



# Draft report from expert appraisal of PRAIA questionnaire\*

Prepared for a national survey on participation in political and public affairs  
\*In print. Must not be cited or referred without authors permission.

TALL

SOM FORTELLER

NOTATER / DOCUMENTS

[20XX/#]

Frode Berglund and Kari-Anne Lund

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

© Statistics Norway

Published: [Dato]

ISBN 978-82-587-[####-#] (electronic)

ISSN 2535-7271 (electronic)

<b>Symbols in tables</b>	<b>Symbol</b>
<b>Category not applicable</b> Figures do not exist at this time, because the category was not in use when the figures were collected.	.
<b>Not available</b> Figures have not been entered into our databases or are too unreliable to be published.	..
<b>Confidential</b> Figures are not published to avoid identifying persons or companies.	:
<b>Decimal punctuation mark</b>	.

## Preface

Statistics Norway was asked to evaluate a questionnaire developed by the Praia Group (Task Team on Participation in Political and Public Affairs) in autumn 2022. The main objective of the questionnaire is to facilitate the production by NSOs of official statistics on Participation in Political and Public Affairs in different countries. These data are to be comparable across time, cultures, languages and development contexts. The questionnaire is intended to be used either as a stand-alone survey or as a module larger surveys run by NSOs.

Statistics Norway, [Godkjenningsdato]

Xeni Kristine Dimakos

## **Abstract**

Researchers' need for details is always guiding the content of the questionnaire. However, such needs must adapt to the respondents' ability and motivation to report. This applies even when the professional and political intentions are the very best. The consideration for the respondents is important for one main reason – burdensome questions (e.g. difficult, sensitive or numerous) reduce the chances to get accurate responses and thus produces data with lower quality.

- o Aim of the testing
- o Test design
- o Ethical considerations
- o Sampling and recruitment
- o Data collection
- o Analyses
- o Findings
- o Conclusion, implications and discussion
- o Strengths and limitations

# Contents

<b>Preface .....</b>	<b>3</b>
<b>Abstract.....</b>	<b>4</b>
<b>1. Background.....</b>	<b>6</b>
1.1. Methodological features of surveys on political participation.....	6
<b>2. Test design .....</b>	<b>8</b>
2.1. Test-design in Norway .....	8
2.2. A proposal for an international test-design .....	9
<b>3. The expert appraisal .....</b>	<b>11</b>
3.1. Content of the questionnaire .....	11
3.2. Regarding sources of information .....	11
3.3. Methods .....	11
3.4. Avoiding errors before the data collection .....	12
3.5. The concepts of respondent perspective and response burden .....	13
<b>4. Analyses .....</b>	<b>14</b>
4.1. General observations .....	14
4.2. Specific observations .....	16
<b>5. Suggestions for further work.....</b>	<b>26</b>
5.1. Possible topics and questions to investigate through user testing .....	27
<b>References .....</b>	<b>28</b>

# 1. Background

Statistics Norway was asked to evaluate a questionnaire developed by the Praia Group (Task Team on Participation in Political and Public Affairs) in autumn 2022. The main objective of the questionnaire is to facilitate the production by NSOs of official statistics on Participation in Political and Public Affairs in different countries. These data are to be comparable across time, cultures, languages and development contexts. The questionnaire is intended to be used either as a stand-alone survey or as a module larger surveys run by NSOs.

The questionnaire is developed on basis of an extensive mapping of relevant global, regional and national survey questionnaires; meaning that most questions are already applied in surveys.

The sample for the survey should be nationally representative (18+)1, no proxy2allowed due to the personal characteristics of the survey questions. The questionnaire is envisioned to be used in a post-election survey. It is recommended that NSOs schedule the survey/module within 12 months of the elections of interest for the survey.

One important aim of this harmonized survey/module is to allow comparability of statistics across times, cultures, languages and development contexts. As such, NSOs should strive to implement the module in its entirety, and to minimize the number of changes made to the questionnaire. Nevertheless, NSOs should consider the applicability of the questions within the socio-cultural and political situations at the national level and may adapt specific questions as needed. NSOs may add additional response options in question items that need more contextualization, provided that they be placed at the end of the list to minimize sequence order effects.

The survey administration mode can be expected to vary across countries. Hence, administration of the tests will also vary to cover both self-administration, interview-administration, and take a mixed-mode approach into account. Problems that may occur depending on administration mode is considered, as well as problems that may occur depending on implementation as a separate survey or a survey module.

In conclusion, there has been put a lot of work and reflections in developing the present PRAIA questionnaire. However, even if the questions already have been applied to produce national statistics, it is useful to test and evaluate both the questions and the questionnaire before introducing it worldwide. Our experience is that even if a question has proven to produce empirical numbers, it does not automatically imply that respondents interpreted the questions in accordance with the intentions of those who ordered the question or those who developed it.

As the PRAIA questionnaire is about to work across countries, cultures, languages, contexts, modes and more, it is useful to have independent evaluation and testing of the questionnaire. We must assure that the questionnaires are producing data that are comparable across many features, and we have out-lined a test design for doing so it. The design is presented below.

## 1.1. Methodological features of surveys on political participation

Estimates on electoral turnout from election surveys often shows a considerably higher share of voters compared to the official turnout. A well-known phenomenon from survey methodology is *'social desirability bias'*: people will try to represent themselves in a way that reflects well on them. But this is not the only source of error that should be taken into consideration. Estimates are not equal to the true values because of variability (due to random effects) and bias (systematic effects). A survey will be exposed to both sampling errors and non-sampling errors. Sampling errors occur because only a subset of the population is selected. Non-sampling errors, like nonresponse and

measurements errors, apply to all statistical processes. Nonresponse error is the difference between the statistics computed from the collected data and those that would be computed if there were no missing values. Measurement errors occur during data collection and cause the recorded values of variables to be different from the true ones.

## 2. Test design

### 2.1. Test-design in Norway

Pre-testing is one of the most critical steps in designing questionnaires. Cognitive interview pre-testing is used to determine how well the questions work for target group respondents. Do the respondents understand the questions well enough to provide accurate answers based on the questionnaire designer's intent?

Based on the distributed materials which have been extensively discussed within our team, we suggest the following procedure to test and quality assure the Praia questionnaire. A pre-condition for all suggested activities, is that a cleaned questionnaire draft is available. The updated questionnaire version is translated into Norwegian. We would strongly encourage that the questionnaire is also tested in the language that it was first developed. This would ensure analytical possibilities to explore the questionnaires cross-cultural equivalence.

The test process is iterative, which means that problems identified in each round are used to modify questions for the next round of tests and so forth. The Praia working group should consider if some (or all) of the tests should be performed using a programmed web test questionnaire to imitate a real-life situation for some of the countries that plan to implement the survey. This would require additional IT resources.

