training, verification of coding, error handling.

Generated/ derived variables
_ What was the purpose of generated variables and which generated variables were provided?

Post-hoc output harmonization
_ What type of harmonization strategy has been employed with the aim of achieving, or at least improving,
the comparability of di_erent surveys, waves, and measures collected (e.g., input vs. output
harmonization, ex-ante vs. ex-post harmonization)?
Explain the criteria for measuring the quality of the harmonization process.

Imputation
_ Were missing values in the data set compensated using algorithms, mean/median values, most
frequent or zero/constant values?

Weighting
_ What weighting procedures were used?
Whenweighting survey samples, themathematical and statisticalmethods used shall be appropriately
described. Provide technical details on, for instance, design, calibration, and panelweighting,
and models used to generate theweights (sources, targets/ quotas, and procedures). The variables
used for the weighting matrices as well as the e_ectiveness of the weighted sample shall be documented.

Anonymization
_ Have variables on personal data been anonymized in the data set?
For data protection reasons, it may be necessary that some variables are only publicly accessible
at a higher aggregate level such as regional information or family relationships in the household.

Analysis so_ware

_ Which so_ware has been used for data cleaning, preparation, and analysis, e.g., R, Stata, SAS, or
SPSS (including version information)?
Detailing with some specificity what so_ware was used to arrive at reported results and where it
may be obtained.

Script/syntax used for cleaning, preparation, and analysis

_ In the case of author-originated code (e.g., in R, Stata, SAS, or SPSS), is it provided in an appendix,
online supplement, or archive? If possible, make documents available to the public.

Bibliography

Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. J. R. van de Vijver & P. Ph. Mohler (Eds.), Cross-cultural survey methods (pp. 35-56). Wiley.

Schaurer, I., Kunz, T., & Heycke, T. (2020). Documentation of online surveys. Mannheim, GESIS - Leibniz Institute for the Social Sciences (GESIS- Survey Guidelines).