

**Comparative Study of Electoral Systems (CSES)
Module 5: Design Report (Sample Design and Data Collection Report)**

September 14, 2016

Country: **India**

Date of Election: **11 April 2019 – 19 May 2019 (7 phases in total)**

Phase I: 11 April 2019

Phase II: 18 April 2019

Phase III: 23 April 2019

Phase IV: 29 April 2019

Phase V: 6 May 2019

Phase VI: 12 May 2019

Phase VII: 19 May 2019

Prepared by:

Dhananjai Joshi (Cicero Associates) and Irfan Nooruddin (Georgetown University)

Date of Preparation: **04 February 2023**

NOTES TO COLLABORATORS:

- Where brackets [] appear, answer by placing an “X” within the appropriate bracket or brackets.
- If more space is needed to answer any question, please lengthen the document as necessary.

Collaborator(s):

Collaborators are the contact persons for election studies that appear in the CSES dataset - they are not necessarily the parties who collected the data. These collaborators and their contact information will be listed on the CSES website.

Name: Irfan Nooruddin Title: Professor Organization: Georgetown University Address: PO Box 571040 SFS Asian Studies Program Georgetown University Washington, DC 20057 USA Telephone: +1-202-687-1626 Fax: N/A E-Mail: in62@georgetown.edu Website: www.irfannooruddin.org	Name: Dhananjai Joshi Title: Founder / Director Organization: Cicero Associates & Consultants Pvt Ltd Address: A7/3 Saket, New Delhi, 110017, India Telephone: +91-9999300870 Fax: N/A E-Mail: dhananjai.joshi@ciceroassociates.co.in Website: www.ciceroassociates.co.in
---	--

Name: Title: Organization: Address: Telephone: Fax: E-Mail: Website:	Name: Title: Organization: Address: Telephone: Fax: E-Mail: Website:

Data Collection Organization:

Organization that conducted the survey field work/data collection:

Organization: **Cicero Associates & Consultants Pvt Ltd**

Address: **A7/3 Saket, New Delhi, 110017, India**

Telephone: +91-9999300870

Fax: N/A

E-Mail: **ghananjai.joshi@ciceroassociates.co.in**

Website: **www.ciceroassociates.co.in**

Funding Organization(s):

Organization(s) that funded the data collection:

Organization: **Georgetown University**

Address: **PO Box 571040**

SFS Asian Studies Program

Georgetown University

Washington, DC 20057

USA

Telephone: **202-687-1626**

Fax: **NA**

E-Mail: **in62@georgetown.edu**

Website: **www.georgetown.edu**

Organization:

Address:

Telephone:

Fax:

E-Mail:

Website:

Organization: Address: Telephone: Fax: E-Mail: Website:
--

Archiving Organization

If appropriate, please indicate the primary location where the full, original election study dataset (not just the CSES portion) will be archived:

Organization: Cicero Associates & Consultants, Pvt Ltd Address: A7/3 Saket, New Delhi, 110017, India Telephone: +91-9999300870 Fax: N/A E-Mail: dhnanjai.joshi@ciceroassociates.co.in Website: www.ciceroassociates.co.in

Please indicate the date when the study is expected to be available at this archive:

15 February 2023

Study Design

1. Timing of the study that the CSES Module was included in:

- Post-Election Study (with interviewing starting within 6 months after the election)
- Post-Election Study (with interviewing starting more than 6 months after the election)
- Pre-Election/Post-Election Panel Study
- Between Rounds

2a. Date Post-Election Interviewing Began:

15-May-2019

2b. Date Post-Election Interviewing Ended:

26-May-2019

3a. Mode of interviewing for the post-election survey in which the CSES Module appeared:
(If multiple modes were used, please mark all that apply.)

- In person, face-to-face - using a questionnaire on paper
- In person, face-to-face - using an electronic/computerized questionnaire
- Telephone
- Mail or self-completion supplement
- Internet

3b. Was there a mode change *within* interviews (e.g., selected self-completion elements within the questionnaire)?

- No
- Yes; please provide details:

4a. Was the survey part of a panel study?