**Table 1. Overview of Test Design**

Steps	Timing	Activity	No of tests
Step 1	Feb23	Expert Appraisal	N=0
Step 2	May23	Cognitive tests	N=8-10
Step 3	Sept/Oct23	Protocol tests	N=40
Step 4	Oct/Nov23	Cognitive tests	N=8-10
Total	-	-	N=56-60

#### Step 1: Expert appraisal of questionnaire

A first step should involve an expert appraisal of the full questionnaire. Such an appraisal is done by two independent questionnaire methodology experts. A joint report is produced based on an analysis meeting where expert share their viewpoints and considerations. Experience shows that this procedure will identify and contribute to solving logical errors, inconsistencies etc. The appraisal will result in a revised questionnaire. A main advantage of starting the process with expert appraisals is that it may contribute to sort out many problems in the questionnaire with relatively little resources, as no respondents are involved at this early stage.

Timing: Report for Praia members review and queries to be distributed by the end of February (A cleaned and translated questionnaire is expected to be available from 6th of February)

Responsible: SSB

#### Step 2: Explorative testing, «classic» cognitive interviews, semi-structured.

A second step would be to conduct 8-10 classic cognitive interviews, 4-5 interviews on both version 1 and version 23. The survey interview will cover all questions in the full questionnaire, but only selected items such as the core questions or other questions where problems are expected, will be prepared for cognitive probing. Such interviews would be explorative, qualitative and include think-aloud sequences and cognitive probes. We aim to recruit test respondents with varied characteristics for these tests, considering age, gender, engagement in politics, which part of the country they live etc. However, their responses to the survey questions must be related to the previous national election in Norway, held in September 2021. After this series of interviews, the questionnaire may be adjusted again. Explorative, qualitative interviews like this is more resource consuming than standardized cognitive tests. This is related to rather complex analyses of qualitative data. However, it's necessary to perform a smaller series of interviews



with these characteristics to be able to identify unexpected problems in the questionnaire. These problems would not be recognized as such by respondents, and impossible to prepare probes for in advance, as we would not know what/where the problems may occur. Herein also lies the main advantage of this type of testing.

Timing: Report for Praia members review and queries to be distributed by the 1st of June

Responsible: SSB

Step 3: «Protocol» cognitive testing including behaviour coding, structured.

A third step would be testing the questionnaire in an extensive series of structured cognitive interviews, led by a strict, fully agreed-upon protocol prepared in advance. We plan to conduct 20 tests of each version, in total 40 tests. It is generally assumed that respondents benefit from a concurrent probing mode as it is easier for them to remember their thought processes from a question asked just a minute ago, compared to probes asked retrospectively at the end of the questionnaire. Which probing mode to use, should be discussed in the Praia working group as the chosen mode would affect the (lack of) possibility to test order effects in a concurrent probing mode. Test respondents answers to the survey questions in step 3 and 4, should be related to the local election in Norway, arranged in September 2023. Interviews may be conducted by especially trained research interviewers, who gets additional training in behaviour coding (codes would assumedly resemble codes used in the «Non-discrimination and Equality» pretest protocol). The analyses will be conducted by two questionnaire methodology experts in SSBs Division for methods. Such an extensive series of standardized interviews would allow us to get comparable, quantitative and reliable information on which questions would be more exposed to unwanted response behaviour than others. Behaviour coding may be summarized with relatively little effort into tables that illustrate problematic behaviour, problematic questions – hence where to put efforts to improve the questionnaire. The survey data should also be included in the analyses, as this would allow for a more complete picture of the data quality.

Timing: Report for Praia members review and queries to be distributed by the 1st of November

Responsible: SSB

Step 4: Follow-up testing, «classic» cognitive interviews, semi-structured. N=8-10

The fourth and last step would be to conduct 8-10 classic cognitive interviews to investigate further and follow-up on possible problems detected in the protocol interviews in Step 3. We plan to conduct 4-5 interviews on each of the two versions. This would be an important final step to make sure that any previous adjustments to the questionnaire are functioning according to intentions and to make sure we have a comprehensive understanding of how the questions perform. It might be possible to use a questionnaire extract for these last tests. That will however, be (re-)considered as we precede in the test series.

Timing: Report for Praia members review and queries to be distributed as part of the report from step 3, by the 1st of December

Responsible: SSB

A report documenting the test series will be published.

The present and first report contains the findings from the expert appraisal

## 2.2. A proposal for an international test-design

Norway might represent a good starting point for testing, but the results will not necessarily be representative to other countries. We therefore suggest that NSOs in other countries are invited to do testing in their respective countries. They do not have to conduct all the four steps that are

outline above for Norway, but maybe one or two steps based on the expertise and resources they have available. SSB might assist with training and implementation abroad.

In short, we see two approaches.

First, it is the standardized protocol-based verbal probing. Here, the interviewer “probes” the respondent with already scripted questions about their thought processes while answering the survey questions (Willson & Miller, 2014). Interviewer is active as s/he “probes” the respondent for the basis of their responses. The approach allows a rather high number of interviews (n = 25 -40). It does not require much training of interviewers, but one might need many interviewers.

Second option is the semi-structured qualitative interview. Interviewers instruct the respondents to “think aloud” as they answer the survey questions – “Tell me what you are thinking.” Respondents “verbalize their thought processes as they went about answering a survey question” (Willson & Miller, 2014). The number of interviews is rather small (5 -10), but it requires training of interviewers. They must be able to manage qualitative interviews, and to formulate spontaneous probes.

Together with UNDP, Statistics Norway can provide support on the test-design and offer:

- Detailed protocol for option 1 and general guidelines for option 2. All material will be available in English, French, Spanish and Portuguese.
- Virtual training sessions for interested NSOs:
- Training on protocol-based verbal probing (Option 1)
- Training on think-aloud interviews (Option 2)
- Peer support from other participating NSOs ('French Group', 'Spanish Group', 'Portuguese Group')
- Technical support

## 3. The expert appraisal

### 3.1. Content of the questionnaire

The voting part do only take into consideration voting at the polls on election day. In many countries it is possible to vote in advance via regular mail or other methods. This must also be possible to report in a questionnaire about voting. Many questions in the current version are not relevant to people who voted in advance, or they must be rephrased to fit the way they voted.

### 3.2. Regarding sources of information

The current questionnaire does not include any question about which sources the respondents use to obtain information – only which one they trust. It will in any way be more important to know which sources they are using, in order to know where they get information about politics. As we suggest in the section on trust in sources, we can't just ask about radio, television or newspaper. There sources are so diverse today that it is we need at least some more information to make sense out the information. In Norway, it would be informative to separate public from commercial television/radio. It is a crude cut, of course, but it is a small step forward to get more accurate information.

