Empire State Poll 2020

Report 1: Introduction & Methodology

Prepared by Lubna Sharab
Survey Research Institute

Introduction

This report outlines the methodology used for the annual Empire State Poll (ESP), a general survey of New York State residents who are at least 18 years of age that is conducted by the Survey Research Institute (SRI). Since it was founded in 1996, SRI has grown into a premier survey research facility and now comprises more than 40 staff and 22 Computer-Assisted Telephone Interviewing (CATI) stations.

The ESP 2020 marks the eighteenth annual poll in an ongoing survey series that will probe residents' views on a range of workplace, social, political, and economic issues. The data derived from this yearly poll are of particular interest to academics, government officials, business and labor leaders, and journalists. The data also help guide policy making, raise issues for civic dialogue, and suggest avenues of future research.

The ESP 2020 contains two sections: the "core" survey of topics that appear annually and a series of questions that are developed by third parties ("omnibus modules").

ESP Core Instrument

The core survey instrument is a mix of questions about community, economic and government issues that reflect the specific research focus of Cornell faculty and the more general needs of policy makers. The core topics are broken down into the following categories:

- Most important issues facing NYS
- Economic perceptions
- Trust in institutions
- Political behavior
- Demographics

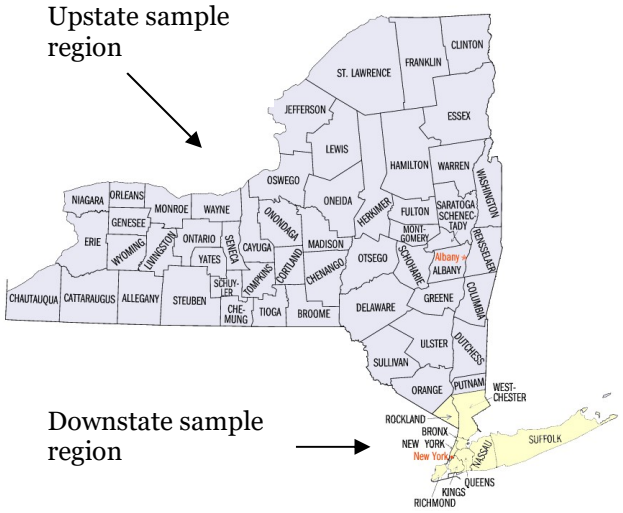
ESP Omnibus Modules

ESP is structured to allow for the inclusion of questions developed by Cornell University faculty and other researchers who are interested in surveying New York State residents on special topics. SRI charges a nominal fee for this service. The data collected from these "omnibus" modules are not reported in the general statewide poll results. For more information on the omnibus section of the poll, please contact SRI directly.

Sampling Methodology

The ESP 2020 survey sample consists of a dual-frame random digit dial (RDD sample covering both cellular and landline exchanges for New York State. Pew Research Center estimates that 11% of New Yorkers use out-of-state cell phone numbers. In an attempt to avoid non-coverage, this year, in 2020, an additional sampling frame was added to cover cell phones of people that live in New York State but have an out-of-state cell phone number. Approximately 11% of all survey completes came from this out-of-state cell phone sample. Additionally, an oversample was used to ensure representation of Black and Hispanic people throughout the state. Marketing Systems Group, a widely used full-service sampling company that provides samples to survey research organizations, supplied the sample, which excluded known business telephone numbers, disconnected numbers, and non-household numbers for landline records. The proportion of cell phone numbers generated for the sample was determined by a count of cell-only households in each county.

The state was divided into two regions, Upstate and Downstate, with sampling in proportion to each region's population totals. "Downstate" was defined as New York, Rockland, Kings, Richmond, Westchester, Suffolk, Queens, Nassau, and Bronx counties, with the remaining counties of the state defined as "Upstate."



Selection of individual respondents came in two steps: for landlines, a household was randomly selected and then a household member who was at least 18 years old was randomly selected from within the household using the "most recent birthday" selection method.[1] Cell phones were treated as personal devices so all adult respondents were deemed eligible.[2] An additional eligibility requirement was that all respondents had to be residents of New York State. These selection procedures ensured that every household with a telephone had an equal chance of being included in the survey; and once a household was selected, each adult in the household had an equal chance of being chosen.