- Yes
- No

4b. If the survey was part of a panel study, please describe the design of the panel study, including the date at which interviewing for each prior wave began and ended:

NA

4c. If the survey was entirely or partly conducted via the Internet, please indicate whether it was based on an access panel (i.e. respondents were selected from a group of pre-screened panelists):

- Yes
- No

4d. If the survey was based on an Internet access panel, please describe the access panel (company, population [does it include persons without initial access to the Internet and how are they interviewed], method of recruiting members, total size of access panel, method of selecting survey respondents from the panel):

NA

Translation

Please provide copies of questionnaires in all languages used as part of the election study deposit. For questionnaires in a language other than English, please also provide a version of each translated back into English. Note: Questions are based on those developed for the ISSP.

5. Was the questionnaire translated?

- Yes, translated by member(s) of research team
- Yes, by translation bureau
- Yes, by specially trained translator(s)
- No, not translated

6. Please list all languages used for the fielded module:

1. **Assamese**
2. **Bengali**
3. **Gujarati**
4. **Hindi**
5. **Kannada**
6. **Malayalam**
7. **Marathi**
8. **Oriya**
9. **Punjabi**
10. **Tamil**
11. **Telugu**

7a. If the questionnaire was translated, was the translated questionnaire assessed/checked or evaluated?

- Yes, by group discussion
- Yes, an expert checked it
- Yes, by back translation
- Other; please specify: _____
- No
- Not applicable

7b. If the questionnaire was translated, was the questionnaire pre-tested?

- Yes
- No
- Not applicable

7c. If the questionnaire was translated, were there any questions which caused problems when translating?

- Yes
- No
- Not applicable

7d. If the questionnaire was translated, please provide a list of all questions which caused problems when translating. For each question listed, describe what problems were encountered and how they were solved:

NA

Sample Design and Sampling Procedures

8. Please describe the population that your sample is meant to be representative of:

Cross-section; all eligible voters in 19 largest states of India covering 96.02% of the electorate. All respondents were sampled from the latest official Electoral Rolls of India.

Eligibility Requirements

9a. Must a person be a certain age to be interviewed?

- Yes
 No

If yes, what ages could be interviewed?

18 years & above

9b. Must a person be a citizen to be interviewed?

- Yes
 No

9c. Must a person be registered to vote to be interviewed?

- Yes
 No

9d. Please list any other interviewing requirements or filters used:

NA

Sample Frame

10a. Were any regions of the country excluded from the sample frame?

- Yes
 No

If yes, what percent of the total eligible population did this exclude from the sample frame?

96.02 %.

If yes, please explain:

India has a total of 543 Parliamentary Constituencies. Of them 509 constituencies are in the 19 largest states that were covered in the survey. The states that were left outside the survey are part of the hill states of the Himalayan range or small Union Territories that have five or less Parliamentary Constituencies.

10b. Were institutionalized persons excluded from the sample?

- Yes
 No

If yes, what percent of the total eligible population did this exclude from the sample frame?

NA

If yes, please explain:

NA

10c. Were military personnel excluded from the sample?

- Yes
 No

If yes, what percent of the total eligible population did this exclude from the sample frame?

NA

If yes, please explain:

NA

10d. If interviews were conducted by telephone, what is the estimated percentage of households without a phone?

0%

Please explain:

NA

10e. If interviews were conducted by telephone, were unlisted telephone numbers included in the population sampled?

- Yes
 No

If no, what percent of the total eligible population did this exclude from the sample frame?

0%

10f. If interviews were conducted via the Internet, what is the estimated percentage of households without access to the Internet?

0%

10g. If interviews were conducted via the Internet, were provisions taken to include members of the population without access to the Internet? And if so, which?

- Yes
 No

If “Yes”, please explain:

NA

If “No”, what percent of the total eligible population did this exclude from the sample frame?

NA

10h. Were other persons excluded from the sample frame?

Yes

No

If yes, what percent of the total eligible population did this exclude from the sample frame?

3.98 %

If yes, please explain:

The states that were left outside the survey are part of the hill states of the Himalayan range or small Union Territories that have five or less Parliamentary Constituencies.

10i. Please estimate the total percentage of the eligible population excluded from the sample frame:

3.98 %

Sample Selection Procedures

12. Please describe, in your own words, how the sample for the study was selected. If the survey is part of a panel study and/or based on an Internet access panel, please also describe the original sample, from the beginning of the study.