In general, it will be necessary to shorten the text to make it fit better into small screens.

The applied scales do lack mid-points. For many of the present scales it will be natural to include a mid -point, and not force respondents to choose side. In addition, one should try to make the distance between point to be more equal than they are here. Some points are close, while the distance to the next is too far.

As the questionnaire is developed to serve all countries, some question are not at all suitable in Norway, while some question should be developed specifically to Norway (and to other countries as well). Questions about violence on election, e.g. is not so relevant here, and will produce little variance. A question like, "is it acceptable that candidate for prime minister can be a woman," also seems a bit strange as the prime minister have been a woman in 18 years since 1981. That might not indicate that all citizens accept female prime ministers, but the view of not accepting is probably very seldom.

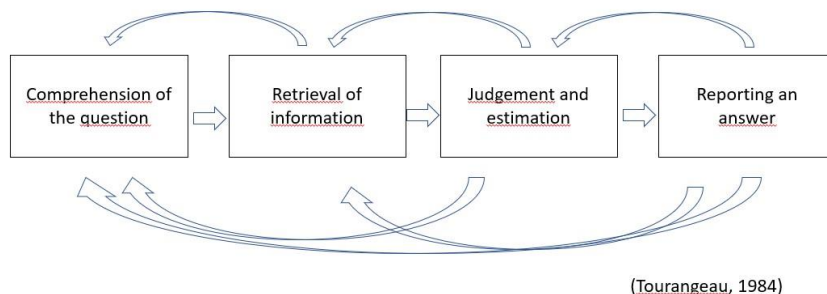
Having said that, we will emphasize that this evaluation at done by the desk. However, the authors are familiar with questionnaires and problems associated with them. So, we know that we will find out more after we have conducted interviews with real respondents.

We do not see any question about which party the respondent voted for – assuming that it would not be relevant to analyse respondents' perceptions of freedom of speech in relation to which party they may sympathize with?

### 3.3. Methods

The current appraisal aims to identify potential problems respondents may experience in the response process. This includes understanding the intent of the question or meaning of terms, remembering relevant information and lack of motivation to invest required efforts to respond accurately. Further, identifying what are sensitive questions or questions people hesitate to respond to, and problems choosing an appropriate response category. The model below model illustrates the process of how respondents produce an answer. It is important to be aware of these steps when evaluating questions; there are a lot of things going on in the minds of respondents from hearing a question to reporting an answer to it.

## A response process model



This evaluation is conducted in accordance with standard procedures for questionnaire appraisals. The Praia questionnaire is reviewed to identify cognitive problems in relation to the interpretation of the questions, information retrieval, the task that the respondent is intended to solve, and lastly to give a response in the format that is offered. The approach results in a review of more general character, and a question-by-question in-depth analysis. To perform this task, we rely on different techniques and current best methods, such as Forsyth's "O-Questionnaire Review Coding System" and Snijkers' comprised "Expert Questionnaire Appraisal Coding System".

The main intention is to identify weaknesses in the questionnaire that can reduce the quality on the responses and increase the response burden. All recommendations in this report aim to reduce the response burden and increase data quality. Even though we may point out some alternatives, it must be emphasized that this is a general appraisal of the existing questionnaires and not a recipe or solution for what the improved version should look like.

Her må vi ha med noe om styrker og svakheter med metoden.

### 3.4. Avoiding errors before the data collection

When considering data quality in the preparation of a survey, one can divide potential issues in sampling errors and measurement errors. The latter being errors related to how you measure whatever it is you decided to measure, and this is where the questionnaire becomes a focus of attention. It is extremely useful to spend resources in developing proper questionnaires. After all there are very limited alternatives for how one can ensure good data quality after the data collection, but a lot more one can do to ensure good data quality before the data collection. When developing questionnaires, we single out four separate tasks:

- In the specification phase, we try to thoroughly specify exactly which data is needed to do our analyses. It is recommended to divide sharply between data that one need to have and data that is nice to have. A rule of thumb is to delete questions that is only nice to have. Further, it is absolutely necessary to ask what kind of data is available to the respondents.
- In the operationalization phase, we transform the theoretical terms from the specification phase, to a more pragmatic language that the respondents can relate to. One makes use of the operationalized concepts from the former phase in actual, complete questions.

- In the questionnaire design phase, all the individual questions are fitted together in an efficient questionnaire.
- User testing is best regarded as an iterative process: One tests the questionnaire, adjust the questionnaire, tests the new version, makes new adjustments and so forth. The more thorough job one did in the specification phase, the fewer rounds of testing is normally required.

### **3.5. The concepts of respondent perspective and response burden**

The present view on questionnaire design was co-developed by psychologists and survey methodologists during the 1980s. The resulting perspective is often referred to as the “cognitive perspective” in contrast to the earlier “mechanic/naive” perspective. This current view acknowledges that answering questions is a complicated cognitive process. That is in contrast to what one used to believe when one “collected” data “out there” and did not reflect on the issue on how the data in a way is shaped in the meeting between the respondent and the interviewer/data collector. This is of course a very critical discussion within qualitative research but is also an important issue in quantitative surveys. One usually treats response burden in one of two ways – “actual” response burden vs. “perceived” response burden. While the “actual” response burden refers to how time consuming the participation was measured in minutes, hours etc., the latter focuses on the respondent’s experience and perception of how demanding he/she experienced it was to participate in the survey.

A high response burden – measured in time or effort - is generally thought to decrease the quality of the data, because it is less likely that the respondents respond to the survey questions as intended in surveys with high response burden. This eventually leads to “satisficing”, which can be understood as making mental shortcuts when answering the survey questions. This is a kind of strategy that some respondents may lean on in surveys that entails high response burden (e.g., difficult questions, time consuming etc). To avoid high response burden, and thereby ensure data quality, we need to craft precise, easy questions, keep the questionnaire as short as possible, strip away every unnecessary question, adjust the questions to the objectives of the survey etc.

## 4. Analyses

### 4.1. General observations

In the following we provide a general review. These are observations made one two or more questions.

