Annenberg Public Policy Center
Civics Knowledge Survey
August 2022

Methods Report

1 Braxton Way Suite 125 Glen Mills, PA 19342
www.ssrs.com

METHODOLOGY

SSRS conducted the *Civics Knowledge Survey* on behalf of The Annenberg Public Policy Center (APPC) from August 2 through August 13, 2022. The goal of this survey was to understand the public's knowledge on U.S civics and government.

For the *Civics Knowledge Survey,* SSRS interviewed a representative sample of 1,113 adults (18 or older), reached via landline (n=224) and cell phone (n=889)[1] random digit dialing (RDD).

This report details the methodological components of the study: sample design, questionnaire design, programming, field operations, data processing, and weighting.

Sample Design

Phone numbers used for this study were randomly generated from landline and cell phone sampling frames, with an overlapping frame design. The random digit dialing (RDD) landline sample included all exchanges (NPA-NXX) associated with phone numbers in the U.S. The sample was generated through Marketing Systems Group's (MSG) GENESYS sampling system. The GENESYS RDD methodology produces a strict single stage, Equal Probability Selection Method (epsem) sample of residential telephone numbers. In other words, a GENESYS RDD sample ensures an equal and known probability of selection for every residential telephone number in the sample frame. The sample was generated shortly before the beginning of data collection to provide the most up-to-date sample possible, maximizing the number of valid telephone extensions.

Using a procedure similar to that used for the landline sample, MSG generated a random list of cell phone telephone numbers. Inactive numbers were flagged and removed utilizing MSG's CellWins procedure. To increase the share of lower-income and non-White respondents in the sample, telephone numbers flagged as prepaid (or, pay-as-you-go) were oversampled so that approximately 25% of the cell phone sample consisted of prepaid numbers (compared with approximately 12-15% that would otherwise be randomly sampled).

Table 1 summarizes the number of interviews completed by sampling method.

[1] Including 665 respondents who could be reached only by cell phone.

Table 1: Completed Interviews by Sampling Method

Sample Type	TOTAL
Landline:	224
Cell Phone	889
TOTAL	**1,113**

Questionnaire Design

The questionnaire was developed by APPC in consultation with the SSRS project team. The questionnaire was translated into Spanish so respondents could choose to be interviewed in English or Spanish, or switch between the languages according to their comfort level.[2]

Prior to the field period, SSRS programmed the study into Confirmit Computer Assisted Telephone Interviewing (CATI) software. Extensive checking of the program was conducted to ensure that skip patterns and sample splits followed the design of the questionnaire.

Field Procedures

Survey Administration

The field period for this study was August 2 through August 13, 2022. All interviews were completed through the CATI system. The CATI system ensured that questions followed logical skip patterns and that complete dispositions of all call attempts were recorded.

CATI interviewers received written materials about the survey instrument and received formal training for this particular project. The written materials were provided prior to the beginning of the field period and included an annotated questionnaire that contained information about the goals of the study, detailed explanations as to why questions were being asked, the meaning and pronunciation of key terms, potential obstacles to be overcome in getting good answers to questions, and respondent problems that could be anticipated ahead of time, as well as strategies for addressing the potential problems.

Interviewer training was conducted immediately before the study was launched. Call center supervisors and interviewers were walked through each question from the questionnaire. Interviewers were given instructions to help them maximize response rates and ensure accurate data collection.

[2] In total, 37 interviews were completed in Spanish.

In order to maximize survey response, SSRS enacted the following procedures during the field period:

- Up to 4 attempts were made to contact non-responsive numbers (e.g. no answer, busy, answering machine)
- Each non-responsive number was contacted multiple times, varying the times of day, and the days of the week that call-backs were placed using a programmed differential call rule
- Interviewers explained the purpose of the study and, when asked, stated as accurately as possible the expected length of the interview (20 minutes)
- Respondents were offered the option of scheduling a call-back at their convenience
- Specially trained interviewers contacted households where the initial call resulted in respondents hanging up the phone
- Cell phone respondents who mentioned the cost of usage were offered $10 to account for their cell phone minutes.

