METHODOLOGICAL NOTE

Understanding Likely Voters

By Jocelyn Kiley and Michael Dimock, Pew Research Center for the People & the Press
January 22, 2009

Each election season, the Pew Research Center for the People & the Press – along with many other pollsters – reports not only on the electoral preferences of all registered voters, but also on the preferences of a sub-group of registered voters identified as "likely" to vote in the upcoming election. Likely voters are identified by their answers to questions about their intention to vote, their past voting behavior, their knowledge about the voting process, and their interest in the campaign. While it is impossible to predict with certainty who will and who will not actually cast a ballot, the process of narrowing the sample to those most likely to vote is an essential step in accurately gauging the balance of voter opinion in an election.

The impact of likely voter estimates is not consistent throughout the election cycle. In particular, likely voter estimates tend to have a more substantial partisan impact – benefiting the Republican candidate – on election weekend than they do earlier in the campaign cycle. Because of the unreliability of early likely voter measures, Pew Research does not produce likely voter estimates earlier than September of the election year, and treats the findings with some caution until the final pre-election survey just days before the election.

That said, the results of the Pew Research Center's final pre-election survey in 2008 reinforced the importance of likely voter modeling in the final election survey. Results based on all registered voters overstated Barack Obama's support substantially, and the likely voter estimate proved to be the more accurate gauge of the status of the election.

This report summarizes the differences in presidential horserace estimates based on likely and registered voters in 2008 and previous election cycles. It then details the components of the 2008 likely voter scale and the impact of small improvements made to the scale this year. Finally, it assesses the impact of both the inclusion of cell phone samples in pre-election polls and the unprecedented amount of early voting on likely voter estimates.

Why Likely Voters Instead of Registered Voters?

While the vast majority of Americans say that they are registered to vote, and virtually all registered voters indicate that they plan to vote in any given election, not all of them will actually do so. According to the best estimates available, between 52% and 62% of eligible voters have cast ballots in presidential elections since 1980.[1] Narrowing a sample of registered voters to "likely voters" allows pollsters to better estimate the preferences of those who ultimately turn out to vote in an election.[2] In three of the last four elections (1996, 2004 and 2008), the Pew Research Center's likely voter estimates of voters' presidential preferences on election weekend provided substantially more accurate estimates of the actual margin than those based solely on registered voters' preferences.

In general, measures of election preferences based on all registered voters tend to be more favorable to the Democratic candidate, while measures based on likely voters are more favorable to the Republican candidate. In 15 of the 19 fall election surveys conducted by the Pew Research Center from 1996 to 2008, likely voters were at least slightly more supportive of Republican candidates (and, it follows, slightly less supportive of Democratic candidates) when compared with registered voters. And in the last five presidential elections, reliance on registered voter preferences from the final survey before Election Day would have significantly overstated support for the Democratic candidate.

[1] http://elections.gmu.edu/voter_turnout.htm

[2] The purpose of reporting on likely voters is not to perfectly determine individual voting behavior (that is, not all "likely" voters will vote, while at least some "unlikely" voters will cast ballots). See Screening Likely Voters: A Survey Experiment, May 18, 2001 for more details on the validation of likely voter measures with actual voting behavior and election estimates.