1. "Who did this?" (Ex. PR 3.1, PR 5.1)  
Questions aiming to determine who exposed the respondent to pressure etc. Would recommend a review of the categories to make sure they give precise information. For example: "A family member" can be interpreted quite differently across cultures and ethnic groups – would make more sense to analyse data from this question if there was a category for "family members IN the household" + "Family members OUTSIDE the household? Other family members" (Close family/wider family?) or "own children, parents of siblings" – family can be very broad?  
Also, category two "A friend, colleague etc." is not very clear – who is "etc."? Would it be possible to find and reuse categories reflecting a similar/scale from other surveys? The categories may not be equally relevant across countries, especially considering the last category "security forces, such as police, the army". Can this be reworked into a more general category that is applicable in every country to ease comparison?
2. Reuse wording from already established surveys (Ex. PR 1, PR 4)  
Reusing questions from other surveys would enable comparison to a much greater extent compared to what we can if we design tailored question versions. That also goes for response scales that has performed well already.
3. Many sub-questions in one battery (Ex. PR 17)  
Questions with many sub-questions – so-called "batteries" - would need to be split up into smaller sequences or separate questions to perform well also on smaller screens such as smart phones. Large question batteries would not be possible to fit into one screen and we would advise against requiring respondents to scroll on the page to see all necessary information. Large question batteries would also risk respondents losing the connection to information provided in the key question/intro text, or not respond to each question as an individual question, but rather see the questions as part of/in relation to other questions in the same battery. This could in turn create order effects that can be more prominent in batteries compared to in individual questions.
4. Double negations (Ex. PR 5)  
On some of the sub-questions in the battery can be prone to response errors, as respondents may be confused by more than one "not". For example, "no" in question 5 could mean "no, this was not part of the reason why I did not vote", or "no, I did not have enough information on how to vote" (hence, yes, this was part of the reason why I did not vote). Such questions would contribute to a higher response burden as it would require more effort from the respondents to interpret the question as intended)
5. Mobile first approach? (Ex. PR 16)  
Shorten amount of text to make sure the questionnaire can also be implemented in self-administered web surveys. The entire questionnaire should be reviewed specifically looking for possibilities to reduce text. If a question performs well on a small screen, it's likely to

perform well also on larger screens. Shortening of text would not only benefit web mode but could also reduce interview length (=costs) in interview administered surveys.

6. Ask one question at a time (Ex. PR 6i, PR 11)

If we ask respondents for more than one thing in a question, it would be difficult to interpret data from the question as we would not know if they responded based on both issues or just one of them, and which one. Two questions in one, could also create confusion and insecurity for respondents who would be unsure how to respond if they can confirm one issue, but not both. We find several examples of this in the questionnaire.

7. Unimode approach? (Ex. PR 6)

Rewrite text that seem to be developed for interview administered mode. Try to develop texts that could be kept as is regardless of mode. For example, "now, I will read a list of..." would be developed for interviewer administered mode and would have to be adapted when administered on web

8. Scales

Lack of mid-category (Ex. PR 23.1, except PR 24 and 25)

Without offering respondents a neutral mid-category, we force respondents to have an opinion they may not have. In some issues, respondents may in fact mean "neither nor", and we should mostly consider giving them the opportunity to report their true answer.

Reconsider labels to ensure similar distance between points in scale. One should strive to spread the response points as evenly as possible over the scale and avoid two points that are very similar in the poles, or elsewhere. Scales may be tempting to use, but often it might be difficult to interpret the meaning. Take this claim (PR 23.2.e): "Men are better leaders than women." You can state if agree or not, if you do not or that you don't want to say. What if you mean that sex doesn't matter? Do "completely disagree" cover this position, or does it only state that you do not believe that men are better than women? What if you believe that women are better than men? It will be interesting to find out how respondents think about comparisons where not both objects are taken into consideration.

9. Mismatch between question and response options (Ex. PR 21)

The question needs to communicate to respondents what type of response is expected. If this is not the case, interviewers would have to improvise additional instruction to respondents to make them understand how they may respond. In modes that present response options to respondents aurally, does not allow respondents to see which options they can choose from. It would have to be clear from the question what type of response is required.

10. Proxy questions (Ex. PR 18)

It would not be recommended to ask for information from someone other than the respondent, for example what the respondent may think that someone else has experienced, or how other persons feel about certain issues. This is because respondents are not likely to have reliable information about the thought processes or feelings of someone other than themselves. Hence, data from such questions would be of questionable quality.

11. Difficult interpretation of concepts (Ex. PR 10.2, 19, 24 and 25)

10.2: "Partisan" – not known to general public?

19: "Political and civic issues" – what is the difference?

24: «Have something to say in what the government does" vs.

25 "Have influence on politics"

We must keep in mind that these questions should perform well also for persons who are not engaged in politics. Both a difficult interpretation task and requires a level of knowledge that we are not sure if we can expect from all respondents in a general survey?

12. Hypothetical questions (Ex. PR 23.1)

Such questions may produce data of questionable quality due to the lack of link to reality. It might be a (huge) gap between good intentions and actual behaviour. Hence, it would be recommended to avoid hypothetical questions or combine them with questions on actual behaviour.

13. Cross cultural comparison (Ex. PR 23.2)

In societies where gender equality is common, it might not seem worthwhile to measure stereotypic attitudes towards women only, without asking similar questions about men. If we ask for such attitudes for both women and men, the results/comparison will be more interesting than if we only measure such attitudes towards women. We also risk "producing" attitudes that do not reflect a "true" or balanced perspective of the society.

14. Questionnaire relevance/coverage

There are no questions to cover voting outside of the election day – would that be something to cover also? E.g., those who voted by prior to the election day (by regular post or electronically) would not have experiences from the election day other than what they may have experienced in their own homes, in front of a computer, or at the postal office. There are also no questions that ask respondents about what time they decided what to vote. That is important information when analysed together with the information on experienced pressure to vote for a specific party or candidate.

There are no questions to cover if respondents voted only once, or if they both pre-voted and overruled their own vote by voting again on the election day. Not sure if that is very interesting, but it is nevertheless worth to be mentioned.

In a Norwegian context, social sanctions are frequently used as "punishment" of people expressing opinions in public – it would be highly relevant, interesting and important to capture how such "punishment" affects the public in general. The questions as pt. seem to reflect more formal and violent sanctions, while in some contexts, social sanctions ("cancel culture") is equally relevant.

15. Reference periods

We do not know how clearly respondents interpret terms like "Outside the election" and "prior to the election". Respondents with a clear political interest might know how to differentiate between before the election campaign, the election campaign period and the election day, while respondents with less interest in politics might only differentiate between the election day and the period that preceded that day.

## 4.2. Specific observations

In the following we provide a question-by-question review. These are observations specific to each question, and may be observed on one particular question only.

### A. Participation in Electoral Processes and Referendums

#### A.3. Voting



Translation issues mainly.

The “if relevant” categories are not relevant in Norway. However, it should be noted that there is a mismatch between question wording and response options and we cannot expect respondents to understand that they should provide their answer more nuanced than “yes” or “no”. This version may work in a self-administered mode but would pose challenges in an interview administered mode because the interviewers would have to improvise follow-up questions to get the response as detailed as needed. Consequently, it can be that a better solution would be to split the question into two separate questions where the first one is only yes/no, and then a follow-up for more detailed information can be asked to those who respond “no” to the first question.