Response Rate

Response rate for the Civics Knowledge Survey was calculated using AAPOR's Response Rate 3 formula. This calculation divides the number of completed interviews in each sampling frame, by the estimated number of eligible phone numbers in the frame. For this study, response rate was calculated as 3%.

Data Processing and Deliverables

After the first 622 interviews, SSRS delivered interim unweighted preliminary data (SPSS).

At the end of the field period SSRS delivered a fully labeled SPSS dataset, two topline documents, and a detailed methods report.

Weighting Procedures

Data were weighted to represent adults 18+ of the United States. There were three steps to the weighting process. First, we applied a sampling weight to account for sampling probabilities. Next, we made an adjustment to account for the oversampling of prepaid cell phones. Finally, sample demographics were balanced to match target population benchmarks.

Sampling Weight

The first stage of the weighting was to apply a sampling weight which corrects for different probabilities of selection associated with the number of adults in each household and each respondent's telephone usage patterns. This adjustment also accounts for the overlapping landline and cell sample frames and the relative sizes of each frame and each sample.[3]

The sampling weight for the i^{th} case can be expressed as:

$$SAMPWT_i = \left[\left(\frac{S_{LL}}{F_{LL}} \times \frac{1}{AD_i} \times LL_i\right) + \left(\frac{S_{CP}}{F_{CP}} \times CP_i\right) - \left(\frac{S_{LL}}{F_{LL}} \times \frac{1}{AD_i} \times LL_i \times \frac{S_{CP}}{F_{CP}} \times CP_i\right)\right]^{-1}$$

S_{LL} = the size of the landline sample

F_{LL} = the size of the landline sample frame

S_{CP} = the size of the cell sample

F_{CP} = the size of the cell sample frame

AD_i= Number of adults in household i, capped at 3

LL_i=1 if respondent i has a landline phone, otherwise LL=0

CP_i=1 if respondent i has a cell phone, otherwise CP=0

Prepaid Cell Oversample Adjustment

The second stage of the weighting is a prepaid cell adjustment to account for the oversampling of prepaid numbers from the cell phone frame. Cell sample respondents were assigned an adjustment so that the proportion of prepaid cell numbers in the entire cell sample equals the proportion of prepaid cell numbers in the base cell sample.

The prepaid cell adjustment for the i^{th} case can be expressed as:

[3] Buskirk, T., Best, J. (2012) Venn Diagrams, Probability 101, and Sampling Weights Computed for Dual-Frame Telephone RDD Designs. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 3696-3710. Retrieved from https://ww2.amstat.org/meetings/jsm/2012/onlineprogram/AbstractDetails.cfm?abstractid=304351

$$PPA_i = \begin{cases} P_{PPD,base}/P_{PPD,total}, & i \in prepaid\ cell\ number \\ (1 - P_{PPD,base})/(1 - P_{PPD,total}), & i \in not\ prepaid\ cell\ number \\ 1, & i \in landline\ number \end{cases}$$

$P_{PPD,base}$ = is the proportion of prepaid cell numbers in the base cell sample

$P_{PPD,total}$ = is the proportion of prepaid cell numbers in the entire cell sample

The final base weight is the product of the sampling weight and the prepaid cell adjustment.

$$\mathrm{BW} = \mathrm{SAMPWT} \times PPA$$

Raking

With the base weight applied, the data were weighted to balance the demographic profile of the sample to the target population benchmarks.

Missing data in the raking variables were imputed using hot decking. Hot deck imputation replaces the missing values of a respondent randomly with another similar respondent without missing data. Hot decking was done using an SPSS macro detailed in 'Goodbye, Listwise Deletion: Presenting Hot Deck Imputation as an Easy and Effective Tool for Handing Missing Data' (Myers, 2011).