The Presidential Horserace and Likely Voters

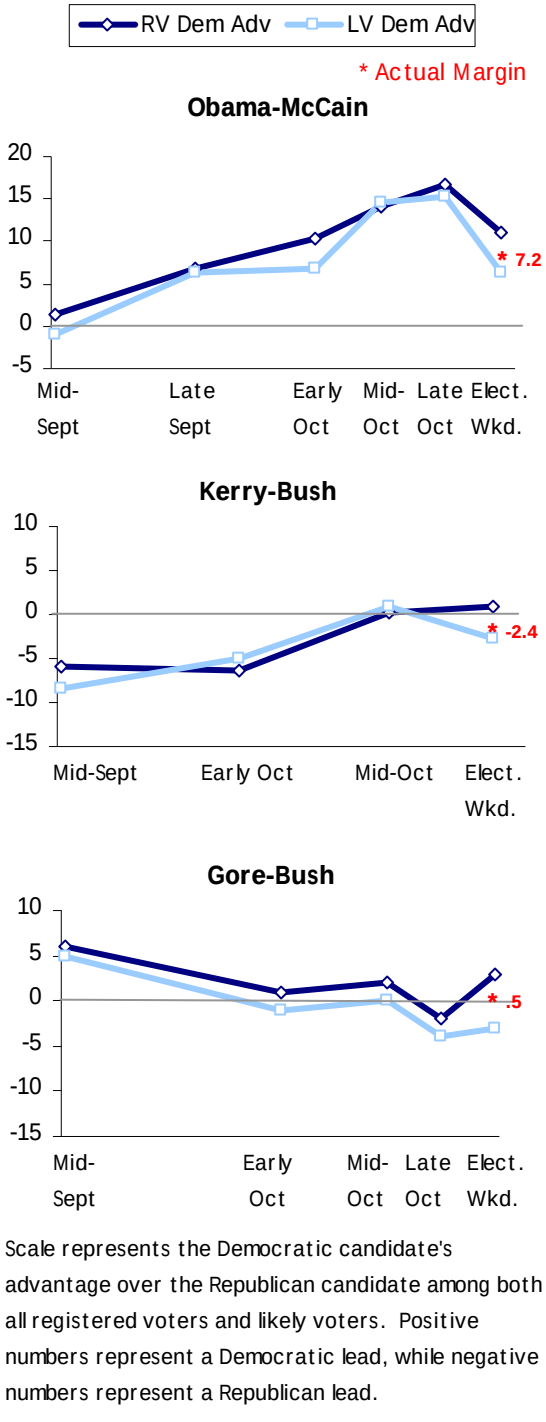
While likely voter estimates are essential to the final election forecasts, their impact is often muted in early election surveys, only to become fully visible in the final survey conducted immediately before Election Day. In the last four election cycles the difference between registered voters and likely voters has consistently been the largest in the final survey conducted the weekend before Election Day.

This pattern was once again evident in the Pew Research Center's 2008 pre-election polling. In the final election-weekend 2008 survey, Barack Obama held an 11-point lead over John McCain among all registered voters (50%-39%). But when the sample was limited to likely voters, his lead was narrowed to seven-points (49%-42%). This four point pro-McCain shift in the margin was the largest difference between the registered and likely voter estimates of the six 2008 pre-election surveys conducted by Pew Research from September through November. But the RV-LV difference was of similar magnitude to that found in prior election weekend surveys.

One week earlier, Obama held a 16-point lead over McCain among registered voters (52%-36%) and a 15-point lead among likely voters (53%-38%). Thus, the overall narrowing of Obama's likely voter lead from 15-points the previous week to seven-points on election weekend reflected both a tightening of the race among registered voters (from a 16-point lead to an 11-point lead) and the greater impact of the likely voter scale (from a one-point to a four-point shift toward McCain).