[1] *O'Rourke, D., Blair, J., "Improving Random Respondent Selection in Telephone Surveys," Journal of Marketing Research, Vol. XX (November 1983), 428-32.*

[2] *AAPOR Standard Definitions*
Final Dispositions of Case Codes and Outcome Rates for Surveys
http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

ESP Data Collection

Telephone data collection began on January 24, 2020 and ended on March 15, 2020. Data collection for the ESP is concentrated in the same period every year, starting in January or February and ending in March or April. All interviews are conducted using a Computer-Assisted Telephone Interviewing (CATI) software system, with an average interview length of 18 minutes. A total of 800 interviews were completed – 400 Upstate and 400 Downstate. The survey was administered in both English and Spanish. The questions were translated from English to Spanish then reviewed by a native Spanish speaker prior to the survey administration.

SRI survey interviewers are employed throughout the year. All interviewers undergo rigorous training by the SRI supervisory team and then complete four weeks of probationary interviewing and follow-up training. For each survey project, including ESP, interviewers are also given a survey-specific training session.

All interviewing staff are monitored and supervised at all times by an SRI supervisory team. SRI employs a computer-based proxy system that allows for audio and video monitoring of all interviewer stations. All supervisors regularly monitor interviewers to maintain data collection quality, provide immediate feedback, and troubleshoot issues as they arise.

SRI uses CASES (Computer-Assisted Survey Execution System) for CATI software. CASES is developed, distributed and supported by the Computer-Assisted Survey Methods (CSM) Program at the University of California, Berkeley and was commissioned by the U.S. Census Bureau. For more than 20 years, CASES has been one of the most widely used interviewing systems by survey centers in the United States. SRI employs programmers to support the CATI software and to ensure data collection quality.

For ESP 2020, 800 respondents completed the survey out of a sample list of 18,007. Overall, the cooperation rate, often referred to as the response rate, was 25.6% and the American Association of Public Opinion Research definition of response rate was 14.6%. The cooperation and response rates differed between the Upstate and Downstate samples. The Downstate sample was drawn from a sample list of 9,191 with a cooperation rate of 21.6% and a response rate of 13.1%. The Upstate sample was drawn from a sample list of 8,816 and had a cooperation rate of 31.3% and a response rate of 16.4%.

These cooperation and response rates are consistent with those obtained by other research organizations such as the Pew Research Organization or CBS News, who regularly survey residents in New York State. Additional data are detailed in Table 1.

Table 1 Final Sample Status for ESP 2020

Status	Downstate	Upstate	**Total**
Completed Survey	400	400	**800**
Refusal	1439	868	**2307**
Non-contact	478	710	**1188**
Partial	14	8	**22**
Incapable	5	8	**13**
Language Problem	92	8	**100**
Unknown Eligibility	2161	1602	**3763**
Not a NYS Resident	300	192	**492**
Nonworking Number	4094	4731	**8825**
Non-Residence	174	260	**434**
Age Ineligible (<18)	34	19	**53**
Ineligible– Region quota	0	10	**10**
Total Sample Used	9191	8816	**18007**
Response Rate[2]	13.1%	16.4%	**14.6%**
Cooperation Rate[3]	21.6%	31.3%	**25.6%**

[3] *American Association for Public Opinion Research (AAPOR) response rate and cooperation rate calculations. The response rate is the total number of survey completions divided by the total eligible sample (total sample minus all ineligible, non-households, and estimated proportion of households where eligibility was not determined). Cooperation rate is the total number of survey completions divided by the number of potential interviews (this includes all instances where contact was made with a properly selected person, but not including those instances where the respondent was incapable of cooperating due to language or physical limitations).*

Sampling Error

The sampling error for ESP 2020 assumes the traditional 95% confidence level, which is equivalent to a "significance level" of .05. This means that for questions with approximately 800 respondents there is no more than a one in twenty chance that variations in the respondent sample will cause the ESP 2020 results to deviate by more than 3.5 percentage points when respondents are asked yes/no questions and an even distribution of responses is assumed (i.e., 50% say "yes" and 50% say "no"). Furthermore, the sampling frame was split between Upstate and Downstate residents, allowing comparisons between the overall state and these geographic regions with a one in twenty chance of sampling error greater than 4.9 percentage points for samples of approximately 400.