Weighting was accomplished using SPSSINC RAKE, an SPSS extension module that simultaneously balances the distributions of all variables using the GENLOG procedure.[4]

Data were weighted to distributions of: age, education, race/ethnicity, census region, population density, phone use. The main demographic benchmarks were obtained from the 2021 Current Population Survey (CPS)[5]. Population density was derived from Census Planning Database 2020[6]. Phone use was projected from the July to December 2021 National Health Interview Survey[7]. Weights were trimmed at the 2nd and 98th percentiles to prevent individual interviews from having too much influence on survey-derived estimates. The table below compared unweighted and weighted sample distributions to target population benchmarks.

[4] https://community.ibm.com/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=17fd2f0b-7555-6ccd-c00c-5388b082161b&forceDialog=0

[5] Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren and Michael Westberry. Integrated Public Use Microdata Series, Current Population Survey: Version 9.0 [dataset]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D030.V9.0.

[6] https://www.census.gov/topics/research/guidance/planning-databases/2020.html

[7] https://www.cdc.gov/nchs/nhis/erwirelesssubs.htm.

1 Braxton Way Suite 125 Glen Mills, PA 19342
www.ssrs.com

Table 1. Sample Demographics

Category	Values	Parameter	Unweighted	Weighted
Gender	Male	48.5%	55.8%	47.0%
	Female	51.5%	41.6%	49.7%
	Other/DK/Ref		2.6%	3.3%
Age	18-29	20.3%	14.6%	20.1%
	30-49	33.2%	24.9%	32.6%
	50-64	24.5%	26.3%	24.3%
	65+	22.0%	33.4%	22.3%
	DK/Ref		0.8%	0.8%
Education	HS Grad or less	37.9%	24.2%	36.9%
	Some college	27.1%	27.3%	27.2%
	College+	35.0%	47.7%	35.3%
	DK/Ref		0.8%	0.7%
Race/ethnicity	White, Non-Hisp	62.5%	66.0%	62.1%
	Black, Non-Hisp	12.0%	9.2%	11.5%
	Hispanic, US Born	8.4%	8.4%	8.3%
	Hispanic, Foreign Born	8.5%	4.8%	7.3%
	Other, Non-Hisp	8.6%	9.1%	8.1%
	DK/Ref		2.5%	2.5%
Region	Northeast	17.2%	17.9%	17.4%
	Midwest	20.6%	22.0%	20.7%
	South	38.3%	36.7%	38.0%
	West	23.9%	23.5%	23.9%
Population Density	1 - Lowest	20.0%	21.8%	19.7%
	2	20.0%	19.9%	19.2%
	3	20.0%	20.6%	19.6%
	4	20.0%	18.1%	19.0%
	5 - Highest	20.0%	16.4%	18.5%
	DK/Ref		3.2%	4.0%
Phone Status	CPO	71.1%	59.9%	70.6%
	Dual Frame	26.4%	37.6%	26.6%
	LLO	2.5%	2.5%	2.7%

Effects of Sample Design on Statistical Inference

Post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. SSRS calculates the effects of these design features so that an appropriate adjustment can be incorporated into tests of statistical significance when using these data. The so-called "design effect" or deff represents the loss in statistical efficiency that results from a disproportionate sample design and systematic non-response. The total sample design effect for this survey is 1.48.

SSRS calculates the composite design effect for a sample of size n, with each case having a weight, w, as:[8]

$$deff = \frac{n \sum w^2}{(\sum w)^2}$$

The survey's margin of error is the largest 95% confidence interval for any estimated proportion based on the total sample — the one around 50%. For example, the margin of error for the entire sample is ± 3.6 percentage points. This means that in 95 out of every 100 samples drawn using the same methodology, estimated proportions based on the entire sample will be no more than 3.6 percentage points away from their true values in the population. Margins of error for subgroups will be larger. It is important to remember that sampling fluctuations are only one possible source of error in a survey estimate. Other sources, such as respondent selection bias, questionnaire wording, and reporting inaccuracy, may contribute additional error of greater or lesser magnitude.

[8] Kish, L. (1992). Weighting for Unequal Pi. Journal of Official Statistics, Vol. 8, No.2, 1992, pp. 183-200.