Size of Horserace Margin

among Registered and Likely Voters



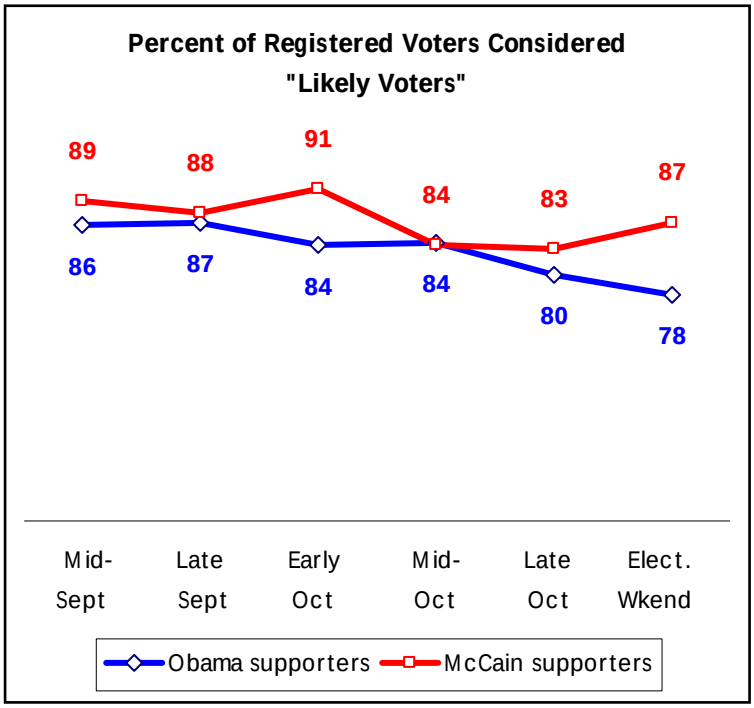
Scale represents the Democratic candidate's advantage over the Republican candidate among both all registered voters and likely voters. Positive numbers represent a Democratic lead, while negative numbers represent a Republican lead.

That the likely voter screen would benefit the Republican candidate more in the final election weekend survey than earlier in the year is typical of previous election cycles. In both 2000 and 2004 the likely voter screen had a far greater impact on estimates in the final survey than at any point earlier in the cycles. However, because neither of these elections also saw a pro-Republican tightening of the race

among registered voters, the week-to-week shift in the margin among likely voters was less pronounced. In fact, in both the 2000 and 2004 campaigns, the final poll showed a slight Democratic bounce among registered voters; as a result, while the gap between likely voters and registered voters was more pronounced on election weekend, the preferences among likely voters were little changed from the week before.

Although the magnitude of this shift in the impact of the likely voter scale on election weekend in each election is relatively small, the consistent pattern of a larger Republican advantage among likely voters on election weekend suggests that the likely voter scale may be operating somewhat differently in the immediate days preceding the election than it does earlier in the cycle.

Yet no single item or set of items in the likely voter scale is clearly responsible for the shift. In 2008, McCain supporters scored slightly better on some individual questions in the scale on election weekend when compared to Obama voters (see appendix for distributions of the individual items from mid-September through election weekend by each candidate's supporters). However, for most items, the gap between McCain and Obama supporters is consistent with the gaps observed in earlier surveys. But the net effect of the scale on election weekend in 2008 (and in previous years) was a set of likely voters substantially more supportive of the Republican candidate and less supportive of the Democratic candidate than had been the case in earlier surveys. For example, while equal proportions of McCain and Obama supporters passed the likely voter screen in mid-October (84% each), McCain supporters were significantly more likely to be considered likely voters on election weekend (87% compared to just 78% of Obama supporters).



In particular, African Americans – a group overwhelmingly in support of Obama – were much less likely to make it through the likely voter screen on election weekend than they had been in previous surveys; just 62% of African American respondents on election weekend were likely voters – far short of the proportion passing the screen on the late October survey (74%) or the two surveys earlier in October (69% in both). As a result, African Americans made up only 11% of likely voters on election weekend, but had been 13% of likely voters a week earlier.

Components of the Pew Research Center's Likely Voter Scale

The Pew Research Center for the People and the Press's likely voter scale relies on a set of questions first developed by the Gallup organization more than five decades ago. These questions measure prior voting behavior, intention to vote, knowledge about where to vote and campaign interest and enthusiasm. While the scale may vary slightly in individual years (for instance, the 2008 scale included an additional indicator of intention to vote to slightly weight the scale more towards interest and engagement), the core questions comprising the scale have remained the same, and are summarized below. The complete set of questions used to determine likely voters in 2008 can be seen here.

Components of the Likely Voter Scale

Question		Response Categories
OFTVOTE	How often do you vote	Always/Nearly always
PVOTE04a	Voted in previous Presidential Election	Yes
PRECINCT	Ever voted in current precinct	Yes
THOUGHT	Thought given to election	A lot/Some
Q1	How closely follow election news	Very/Fairly closely
Q2	How much general interest in politics	A great deal
PLAN3	How certain are you you will vote	Absolutely certain
SCALE10	Likelihood of voting (10-pt scale)	9 or 10
WHERE	Know where to vote	Yes

Respondents get one point for each question.