PR 5. This is a sensitive question and the wording should be developed carefully. Hence, the “May WE ask what is the reason...” may not be optimal? Another way to ask, could be to start with a desensitizing statement like “There are many (equally reasonable) different reasons why persons do not vote. What is the reason why you....”

Reconsider the order of the sub-questions to make sure the most relevant alternatives come on top?

Too many sub-questions in one battery – will not work on smaller screens such as smart phones.

Too many sub-questions in one battery can also increase chances of response order effects such as for example anchoring or satisficing.

Double negations on some of the sub-questions in the battery can be prone to response errors, for example interpretation of PR5i. “You did not think that your vote would make a difference” – A “no” can mean “no, I did not think that my vote would make a difference”, or “No, this was one of the reasons why I did not vote”.

PR 5a. Unclear if it should be interpreted as not having ID at all, or just not available at the for voting?

PR 5d. Should be split into two separate questions to differentiate between access to the voting facilities or access to transport to get there?

PR 5f. “Take time of” may be applicable to only those who work/study, reword to a more general “Could not set aside enough time to...” or similar to make it relevant for everyone?

PR 5g. Is it possible that fear for personal security may have hindered persons from voting, NOT restricted to “vote related violence”? If yes, may we drop that part of the question?

PR 5h. Partly overlap with 5d. Also, several conditions in one question (ill, disability and access). What are the data requirements? Could it be reworded to reflect “health reasons” prevented you from voting or similar?

PR 5i. Reword from “thought” to “meant” or something. I am not sure, but the wording pt. seems somewhat “patronizing”?

PR 5k. Reword to “Someone prevented you from voting” Are we sure we need examples to illustrate what “someone” means? To make sure the questionnaire can also perform well on smaller screens, it would be good to keep the length of the texts as short as possible.

PR 5.1 Not worded as a question, but still accompanied by a question mark (?). If we design the questionnaire in such a way that the question pops up as an instant follow-up to the filter question (5k), we can avoid a lot of text and increase the web usability. For example, “Who prevented you from voting?”

The categories should be reviewed (see general observations)

Should this question also be accompanied by a follow-up question to determine how this happened, or do we not need to know if this was by physical barriers, convincing or threats (or other?)?

PR 6. Double wording “potential” and “can” means the same. In Norwegian, we would suggest the following: “.. problemer velgere kan støte på ...” Reword from “gå for å stemme» to «skal stemme» and from “gikk for å” to

“skulle”, to avoid risking persons to interpret it as walking to the voting facility. Not sure if this is only valid remarks for Norwegian translation.

PR 6a. Remove, “too”, as they did vote, hence the lines were not too long, but might have been long.

PR 6b. Problem with double negation.

PR 6c. Problem with double negation.

PR 6d. Reword so that the sub-questions perform better as alone-standing text. From “You had problems...” to “Did you have problems with...” Questions instead of statements.

PR 6g. Should the question cover “violence” only, or be reworded to capture also other forms of threats to personal security?

PR 6h. Split in access to facility and access to transport?

PR 6i. Captures several aspects in one question – we should only ask one thing at a time. Suggest revisiting the data requirements for this one in particular. Having problem getting into the polling station due to disability is something else than being in a hospital on election day.

PR7. The wording of this question may produce data to support misleading analyses – what is the difference between “safe” and “very safe” and do we know how safe the respondents usually feels on a not-voting setting? Can it be that some might report that they felt “safe”, but not “very safe” because they do not go around feeling “very safe” all the time anyway? Especially considering cross cultural comparison, this might be reconsidered. This is a question that should be tested in different countries and cultures.

#### A.4. Participation in Election-Related Activities

PR 8. Alternatively, we suggest that respondents are asked if they were engaged in the electoral campaign for a candidate or a party. Working in the electoral campaign is more demanding and enduring than the other activities mentioned in the next section. By asking this way, 9b are moved up, which is nice to those who already has said that have worked for pay. Section 9 will then consist of activities that is not as time consuming as campaigning.

PR.9 «Now some questions about volunteer, unpaid, vote related activities” Try to shorten the text and avoid asking a question that we do not want the respondents to answer. The questions come in PR 9a etc.

PR 9b. Change from “Did you work as...” to “Did you engage in...” to avoid respondents misinterpret the question as to be about paid work? (The argument is also valid if we move the question to section 8.)

PR 9c. Add “Engage in” to make sure it’s an activity we aim to capture and not just any convincing around the dinner table? Did you engage in... or did you actively....? In NO its not relevant to ask if helped to register.

PR 9e. The sign in front of the house is not relevant in a Norwegian context as practically no one does that here. Maybe consider add a sticker to the car or something instead? Also wearing a pin, the party symbol, or if extreme – wearing a t-shirt.

PR9f. The “naturalytelser” may seem a bit weird, can we provide examples of what that might be or is the implication intended? Is it e.g. supply of food?

PR 9g. Unclear what it means to voluntarily support a campaign IS it like signing up, give money to candidate, or what?

PR 9j. The specification here can be used actively in the tests to investigate how well the categories perform and to evaluate the need to add more categories.

#### A.5. Experience of Political Intimidation or Violence During Elections

If these questions should be presented to all respondents, then how would data from these questions be interpreted? Those who report “not at all” can mean two different things – not at all, because they did not plan to vote hence not relevant - or not at all, hence they did not fear these phenomena’s when voting?

PR.10.1 Several dimensions and very different level on the examples. Being asked about voting intention does not necessarily imply a threat, so they should not be mentioned as equal examples. When is this? In the campaign period or only on the election day?

“Fear” is a very strong word, can be reworded to something else? We wonder if it is “fear of threat” that is the relevant information or is it “having experienced threats” that is relevant. “Fear of” does not even imply that it has happened.

PR 10.2 Unclear if it’s a question about fear that something like this might happen to anyone, or to the respondent?

We acknowledge that it can happen some places, but to make the question more generally applicable in all countries it might be reconsidered to moderate the part about state attacking those who demonstrate as this is not what usually happens – a more common event could be that they interfere or block the demonstration, removing those who demonstrate.

It would be better if we managed to use the word “fear” also in the response categories. However, the scaling of fear seems a bit strange; how much do you fear ... It is better to apply another term than fear.

«Partisaner» is a difficult word. The English word “partisans” have been translated into “partisaner” in Norwegian, and give connotations to political activists. That suits the question, but is the intention with using the phrase partisan? In English, it is often used as a reference to person belonging to a regular, political party

PR 10.3 When?