The survey employed a Multi-stage Stratified Random Sample Design. At the first stage 336 Assembly Segments were sampled by employing the Probability Proportionate to Size (PPS) random sampling technique. The Assembly Segments were finalized after checking demographic and political representativeness against demographic data from the Census of India 2011 and past election results from the Election Commission of India. Within each sampled Assembly Segment, 2 Polling Booths were randomly sampled by employing the systematic random sampling method. The latest Electoral Rolls for all 672 sampled Polling Booths were downloaded from the Election Commission of India website. Sampled Respondent’s Lists of 30 voters were made from the Electoral Rolls again by employing the systematic random sampling method.

12a. What were the primary sampling units?

Assembly Segments within states

12b. How were the primary sampling units selected?

PPS (Probability Proportionate to Size)

12c. Were the primary sampling units randomly selected?

Yes

No

Please explain how the units were randomly selected. If the units were not randomly selected, please provide a justification for why the units were not randomly selected.

At the first stage 336 Assembly Segments were sampled by employing the Probability Proportionate to Size (PPS) random sampling technique. The Assembly Segments were finalized after checking demographic and political representativeness against demographic data from the Census of India 2011 and past election results from the Election Commission of India.

13. Were there further stages of selection?

Yes

No

13a. If there were further stages of selection, what were the sampling units at each of the additional stages?

Polling Stations within Assembly Segments

13b. If there were further stages of selection, how were the sampling units selected at each of the additional stages?

Within each sampled Assembly Segment, 2 Polling Booths were randomly sampled by employing the systematic random sampling method.

13c. If there were further stages of selection, were units at each of these stages randomly selected?

Yes

No

Please explain how the units were randomly selected. If the units were not randomly selected, please provide a justification for why the units were not randomly selected.

NA

14a. How were individual respondents identified and selected in the final stage?

The latest Electoral Rolls for all 672 sampled Polling Booths were downloaded from the Election Commission of India website. Sampled Respondent's Lists of 30 voters were made from the Electoral Rolls again by employing the systematic random sampling method.

14b. Could more than one respondent be interviewed from a single household?

- Yes
 No

If yes, please explain:

It is possible that more than one respondent was interviewed from a single household. This would happen primarily because 'household' was not a sampling unit. The individuals were listed from the Electoral Rolls and there is a possibility of more than one respondent from a single household in case of bigger households in which large families reside.

15. Did the sample design include clustering at any stage?

- Yes
 No

If yes, please describe:

NA

16. Did the sample design include stratification?

Definition: Stratification involves the division of the population of interest according to certain characteristics (for instance: geographic, political, or demographic). Random selection then occurs within each of the groups that result.

- Yes
 No

If yes, please describe (please include the list of characteristics used for stratification, and in the case of multi-stage selection processes the stage[s] at which stratification occurred):

- **The 19 largest states are all selected for the sample.**
- **Within each of these states, a random sample of assembly constituencies is drawn.**
- **Within selected assembly constituencies, a random sample of polling stations is drawn.**
- **At the polling station level, a random sample of eligible voters is drawn from the electoral rolls.**

17. Was quota sampling used at any stage of selection?

- Yes
 No

If yes, please describe:

NA

18. Was substitution of individuals permitted at any stage of the selection process or during fieldwork?

- Yes
 No

If yes, please describe:

Substitution was permitted within a polling station. A separate list of randomly sampled substitute respondents was maintained.

19. Under what circumstances was a household designated non-sample? Please check all that apply:

- Non-residential sample point
- All members of household are ineligible
- Housing unit is vacant
- No answer at housing unit after 2 callbacks
- Other (Please explain):

20. Were non-sample replacement methods used?

- Yes
- No

Please describe:

NA

21a. For surveys conducted by telephone, was the sample a random digit dial (RDD) sample?

- Yes
- No

21b. For surveys conducted by telephone, was the sample a listed sample?

- Yes
- No

21c. For surveys conducted by telephone, was the sample a dual frame sample?

- Yes
- No

If yes, what % list frame _____ and what % RDD _____

NA

22. For surveys conducted by mail, was the sample a listed sample?

- Yes
- No

Please describe:

NA

23. For surveys conducted on the Internet, did respondents self-select into the survey, at any stage?

- Yes
 No

Please explain:

NA

Incentives

24a. Prior to the study, was a letter sent to the respondent?