Respondents are automatically coded a zero (0) on the scale if:

They are not considered registered voters (already registered, or able to same-day register/vote) OR
They say they do not plan to vote

Respondents are automatically coded a 9 (definite likely voter) if:

They say (in PLAN1) that they have already voted

Respondents under age 30 are provided additional points (or partial points) to compensate for past voting behaviors and lower levels of knowledge about the voting process as follows:

18-19 year olds +1.5 points
20-24 year olds +1.0 points
25-29 year olds +0.5 points

African American respondents (30 and older), receive an additional half-point (+0.5) to compensate for underepresentation among likely voters in previous election-year surveys (compared with Current Population Survey reports of voting).

Those considered likely voters represent a proportion of respondents scoring highest on the likely voter scale. The proportion of Americans estimated to be likely to vote is determined based on estimates of voter turnout, as detailed below. In both late-October and on election weekend 2008, the full nine-point likely voter scale was used, while – as in previous years – surveys early in the fall used a somewhat shorter set of questions in order to save time on the surveys. (In the mid-September and mid-October

surveys likely voters were determined using a seven-item scale, while in late September and early October a five-item scale was used).[3]

The shift from shorter versions of the scale to a longer version has little impact on likely voter predictions. For instance, calculating likely voters based upon either of the truncated scales on election weekend would not have significantly shifted the margin of Obama's lead (Obama's margin over McCain with a nine-point scale was 6.4 points, compared with a 6.1 point lead using the seven-point scale and a 5.8 point lead with a five-point scale).

Impact of Different Versions of Likely Voter Scales

	Obama	McCain	*Obama Adv*
	%	%	
All registered voters	50.4	39.4	+11.0
Likely voters, as determined using a…			
9-item scale	48.7	42.3	+6.4
7-item scale	48.5	42.3	+6.1
5-item scale	48.1	42.3	+5.8

Election weekend, 2008

Younger respondents who have not had the opportunity to vote in the past are at a severe disadvantage in scales that include measures of past voting behavior such as PVOTE, OFTVOTE and PRECINCT. To adjust for this, likely voter scales typically do not count these questions against younger voters and/or provide extra "points" to them to compensate for fewer opportunities to vote in the past. For example, Pew Research's likely voter indexes in 2000 and 2004 excluded the question about voting in the prior presidential election (PVOTE) from the scale for all respondents under age 24 and provided them with two additional points to compensate for this question and their disadvantage on other questions about past behavior.

The youth adjustment was modified slightly in 2008 to correct for known problems. In particular, the adjustment made in 2000 and 2004 which gives every voter up to age 23 two extra points had created a sharp cutoff that had the effect of disadvantaging slightly older voters – especially 24 to 29 year olds – disproportionately. Comparisons between the Pew Research Center's election weekend surveys and the measures of self-reported voting from the Current Population Survey (CPS) Voter Supplements in 2000 and 2004 suggested that 24-29 year olds had been underrepresented among the Pew Research Center's likely voters. In 2008, the adjustment was made more gradual and extended as follows: 1.5 points for 18 and 19 year olds, 1 point for 20 to 25 year olds, and ½ point for 26 to 29 year olds. In addition, since the bonusing was extended up to age 29, respondents were given credit for their answers to all measures of prior voting behavior, rather than excluding the PVOTE question. With all questions included, the maximum "bonus" was reduced from 2 points to 1.5 points.

Comparisons to the 2000 and 2004 Current Population Surveys also indicated an underestimation of African American turnout in previous years. As a result of that finding, an additional ½ point was given to African Americans respondents ages 30 and older.

[3] The seven-item and five-item scales did not include WHERE or Q.2. Surveys using the 5-item scale also did not include Q.1 or SCALE10.