Can be perceived as irrelevant if one has reported to not have engaged in such activities? Further on, as unsafe is in the response categories, it should be mentioned in the question. Then, the question will be balanced, saying that it is “allowed” to feel unsafe. Having said that, it will be better to ask, “How safe was it?” than “how safe did you feel?” it is not to underestimate feelings, but the feeling component makes the intention of the question less clear.

PR 11 “You or someone else” – unfortunate, too broad, that is also valid for the reference period. Some of the questions partly overlap already asked questions.

Too many questions to be presented on one small screen, should be split in two screens at least – that would be logic anyway as the first 3 sub questions are more related to each other, and so also the last 3.

The question wording “you or someone else you know”, does not correspond to the “you” in the sub-questions.

It's problematic that the questions are worded as yes/no, but that the responses are expected to be more nuanced. The question should be split into several individual questions or cut down to cover the respondent only, and ask y/n first, then follow-up for "yes" reports.

The current response categories imply that if you experienced it, we do not know if you know someone who experienced it. That seems to be somewhat strange information.

PR 11f. Is the question intended to capture personal experience towards oneself, or also threats to someone else? If also someone else, this is extremely widespread/happens a lot.

PR 11.1 The wording does not correlate to the previous question, as it might have been someone else the respondents know who experienced this? This question should be a direct follow-up of the filter question. Then we can cut down on the text length. Regarding the categories, we will suggest that member of the family is separated into household and relatives not in household.

PR 11.2 Who tried to prevent you from voting? Should be presented directly after the filter question. Separate member of the family is separated into household and relatives not in household. State soldiers attacking politicians is not very common in parts of the world, ref. PR 10.2.

PR 11.3 Who tried to prevent you from voting? Should be presented directly after the filter question. Separate member of the family is separated into household and relatives not in household.

## **B. Participation in Political and Civic Life**

### B.1. Political parties

PR 12. Skip the "hvilket som helst" in Norwegian version – does not seem reasonable, as respondents are probably not member of "any" (random) political party. Is it necessary to specify that it needs to be in a specific country, for example Norway?

The question is worded as a yes/no-questions while the response options are more nuanced. This is not optimal because there is no guide to respondents that let them know that we do want more information than a yes or no answer. In a interview administered survey this would require interviewers to improvise follow-up questions to get as much details as required to respond properly.

There is a mismatch between question and response options. If we need to know if respondents have been members earlier, this should be asked about in a separate follow-up question.

PR 13. "Outside of elections" is an unclear reference period – it's very general/vague and it could cause respondents experiencing problems when interpreting the question. Does it mean outside the election period or the election day?

The term "civic" activities are unclear/unknown to many people and would cause interpretation problems for respondents. What kind of activities is this? Would it make sense to ask the question without the condition civic part? It's also unclear what is meant by being "involved in activities" – does it imply ongoing activities, such as being

member of a working group who meets regularly or is what we want to know if respondents have participated in any activities organized by a political party? It should be considered, if it would ease the response burden for many persons if we asked a general question first, and then specify in a follow-up for those who respond "yes", to get a clearer understanding of what kind of activities this was. Also skip the any?

PR 14. Reconsider the response scale to "Veldig, Ganske, Litt, Ikke?" as the Norwegian translation does not provide a balanced scale. Valid for both English and Norwegian version?

A more prominent question is what kind of information we try to capture with this question. We already know that the respondents who get this question is not member of a political party. That's the reason why we suggest considering asking (at least those who report to be "very interested" in becoming a member, if they do in fact consider signing up as a member. The rationale is that we by that follow-up question try to get more information about how realistic the "very interested" is about actually becoming a member – is it just a vague idea or a realistic plan? Or did they try to become a member, but did for some reason not succeed?

PR 15. The language in "represents my views reasonably well" is quite formal, is it possible to ease up the language a bit to make it less formal? Change from "at least one" to "one or more".

We suggest that question b15 and b14 should be swapped. Then we will know the rationale behind the (lack of) interest to become a member. And, instead of asking, "how interested are you ...," the question could be, "do you consider ..." or "have you ever thought of becoming a member" of one these parties.

## B.2. Membership in Political and Civic Organizations

PR 16. Shorten the text in the introduction, not least to make it work on smaller screens like smartphones.

Reword to mode independent (Now I will read (...) – Now, some questions about (...))

Phrase it as a to the point question - "Are you a member of..."

Split this question about memberships yes/no and the follow-up question 16.1 about participation in activities. First membership, then a direct follow-up question on participation in activities. This inter-leafed design takes the respondents cognition into account. Once they think about a certain type of organization they answer on membership and activities (in case of a yes-response to the first question), before thinking about the next type of organization. That way they do not have to rethink the same organizations (twice). A downside of interleaved design could be that respondents quickly learn that they get follow-up questions for every "yes" they produce, hence start reporting false "no" to avoid more questions. The extent of this deliberate underreporting is however likely to be low, especially considering the target group who are politically active.

Consider the order of the type of organizations to make sure the most relevant once are presented first?

PR 16.1. Is the condition about which type of activities the respondents may have been participating in necessary? Can we ask generally if they participated in any activity first, then ask about type of activity for those who report to have participated? Could it be hard to judge if an activity organized by an organization is "political or civic"? The response options should be changed to yes/no (instead of member/not-member), as this corresponds to the wording of the question.

In a web mode we might want to leave the introduction text on the top of the page, then question about membership in orgtype 1 y/n, then participation in activities in orgtype 1 y/n. Then replace orgtype1 with orgtype 2 and repeat the pattern.

PR 16a. What does this mean? Unclear.

16.1c and d. Repeated text, if possible cut down to avoid unnecessary repetitions and also make sure the format is

suited for small screens.

It could be hard to both interpret and retrieve relevant information to gain a proper understanding of which organizations should be captured in each of these follow-up questions. Hence, examples of the most dominant organizations may be added to ease the response burden.

16e. Unclear what “kind of activities” should count – attending a member meeting at work/during work hours for example, is that a kind of activity that should qualify to say yes?

16.1f This “other”-questions should be worded as a complete question in full sentences and a yes-response should be followed-up by a “please, specify” field.

### B.3. Participation in Other Political and Civic Activities

PR 17. Shorten the intro text and reword to fit any mode. Separate intro text from the question text. During the last 12 months, did you...

...contact a (...)

Also consider the order in which we present the different events, there should be a thought through system, for example moving from the least “active” event, to the most “active” event at the other end of the list?

PR17a. Hard to interpret if it needs to be a person or a service? Are there any other levels than national and local? If no, that means we do not have to specify which level and can shorten the question? Also, unclear if this needs to be done as a private individual – many persons do this and more of the activities on the list as part of their work, but should that be part of the response?

It's not clear in the questions that the actions on the list needs to be due to a wish to “improve things” etc. as mentioned in the introduction text. If that is a condition, we need to work this into the question wordings.