- Yes
 No

(If yes, please provide a copy of the letter.)

NA

24b. Prior to the study, was a payment sent to the respondent?

- Yes
 No

If yes, please describe (including amount of payment):

NA

24c. Prior to the study, was a token gift sent to the respondent?

- Yes
 No

If yes, please describe:

NA

24d. Did respondent receive an additional payment after their participation? (Do not include any payment made prior to the study.)

- Yes
 No

If yes, please describe (including amount of payment):

NA

24e. Were any other incentives used?

- Yes
 No

If yes, please describe:

NA

Interviewers

25. Please describe the interviewers (e.g., age, level of education, years of experience):

Cicero Associates hires a pool of interviewers. The pool consists of university students or recent graduates. The age range of the pool typically is 19-30 years. All interviewers are at least high school graduates pursuing a university education or have their university degree. Years of experience varies from 0 to 10 years. Due to the language diversity, each state has a unique set of interviewers.

Approximately 100 interviewers were hired in total to conduct this survey.

26. Please provide a description of interviewer training. If possible please differentiate between general interviewer training and study-specific components:

Cicero Associates provides a mandatory day-long training held at the state-level. Trainings are in-person and conducted one-on-one by the state supervisor and a trainer from the central team based out of Delhi.

General training is provided to all interviewers on how to approach respondents; how to obtain informed consent; and how to deal with unexpected situations that might arise at the local level.

For the CSES survey, each question is discussed in detail after which the interviewer conducted a mock interview and received feedback.

26a. Please provide a description of the content, structure and time used for general training of interviewers:

See response to Q26.

26b. Please provided a description of the content, structure and time used for training interviewers in the specifics of the study within which CSES was run:

See response to Q26.

Contacts

27a. What was the average number of contact attempts made per household, for the entire sample?

1

27b. For households where contact was made, what was the average number of contact attempts prior to first contact?

0

27c. During the field period, how many contacts were made with the household before declaring it a **non-sample**?

2

28d. During the field period, how many contacts were made with the household before declaring it a **non-interview**?

2 (see 27c)

28e. During the field period, what were the maximum number of days over which a household was contacted?

2

28f. During the field period, did interviewers vary the time of day at which they re-contacted the household?

Yes

No

If yes, please describe:

Interviewers typically spend 2 days in each Assembly Segment before moving to the next. If a household was not contactable on day 1, the interviewer tried again on day 2. The timing of the re-contact depended on the schedule for the day, which introduces natural variation in the timing.

Refusal Conversion

29a. Were efforts made to persuade respondents who were reluctant to be interviewed?

Yes

No

Please describe:

Interviewers are trained to gently persuade reluctant respondents but if facing resistance are instructed to thank the respondent for their time and move to the next household on their list.

29b. Were respondents who were reluctant to be interviewed sent a letter persuading them to take part?

Yes

No

(If yes, please provide a copy of the letter or letters.)

If yes, please describe:

NA

29c. Was payment offered to respondents who were reluctant to take part?

- Yes
 No

If yes, how much?

NA

29d. Were respondents who were reluctant to take part turned over to a more experienced interviewer?

- Yes
 No

29e. What was the maximum number of re-contacts used to persuade respondents to be interviewed?

None. See response to 29a above.

29f. Were any other methods used to persuade respondents reluctant to be interviewed to take part?

- Yes
 No

If yes, please describe:

See response to 29a above.

Interview/Survey Verification

Definition: Interview/survey verification is the process of verifying that an interview was conducted and that the survey was administered to the correct respondent, for quality control purposes.

30. Was interview/survey verification used?

- Yes
 No

If yes, please describe the method(s) used:

- **GPS coordinates of the interview must be in the geographical fence of the Polling Booth**
- **15% interviews Back Check by the appointed supervisor**

If yes, please indicate the percent of completed surveys that were verified:

15%

Response Rate

Note: If multiple modes of interviewing were used for the post-election survey in which the CSES Module appeared, please repeat the following questions as appropriate for each of the modes used.

