

Comparing Phone and Web Surveys: How Methodology Shapes Educational Distributions and Election Interest

STA304 - Winter 2025 - Assignment 1

Umi Yamaguchi - 1008977369

1 Introduction

This report compares phone and web survey methodologies using data from the 2019 Canadian Election Study (CES) [1]. The central goal is to explore how the mode of data collection might influence demographic representation and reported political attitudes. Differences in respondent demographics and reported levels of political interest can lead to distinct inferences about Canadian voters.

For the demographic variable, I selected Education (categorized into four groups), which is a key predictor of many social and political outcomes. For the outcome of interest, I chose Interest in the Canadian federal election (scored from 0–10). Since election interest strongly correlates with political participation and voting behavior, comparing it between phone and web respondents illustrates how survey mode may shape one’s understanding of the electorate. In these analyses, the target population is Canadian adults eligible to vote (18+), the frame population is individuals reachable by phone or with reliable internet access, and the sample population consists of respondents who agreed to complete the phone or web-based CES questionnaire. Any differences across these populations will be central to interpreting the results.

2 Data

The analysis utilizes two cleaned CES datasets: one from the phone survey (`ces_phone.csv`) and another from the web survey (`ces_web.csv`). Both contain demographic information (e.g., age, education) and a measure of interest in the 2019 Canadian federal election (0 = No interest, 10 = High interest). Several steps ensured that the data focused on eligible and valid responses. Participants under 18 were removed, “Don’t know” and “Prefer not to answer” responses for education were excluded, and invalid or missing interest responses were dropped. Education was then grouped into four categories (Below High School, Secondary Diploma, Some College/University, Postsecondary Degree), reflecting broad levels of educational attainment [2]. Interest in the election was classified into Low (0–4), Moderate (5–7), or High (8–10).

Because the web survey dataset initially had more respondents than the phone survey, the both dataset was randomly downsampled to match the same lengths. This balanced approach helps ensure that any differences observed are not simply due to vastly different sample sizes. The resulting

datasets contain equally sized phone and web samples, allowing direct comparisons between the two survey modes [3].

3 Demographic Variables

The demographic variable under examination is Education, which is significant because individuals' educational attainment is associated with political knowledge, voter turnout, and policy preferences [5]. By comparing the distribution of education levels in the phone survey versus the web survey, we can observe whether one mode tends to draw participants of a particular educational background. Measuring the target population (all eligible Canadian voters) through different survey modes introduces potential biases in how well each method represents the broader electorate. The frame population, consisting of individuals reachable by phone or with reliable internet access, may not fully capture certain demographic groups, such as older individuals who may be more likely to respond to phone surveys or younger, tech-savvy individuals who prefer web surveys. These accessibility differences can influence the educational composition of respondents, leading to an overrepresentation of individuals with higher education levels in phone surveys and those with intermediate educational backgrounds in web surveys.

Figures present side-by-side bar charts of education categories (Figure 1 and Figure 2). One chart displays the raw counts for each education group, while the other shows the proportions within each survey type. The sample population, composed of those who actually participated in each survey, reflects the biases inherent in each mode. The phone survey, for instance, tends to include a larger share of respondents in the Postsecondary Degree category. In contrast, the web survey includes more participants in the Some College/University group. These discrepancies suggest that different sampling or accessibility mechanisms—whether related to age, technological familiarity, or respondent preferences—are influencing who ends up in each dataset. Recognizing these differences is important when drawing broader conclusions about Canadians' educational profile. Understanding the effects of survey mode on population representation highlights the importance of considering these biases when interpreting results and making inferences about the entire population [4].

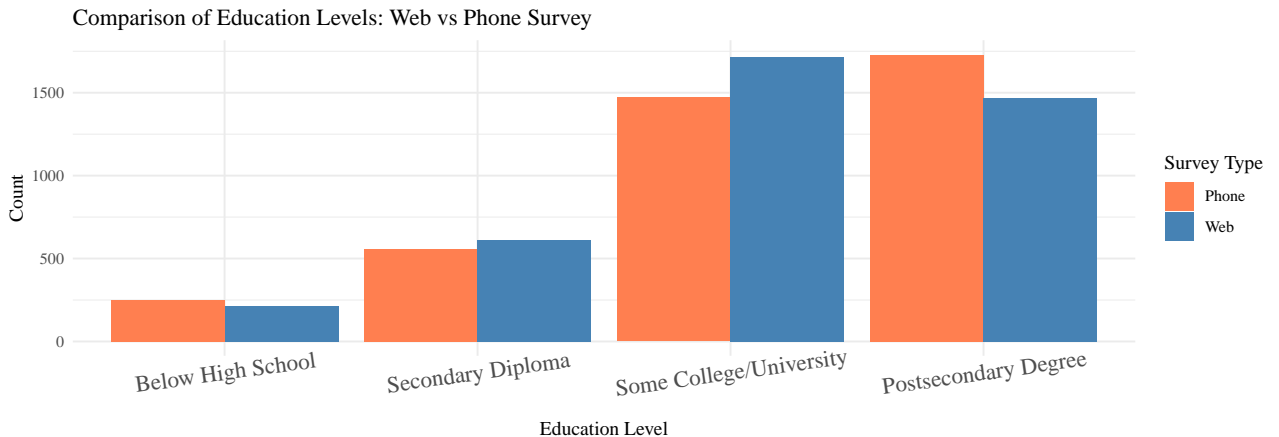


Figure 1: Comparison of Education Levels by Survey Mode. The bar chart shows the distribution of education levels across phone and web survey respondents, highlighting differences in sample composition.