PR17d. Hard to understand which level one needs to meet to say “yes” – and the examples covers very different type of activities – participation in a strike can for many people be a lot less “active” and done out of a personal desire compared to participation in a demonstration. This might be a cultural difference in Norway only? (Can be investigated further in testing)

PR17h. To solve or discuss? And reword to “Met with”. Also, it the last part of the sentence about who arranged the meeting and why, necessary? If no, we may shorten the text to fit smaller screens. How formal must this meeting be to be counted?

PR17i. “Refused” is a very strong word – maybe change to “avoided”?

PR 18. Here we ask for information about other persons than the respondents, and many persons will perhaps experience that it's hard to judge on behalf of others. It depends on your viewpoint, for some groups/types of political viewpoints it may be unsafe, for others no. Reconsider. (Can be investigated further in testing)

## C. Enabling Environment of Participation

Reword intro text to be mode independent – “Now some questions about political interest”

### C.1. Interest in Political and Public Affairs

PR 19. Add “overall” to more clearly communicate to respondents that we want them to give us a general answer, not depending on the example given?

Reconsider using example here, as this could influence the respondents answers more than we want (can be explored in tests).

In a general survey, it may not be equally known to all respondents how to interpret the difference between politics and civic matters?

The question is a two-in-one asking for both political and civic matters, how about sticking to political only?

The scale may be reconsidered as it should be easy for respondents to see the difference between the categories.

## C.2. Information on Political and Public Affairs

PR 20. Many issues in one question – it’s both news AND information, ON political AND public affairs, that respondents get access to through the sources mentioned.

It would be relevant for all these sources for information that it depends on which distributor one considered – not all radio stations are reliable, but some – and the same goes for all the other sources. How will respondents relate to this? May be difficult (or does not make sense) to make an “average” opinion based on all in one answer? As a minimum, we recommend splitting between state and commercial distributors on TV and radio.

PR 20b. TV – what does that mean today? Web-TV, on a TV-set, real-time TV, TV on demand? All? Commercial or public?

PR20e. What is video hosting? Is that a phenomenon that all respondents know? And podcasts, are they published on websites? Are video hosting and podcasts two sides of the same coin so that it makes sense to consider them in one question?

There is no question about which sources the respondents use to obtain information – would that be (even more) relevant here? This would also reflect which sources they find trustworthy but in a more concrete way? Maybe before PR 20?

Questions on trust has an established “standard”, may we reuse that here?

PR 21. There is a mismatch between question wording and offered response categories. May consider rewording the question to “How participatory are you in...”, also reshape the categories to be easier to see the difference between them. Move the most important element first. Remember that respondents often read very superficially. Response options could for example be like

“Active participation, contributes with for example sharing, distribution etc.”

“Passive participation, monitors discussions, but do not publish own material, and

“No participation, neither follow or publish”

Is the response options an exhaustive list? We should also consider a category to capture those of us who are somewhere between passive and no participation at all – do not “follow”, but to say that this is total absence is not correct either. This is something that can be explored further in the tests, but pt. we assume that many persons would relate more to a category that reflects a midpoint between “passive” and “no” – a 3-point scale might not reflect the reality. This can be solved by adding a fourth category, or split the question into several y/n questions – are you active in online discussions, by publishing, sharing etc? y/n, - do you follow online discussions, without sharing, publishing etc.? y/n,

Do you follow? Yes/no if yes,

Do you contribute? Yes/no if yes, to what extend? Often, sometimes, seldom

PR 22. There is a lot of repeated text here – the same is said first in the key question and then repeated in all the sub-

questions. Especially considering a web mode (but also unnecessary long interviews F2F or on the phone, increased interview time= increased expenses) where the text should be as short and concise as possible. Decide if better to keep the conditions in the key question OR in the sub questions. The advantage of keeping it in the key question would be that we can reduce a lot of text due to the one-time-only presentation. The advantage of keeping it in the sub questions would be that the respondents might be more conscious about the condition when infiltrated in each and every question.

May reconsider the order of PR 22a-d, as a, b, and d are similar while c is the odd one out. It would make sense to present in the following order: a, b, d, c?

PR22c. Change from “violent” to “threatening” or “aggressive”? Valid for both versions?

PR22d. Considering the response scale – would it communicate an «equal» distance between each category better if we added “partly” to the two mid categories? “Full of” might be too strong?

### C.3. Adult Population with Stereotyping Attitudes and Values: Three Versions to Test

PR 23. This must be redesigned in a Norwegian context as we do not have direct voting on person level. Could for example reword to “support a candidate” or similar?

Considering the response scale – do we offer an exhaustive scale? Is “very willing” the contrast to “not willing at all”? There is no alternative that reflect the position that it does not matter at all, this should be considered.

The distance between the categories in the scale might be clearer if we use consistently the same prefix, such as “Veldig, ganske, lite, ikke”. Is this a NO translation issue or relevant for original questionnaire to?

In Norway its quite common to also measure attitudes or discrimination of persons due to their age – this should also be reflected here – high age and low age.

PR 23a. Should we also ask the same but for men? That could be a position that some hold in Norway, and probably also in other countries. We should give the respondents an opportunity to report their stereotypic attitudes and values as objectively as possible.

PR 23b. Many different minorities covered in one question. Can this be difficult for someone who might be reluctant to religious minorities, but not to ethnic minorities – hence potentially lead to underreporting because they do not want to report that they are unwilling towards all these groups?

PR 23d. Shorten to “sexual minority”, if we use terms in the other sub-questions too?

PR 23.1 Does not seem like a good option as it is both hypothetical and proxy – responding for someone else. It would give us a lot of less reliable data. We advise against using this version.

PR 23.2 Of the three versions, our recommendation would be to further develop and use PR 23.2 V3. Statements should be developed to cover all minority groups as reflected in V1/V2 and reworded to “equal opportunities” to avoid references to probability calculations. Skip 23.2a, e, f and g as they would not be relevant. If kept, statements should be added to reflect the reverse attitudes also. Unless, we risk being accused of producing data that does not reflect a true situation.

PR 23.2a. Difficult to interpret – the point is the equal opportunities, not that it needs to be the exact same number? Would not that be more controversial to support, than liberal? Very sensitive to the words, terms used – large room for different interpretations – it depends on how this is achieved – through force, quotations or ideally? Can be explored further in tests, if relevant.

PR 23.2d-g. We should consider adding questions with the opposite value as these are also opinions that should be reflected in the survey, despite the probability that there is fewer to support these/such views.



#### C.4. Self-Reported Political Efficacy

PR 24. Reuse established wording of question? The word “allows” makes it sounds like a legal question, and that makes it less relevant/less a question of the respondents’ opinion, but a fact? In a democracy, people are allowed to have something to say?