With early projections that youth voting and African American voting could spike disproportionately in 2008 we were particularly conscious of improving the accuracy with which Pew Research's likely voter models represent these groups. But the changes to the scales were relatively 'conservative' in that they attempted to correct for observed underrepresentation of these groups in previous election cycles.[4] The adjustments made no effort to increase the share of younger or African-American voters beyond what was the case in 2004. For example, the maximum bonus for the youngest people was actually reduced slightly from 2004 in order to achieve a more consistent estimate across all age ranges. Estimates of the actual levels of youth and black turnout in 2008 remain somewhat unfixed until the release of the 2008 CPS post-election data. At that time the Pew Research Center will once again benchmark the composition of our likely voters with CPS findings

Estimating Turnout

The likely voter index assigns each respondent a score based on their answers to the various survey questions, but determining what share of respondents should be considered to be likely voters is another step in the process. The easiest solution would be to set a minimum value on the scale – such as scoring seven points or higher on a nine-point scale – as a cutoff. However, this approach poses two problems. First, if the number of items in the scale varies from survey to survey, the cutoff would have to be adjusted, and might result in different proportions of respondents "passing" the scale from survey to survey. Second, many items related to interest and engagement in the election, as well as self-reported registration, increase as Election Day draws nearer. This means that in each survey a higher share of respondents score at or near the top of the likely voter scale. As a result, a fixed cutoff set early in the election cycle would result in an overly inclusive definition of likely voters by election weekend (or an overly exclusive definition of likely voters earlier in the campaign).

To avoid these problems, Pew Research determines a fixed share of respondents that will be considered likely voters ex-ante, and this proportion is taken off the top of each survey's likely voter index. The estimate is determined based on several sources including: estimates based on voter interest and engagement measures from the Pew Research Center's June voter attitudes survey conducted in the summer of each presidential election year, external sources (in particular, the United States Elections Project) and a regression model that includes many of the likely voter questions as variables to predict turnout in modern presidential elections. Once a turnout estimate is derived, it is applied consistently across all surveys in that election year. In 2008, analyses and early projections estimated a 62% turnout

[4] Based on the comparisons with the 2004 CPS data, the adjustments made in Pew Research's mid-September 2008 survey – the first with a likely voter scale –were more generous than what is described here (3 points for 18 and 19 year olds, 2 points for 20-25 year olds, 1 point for 26-30 year olds, and 1 point for African Americans ages 30 and over). Beginning in late September the adjustment was reined back for the remainder of the campaign. We chose to reduce the bonusing due to the use of shorter likely voter scales in a number of surveys, and also due to substantially increased registration rates and scores on other likely voter questions by young people and African Americans that were resulting in disproportionately high turnout estimates for these groups relative to 2004 under the mid-September bonusing system. Results for mid-September in this report are computed using the revised adjustment and differ slightly from what was originally reported. (Likely voters in mid-September were originally reported as 46% Obama, 46% McCain. This change results in a 45% Obama, 46% McCain margin based on likely voters.)

rate, compared with an estimate of 57% turnout in 2004 and 50% in 2000).[5] In this respect, the predicted growth in voter turnout from 2004 to 2008 was larger than what actually came to pass.[6]

New Developments in 2008: Rising Use of Cell Phones and Much More Early Voting

The growing number of cell phone users (and the use of cell phone sampling) and the record breaking numbers of early voters this year both raised questions about the potential impact on the calculation of likely voters. Although both posed methodological challenges, analysis suggests that neither had a significant impact on projections of likely voters.

While the inclusion of cell phone samples in the 2008 election surveys may have shifted the overall horserace numbers (for more on this, see Calling Cell Phones In '08 Pre-Election Polls), it does not appear to have had an impact on the differences between registered voters and likely voters in individual polls. Landline samples (respondents reached on landline telephones) would have produced a nearly identical pattern in the observed gap between Obama and McCain over the course of the election (e.g., on election weekend, a landline sample would have shown Obama's lead over McCain decrease by four-points among likely voters compared to registered voters; a week earlier, his lead would have decreased by a single point).