31. What was the response rate of the survey that the CSES Module appeared in? Please show your calculations. (If the CSES Module appeared in a panel study, please report the response rate of the first wave of the study, even if the CSES Module did not appear in that wave.)

Sl.No.	State_Name	Target n/State	Achieved n/State	Achieved n/PS
1	ASSAM	720	544	75.56%
2	BIHAR	1440	1057	73.40%
3	GUJARAT	960	665	69.27%
4	HARYANA	720	520	72.22%
5	KARNATAKA	960	704	73.33%
6	KERALA	960	640	66.67%
7	MADHYA PRADESH	960	643	66.98%
8	MAHARASHTRA	1440	990	68.75%
9	ODISHA	960	617	64.27%
10	PUNJAB	720	503	69.86%
11	RAJASTHAN	960	674	70.21%
12	TAMIL NADU	1440	1047	72.71%
13	UTTAR PRADESH	1920	1291	67.24%
14	WEST BENGAL	1440	917	63.68%
15	DELHI	720	525	72.92%
16	JHARKHAND	720	484	67.22%
17	CHHATTISGARH	720	487	67.64%
18	TELENGANA	960	607	63.23%
19	ANDHRA PRADESH	1440	1048	72.78%
	ALL INDIA	20,160	13,963	69.26%

32. Please provide the following statistics for the survey that the CSES Module appeared in. (If the CSES Module appeared in a panel study, please report the statistics for the first wave of the study, even if the CSES Module did not appear in that wave.)

- | | |
|---|-----------|
| A. Total number of households in sample: | <u>NA</u> |
| B. Number of valid households: | <u>NA</u> |
| C. Number of invalid (non-sample) households: | <u>NA</u> |
| D. Number of households of unknown validity: | <u>NA</u> |

Note: 'Household' was not a sampling unit. A Polling Booth which has approximately 1000 to 1500 voters was the lowest sampling unit. Within each Polling Booth 30 respondents were sampled. Thus there is no data at the Household level.

- | | |
|--|--------------|
| E. Number of completed interviews: | <u>13963</u> |
| F. Number of partial interviews:* | <u>0</u> |
| G. Number of refusals and break-offs: | <u>6167</u> |
| H. Number non-contact (never contacted): | <u>30</u> |
| I. Other non-response: | <u>0</u> |

**** Partial or incomplete interviews are not part of the data set. Such interviews were terminated after 2 attempts to complete the interview.***

The sum of B+C+D should equal the value of A. If not, please describe why:

If statistic D (number of households of unknown validity) has a value greater than zero (0), please estimate the proportion of households of unknown validity that are valid:

The sum of E+F+G+H+I should equal the value of B. If not, please describe why:

If statistic I has a value greater than zero (0), please describe what cases fall into this category:

33. If the CSES Module appeared in a panel study, how many waves were conducted prior to the wave that included the CSES Module?

Not a panel study. This was a single wave study.

34. If the CSES Module appeared in a panel study, what was the total panel attrition between the first wave of the study and the wave that included the CSES Module? Please show your calculations.

Not a panel study. This was a single wave study.

35. If the CSES Module appeared in a panel study, please provide the number of completed interviews for the wave that included the CSES Module:

NA

36. If the CSES Module appeared in a panel study, please provide the following statistics for panel attrition by age and education. In each cell, indicate the percent of all completed interviews in each category for the indicated wave.

Age	First wave of study	Wave that included CSES
18-25		14.40 %
26-40		42.27 %
41-64		35.79 %
65 and over		07.55 %

Education *	First wave of study	Wave that included CSES
None		5%
Incomplete primary		18%
Primary completed		18%
Incomplete secondary		15%
Secondary completed		19%
Post-Secondary Trade/Vocational		7%
University incomplete		4%
University degree		14%

*** Response categories in the survey are different from the ones given above. These are equivalents at best.**

37. Are weights necessary to make the sample representative of the population being studied?

- Yes
 No

If yes, please explain:

NA

38. Are weights included in the data file?

- Yes
 No

Note by the CSES-Secretariat: Collaborators kindly deposited a combined sample, demographic and political weight upon request. The weight adjusts the sample to match known demographic characteristics (state, gender, locality, religious denomination, caste groups) and vote choice. A detailed description of the weight as deposited by Collaborators is available in the Appendix (p. 23 and 24).