Sampling error is determined by the assumed distribution of responses and by the size of the sample. An extreme distribution of question responses has a smaller error range. If the distribution of responses were 80/20, for example, the sampling error would be 2.8% for the total sample of 800. See Table 2 for additional distributions and sampling error calculations. The size of the sample or subpopulation is also important because the margin of sampling error increases as the sample size decreases.

The margin of error from responses of demographically distinct subgroups within ESP 2020 will vary depending on the size of the group in question. Again, Table 2 provides some standard sampling errors for groups of different size.

Beginning this year in 2020, the data have been weighted using an iterative "raking" process, as provided by the ipfweight command in STATA 16.1. The weight variable uses population proportions for Upstate/Downstate New York regions, gender, age, Black population, Hispanic population, education, household income, and party affiliation, based on the 2018 American Community Survey 5-year estimates, and New York State Board of Elections data.

Table 2 Sampling Error Margins by Question Response Distribution and Sample Size[4]

		Sample Size (N)				
		800	600	400	200	100
Question Response Distribution (%)	50/50	***3.5***	***4.0***	***4.9***	***6.9***	***9.8***
	60/40	***3.4***	***3.9***	***4.8***	***6.8***	***9.6***
	70/30	***3.2***	***3.7***	***4.5***	***6.4***	***9.0***
	80/20	***2.8***	***3.2***	***3.9***	***5.5***	***7.8***
	90/10	***2.1***	***2.4***	***2.9***	***4.2***	***5.9***

[4] *Calculations made through the Survey System sample size calculator. http://www.surveysystem.com/sscalc.htm*

Lastly, besides the possible sample error mentioned above, all public opinion polls may incur other sources of error associated with telephone data collection procedures, including the sampling error from the systematic exclusion of households without telephones, question wording, question order, and interviewer-induced bias

Respondent Demographics

The accuracy of ESP 2020 can be evaluated by comparing selected characteristics of the survey respondents to data from the U.S. Census. A weight variable was developed based on geography (Upstate versus Downstate) in order to approximate actual population distribution within New York State. All substantive results described within any ESP 2020 report are weighted using this variable.

Table 3 compares the weighted distribution of ESP 2020 respondents' characteristics with the actual statewide distributions reported by the U.S. Census in the 2012-2016 American Community Survey 5-Year Estimates.

Table 3 Key Respondent Demographics for ESP 2020 (% reported)

Characteristics	Downstate (N=400)		Upstate (N=400)		Pooled[5] (N=800)		U.S. Census/ ACS[6]
Age	(raw)	(weighted)	(raw)	(weighted)	(raw)	(weighted)	
18-24	11	12	8	5	10	9	13
25-34	18	21	14	20	16	21	18
35-44	18	19	13	14	15	17	16
45-54	18	19	16	16	17	18	18
55-64	17	14	18	18	17	16	16
65 and older	18	15	31	28	24	20	19
Gender							
Male	51	46	53	53	52	49	48
Female	49	54	47	47	48	51	52
Race							
White	65	63	88	89	77	73	64
Non-White	35	37	12	11	23	27	36
Ethnicity							
Hispanic (any race)	25	26	8	7	16	19	19
Non-Hispanic	75	74	92	93	84	81	81
Employment Status							
Employed	69	65	61	53	65	60	59
Unemployed	13	16	10	11	12	14	8
Not in labor force	18	19	29	36	23	25	33
Annual Household Income							
Less than $10,000	3	7	2	7	2	7	8
$10,000-49,999	26	31	28	36	27	33	35
$50,000 -99,999	31	28	34	25	33	27	28
$100,000 or more	40	34	36	32	38	33	29
Education							
Less than Bachelor's Degree	45	58	55	74	50	64	65
Bachelor's Degree or higher	55	42	45	26	50	36	35

[5] *Weighting applied to match actual distribution of Upstate vs. Downstate. Due to rounding, distributions may not add up to 100. Reported percentages exclude non-responses.*
[6] *From the 2012-2016 American Community Survey 5-Year Estimates.* *http://factfinder2.census.gov*

For More Information:
Stephanie Slate, Administrative Manager
Survey Research Institute at Cornell University
Ithaca, New York 14850
Email: srs287@cornell.edu, Web: www.sri.cornell.edu
Phone: 607-255-3786, Fax: 607-255-7118