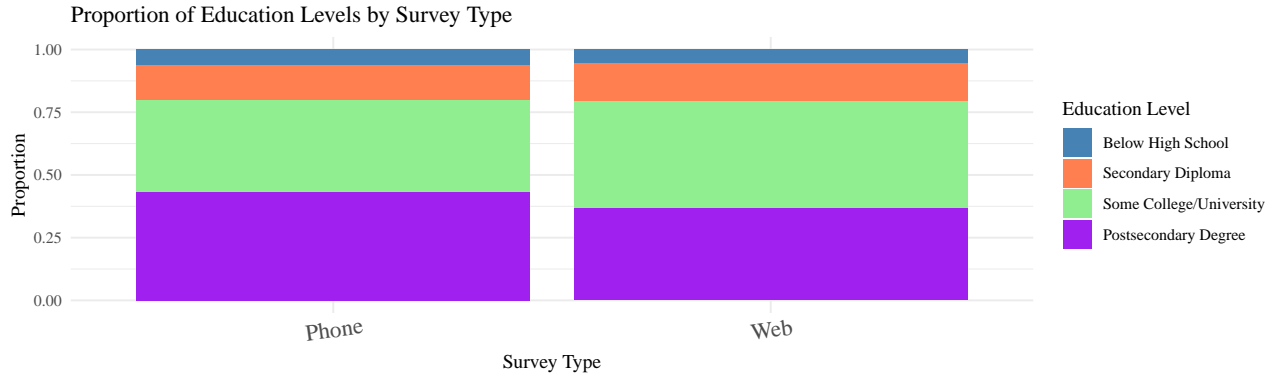


Figure 2: Proportion of Education Levels by Survey Type. This stacked bar chart displays the relative proportions of education levels within each survey type (phone vs. web), highlighting differences in respondent composition across educational categories.

4 Outcome of Interest

The outcome variable is Interest in the Canadian federal election, measured on an 11-point scale (0–10). This outcome sheds light on public engagement in politics; in general, higher interest predicts greater likelihood of voting. Given that survey mode can affect who responds, we compare whether phone or web surveys produce distinct estimates of election interest.

A table in the final report summarizes the proportion of respondents classified as “High Interest” (8–10) for each survey mode. A 95% confidence interval (CI) for these proportions was computed using the typical formula for a proportion:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

where \hat{p} is the sample proportion of “High Interest” respondents, $z_{\frac{\alpha}{2}} \approx 1.96$ for a 95% confidence level, and n is the sample size.

The Figure 3 illustrates differences in reported political interest levels between respondents from phone and web surveys. The phone survey shows a higher number of respondents reporting High Interest, whereas the web survey captures more participants in the Low Interest category. This suggests potential differences in self-selection bias, with phone respondents possibly being more engaged in political matters than their web-based counterparts. The moderate interest levels appear relatively similar across both survey types, indicating some consistency in mid-level engagement. These findings suggest that survey mode may influence the distribution of political interest responses, potentially affecting conclusions drawn from the data regarding overall public engagement in political processes.

The 95% confidence intervals (CIs) for the proportion of respondents reporting High Interest in the election reveal critical differences and insights:

	Proportion of Outcome Variable	95% Confidence Interval of Outcome Variable
Phone Survey	0.565	(0.55, 0.581)
Web Survey	0.499	(0.484, 0.515)