39. If weights are included in the data file, please describe in detail how the weights were constructed:

NA

40a. If weights are included in the data file, are the weights designed to compensate for disproportionate probability of selection?

- Yes
 No

If yes, please describe:

NA

40b. If weights are included in the data file, are the weights designed to match known demographic characteristics of the population?

- Yes
 No

If yes, please describe:

NA

40c. If weights are included in the data file, are the weights designed to correct for non-response?

- Yes
 No

If yes, please describe:

NA

40d. If weights are included in the data file, are the weights designed to correct to the official election results?

- Yes
 No

If yes, please describe:

NA

41. Comparison of Completed Interviews to Population (please provide as percentages of the total):

Characteristic	Population Estimates	Completed Interviews	
		Unweighted Distribution	Weighted Distribution
<u>Age</u>			
18-25	%	14 %	%
26-40	%	42 %	%
41-64	%	36 %	%
65 and over	%	08 %	%
<u>Education *</u>			
None	%	5%	%
Incomplete Primary	%	18%	%
Primary Completed	%	18%	%
Incomplete Secondary	%	15%	%
Secondary Completed	%	19%	%
Post-Secondary Trade/ Vocational	%	7%	%
University Incomplete	%	4%	%
University Degree	%	14%	%
<i>* Response categories in the survey are different from the ones given above. These are equivalents at best.</i>			
<u>Gender</u>			
Male	%	56 %	%
Female	%	44 %	%

42. Please indicate the source of the population estimates in the prior question. English language sources are especially helpful. Include website links or contact information if applicable.

See response to Q41 above.

Appendix: Description of Combined Sample, Demographic and Political Weight

Note on Weights for India Dataset

Variable Name: "Ptw"

Variable Label: "Ptw: Multivariate Weights for Demographics and Vote 2019".

A variable for multivariate weights has been added to the India dataset. The weights control for discrepancies in demographics and vote estimates. In total six variables have been weighed. As weights are applied on multiple variables small discrepancies (less than one percent) continue.

The first variable on which weights are applied is 'State'. Weights on actual proportion of electorate in each sampled state have been applied to control for over-representation or under-representation of any state in the dataset. The survey covered 19 big states of India that represent 96.02 percent of the total electorate. India has a total of 543 Parliamentary Constituencies. Of them 509 constituencies are in these 19 largest states covered in the survey. The states that were left outside the survey are part of the hill states of the Himalayan range or small Union Territories that have five or less Parliamentary Constituencies. These states represent 3.98 percent of total electorate. The weights adjust this 3.98 percent proportionately to the size of the sampled states.

After applying weights to State, weights are applied to a series of demographic variables. These are; 'Respondent's Gender', followed by 'Locality', followed by 'Respondent's Religion' and then 'Respondent's Caste Group'. All these variables are weighed at the national level and not state level. Once all demographic variables are adjusted, final weights are applied to the voting variable. For this purpose, parties are recoded into major alliances at the national level. Weights are finally applied to the recoded variable.

Table 1:

Gender	Census 2011	Raw-Survey	After Weight
Male	51.5	56.6	51.3
Female	48.5	43.4	48.7

Table 2:

Locality	Census 2011	Raw-Survey	After Weight
-----------------	--------------------	-------------------	---------------------

Rural	68.9	69.1	68.5
Urban	31.1	30.9	31.5

Table 3:

Religion	Census 2011	Raw-Survey	After Weight
Hindu	79.8	82.0	80.6
Muslim	14.2	12.4	13.4
Christian	2.3	1.7	2.3
Sikh	1.7	2.5	1.7
Others	2.0	1.4	2.0

Table 4:

Caste Group	Census 2011	Raw-Survey	After Weight
SC	16.6	16.8	16.4
ST	8.62	8.5	8.8
Others	74.8	74.7	74.8

Table 5:

Party Vote	Election Commission	Raw-Survey	After Weight
Congress	19.5	23.4	19.5
Congress Allies	7.3	7	7.3
BJP	37.4	34.2	37.4
BJP Allies	7.5	7.2	7.5
BSP+	6.5	6.6	6.5
Left	2.1	2.3	2.1
Others	19.7	19.3	19.7