PR 25. Make the difference between this question and the previous, clearer or skip PR 24 and keep 25 only, as 24 could be dependent on which party one sympathizes with. What is the difference between them? Especially late in a survey when respondents start to get tired, they could have difficulties getting the difference without taking a close look (which they would be reluctant to do).

#### C.5. Perceived Levels of Freedom of Expression, to Join Political Organizations, and to Criticize Government Actions or Performance

PR 26. This is a version of a question we have tested earlier, and we experienced that respondents had problems judging who is “people like me”? Would also be more specific if asked “How free do you feel to...” instead of others, we let people speak on their own behalf.

What does “without fear” mean? Fear from whom or what? It is rather unclear.

PR 27. There is a mismatch between question and response options. Maybe cut the “following institutions” and replace with the different institutions listed in the sub-questions? How free would you say that MEDIA in this country is to...

We suggest harmonizing the text on all three questions a-c for example “speak freely and openly criticize the government” and reuse this text for all three different institutions. The point about these questions is to compare how free the different institutions are – hence the context should be presented in the same way – now, the sub questions a-c are worded differently and this could affect how respondents answer (and make the responses not comparable). If not worded in the same way, it will not make sense to compare them? This way the question comes only once, and the type of institutions are read/appear on the screen in turn.

PR 27b. “organizations” twice (+ in intro), cut one of them, and shorten the rest of the text. Too much text to fit a small mobile screen.

#### D. Self-Identification

PR 29. This is a version of an already established question. Consider using the standard wording (In your experience, do you...” or “Do you feel a belonging to...”

PR 30. Instead of repeating “minority” in all the alternatives, it can be used in the question once, for example “Which minority group would that be?”

## 5. Suggestions for further work

As the survey will be implemented in different modes depending on for example the infrastructure of each country, it would be important to adjust texts in the questionnaire so that they make sense regardless of mode. A related questionnaire design issue is that we recommend reducing length of texts in the questionnaire for several purposes. Firstly, to make sure the questions can be implemented also on small mobile screens without further adaptations. Secondly, because it reduces interview time and survey costs. Thirdly, because respondents tend to look for the information they need to respond to a question, they rarely read text thoroughly if it's considered too long, superfluous etc.

The order of how response options are presented should be subject to thorough considerations as the presentation order might affect how the respondents perceive, relate, and respond to questions. In some cases (this also depends on mode) the most common response option should be presented first, while in other situations (for example if we ask for negative/sensitive behaviour) we might want to start the list of response options with the least socially acceptable one. How we order response options may be affected by how sensitive we regard the question to be, and in which mode we plan to present the survey in. This should be considered for each individual question.

There are two ways to reduce the risk of measurement errors. Firstly, to develop a questionnaire in accordance with what is considered current best methods. Current best methods are what we advocate for in this report. Secondly, to user test the questionnaire. It is not possible to design an optimal questionnaire solely based on desktop development and reviews. Respondents thought structures are too complex to be taken fully into account without actually talking to them. If it is needed to accelerate the questionnaire development and testing process, we would recommend prioritizing qualitative/cognitive testing over protocol testing (see Appendix 1 – “Test Design for Praia Questionnaire” for a description of the different test steps). The recommendation is based on our experience of which methods produce the most valuable data and insights about response quality in a questionnaire development process. From cognitive testing we learn about how respondents' response process looks like on each question in an in-depth way that reveals to methodologist if the respondent provides a response that is in accordance with the intentions of the surveyor.

This report includes a consideration of two different versions of the same questionnaire. All countries implement the same questions regardless of registration system, it is only the sequencing of questions that differ. Option 1 starts with a question about voter eligibility. Whereas option 2 starts directly with a question about whether the respondent voted or not. User testing these two different options, will help us answer two important questions – 1) Would it produce data of better quality if the questionnaire start with question on voting or question on eligibility?, and 2) Should we recommend a rigid questionnaire structure for all implementing countries, or can the order of questions differ from country to country depending on what one might find evidence for in the user tests? An argument that supports a structure tailored to each country could be that an initial question to measure eligibility could be very relevant in some countries (for example with manual registration systems, or non-existent public registers to draw the sample from) but would be equally irrelevant in others. On the other hand, in a country specific tailored structure, we risk a problem of cross country/survey comparison.

Regarding the two questionnaire options, we do not currently have evidence that suggest that these questions (about eligibility to vote/whether a respondent voted or not) are sensitive in a way that makes us concerned about any significant order effects from different versions being implemented in different countries. Hence, we cannot currently recommend neither a rigid nor a flexible structure. This is however, impossible to decide before the user testing.

Cognitive testing might tell us if respondents would find it logical that a survey that is titled “Survey on Participation in Political and Public affairs” start with a question to determine if they voted or not. On the other hand, a question on voting behaviour is likely to be perceived as sensitive in some (many?) countries, and in such cases better be placed later in the questionnaire (question number 2), after a relationship of trust is established between the respondent and surveyor/interviewer. To complicate the question of order even more, a start question on voting behaviour might trigger respondents' interest to (go through with) participate in the survey, and this question might also match well the expected content to the title of the survey. This is however likely to vary depending on political interest and participation also across different subgroups in a population. Persons who are politically active, with socially acceptable political views and actions, might be triggered while for persons who do not care about politics or who

might have controversial attitudes will find it threatening to get a question on voting behaviour without a “warm-up” on eligibility first.

We cannot conclude on possible order effects and a consequential cross survey comparison problem without having performed any user testing of the instrument. Hence, we suggest that this rather important question of how to structure the optimal questionnaire should be subject for user testing and revisited later in methodology discussions. As sensitivity issues should be considered culture specific, it would be of utmost importance for insurance of the questionnaire and data quality, to do proper user testing in several – at least 2-3- countries with different political systems and structures. SSBs methodology division would be happy to provide further guidance in how this may be planned and implemented to guarantee a solid outcome of this very important development process.

### **5.1. Possible topics and questions to investigate through user testing**

- Interpretation of “political and civic”
- Interpretation of scale for feeling of safety
- Volunteer election related activities – relevance and exhaustiveness of alternatives offered
- Active/passive “following” of online discussions, exhaustive response categories?
- Interpretation of 9c – how active or systematic must this be to be counted as a yes?
- Relevance and coverage – are there other topics than those mentioned in the questionnaire that the respondents miss, or believe should be covered in the questionnaire to make sure their responses produce a correct image of them?
- Attending strike or demonstrations – respondents’ interpretation of level of engagement
- Perceived safety of other people related to participation in demonstrations etc
- Examples and their effect on respondents’ interpretations

## References

[Brødtekst]