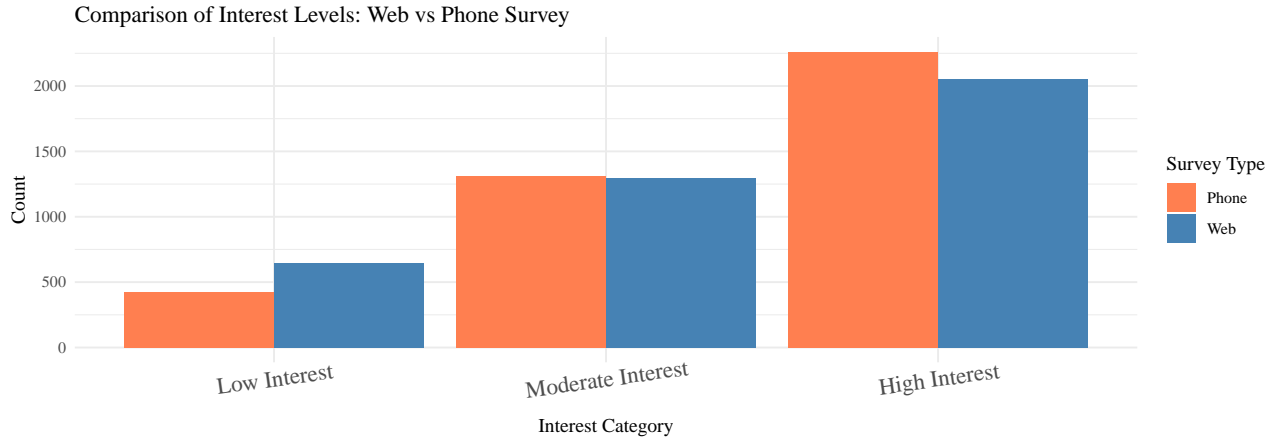


Figure 3: Comparison of Interest Levels by Survey Type. The bar chart compares political interest levels across phone and web survey respondents. The phone survey has more respondents with High Interest, while the web survey shows a slightly higher proportion of Low Interest respondents.

The CIs for phone ($0.550 \sim 0.581$) and web ($0.484 \sim 0.515$) surveys do not overlap, indicating a statistically significant difference in reported interest between modes. This suggests that survey mode systematically influences responses, rather than random sampling variation. The phone survey’s proportion of high-interest respondents (56.5%) is 6.6 percentage points higher than the web survey’s (49.9%). This gap exceeds the margin of error for both surveys (~ 3.1 percentage points), reinforcing the practical significance of the disparity. Lastly, both modes exclude populations without phone/internet access, potentially underestimating interest among marginalized groups.

While the phone survey suggests higher electoral engagement, this finding may not generalize to the broader Canadian electorate due to frame bias (e.g. exclusion of non-phone/internet users). The web survey’s lower interest estimates may better reflect younger demographics but underrepresent older voters. These discrepancies highlight the need for mixed-mode surveys or weighting adjustments to improve population-level inferences.

5 Comparative Analysis

A comparison of the two survey modes points to potentially substantial differences in demographic composition and political engagement. The phone sample contained a larger fraction of participants with higher formal education, which may reflect patterns of landline usage or the fact that older, more highly educated individuals are more likely to take part in phone surveys. Meanwhile, the web survey included more respondents who had only partial postsecondary education. This mode might be more appealing or accessible to younger, more technology-driven individuals, leading to a different educational distribution.

Various biases emerge from these observations. Mode differences can generate coverage errors, as each approach might miss segments of the population without consistent phone or internet access. Selection bias is another concern, since opting into one survey type may correlate with traits like age, education, or political predisposition. These discrepancies have direct implications for analyzing

election interest. If the phone survey systematically reaches more politically attentive participants (or those inclined to say they are interested), the reported interest levels may appear higher than in the web survey. Consequently, policy researchers or political campaigns relying on only one mode risk overestimating or underestimating true public engagement.

Despite these challenges, both surveys add valuable information when interpreted in context. In principle, analysts can employ weighting or combine multiple survey modes to mitigate skewed distributions. Reflecting on both phone-based and web-based responses enriches our understanding of Canadian voters by exposing the diversity of perspectives and, crucially, the influence of the data-collection method itself.

6 Generative AI Statement

I used an AI-assisted writing tool (ChatGPT) for language suggestions, sentence restructuring, and refining the clarity of my writing, including adjustments to visualization descriptions. AI was consulted to enhance readability in sections such as the introduction, comparative analysis, and figure captions. However, all statistical analysis, data cleaning, and visualizations were independently performed using RStudio and verified against course materials and external references.

AI suggestions were carefully evaluated and incorporated only when they aligned with assignment requirements and my understanding of the subject. The final work reflects my own effort and comprehension, ensuring compliance with academic integrity standards while acknowledging AI as a supplementary tool.

7 Bibliography

1. Hodgetts, P. (2023). cesR: Canadian Election Study Data in R. <https://hodgettsp.github.io/cesR/> (Accessed January 20, 2025).
2. Statistics Canada (2021). Educational attainment classification. <https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=305734&CVD=305735&CLV=0&MLV=4&D=1> (Accessed January 20, 2025).
3. Murel, J. IBM (2022). Downsampling in data analysis. <https://www.ibm.com/think/topics/downsampling> (Accessed January 20, 2025).
4. Wickham, H., Chang, W. et al., (2016). ggplot2: Elegant Graphics for Data Analysis. <https://ggplot2.tidyverse.org/reference> (Accessed January 20, 2025).
5. Weinschenk, A., Dawes, C., et al., (2021). Explaining the Education Gap in Knowledge of Political Facts. <https://www.tandfonline.com/doi/abs/10.1080/17457289.2021.1952416> (Accessed January 24, 2025).