Beginning in mid-October, respondents who reported that they had "already voted" when interviewed were taken at their word and automatically considered likely voters. As Election Day approached, these early voters represented a larger and larger share of likely voters. On election weekend, more than a quarter of registered voters passed the likely voter screen based not on their responses to the questions in the likely voter scale, but because they reported having already cast their ballots (put differently, this meant that only 44% of those who had **not already cast their ballots** could be considered likely voters). However, the record breaking amount of early voting in the 2008 election did not significantly impact the horserace estimates.

An alternative scale, computed without automatically considering those who had already voted as likely voters, does result in differences at the individual level (that is, some of the individuals considered "likely" to vote differ in each version of the scale), but both versions of the likely voter scale would have resulted in identical likely voter projections for both election weekend (49%-42%) and the late October survey (53%-38%).

---

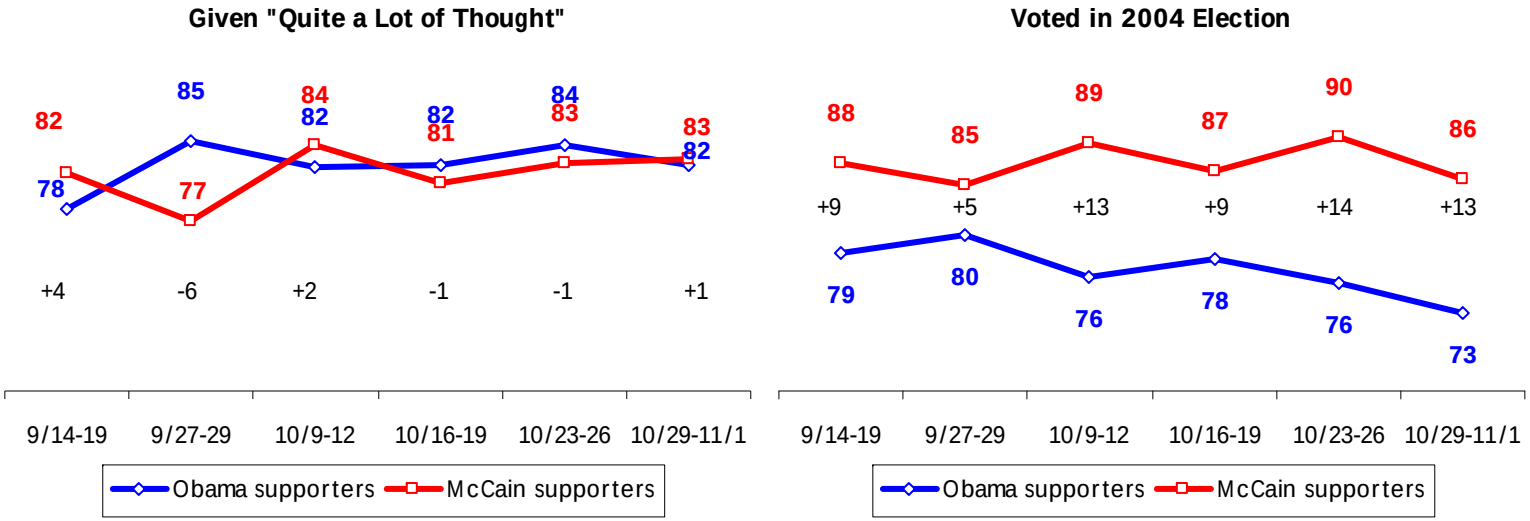
[5] In applying these estimates to the surveys themselves, we include a slightly larger percentage of respondents as likely voters in order to account for the known response bias in political surveys toward more highly politically engaged respondents. In 2008 when we projected 62% turnout we counted the top 67% of respondents as likely voters, and have made a similar 5% adjustment in previous years.

[6] None of these estimates align directly with either VAP (voting age population) or VEP (vote eligible population) measures of actual voter turnout because the sampling base of our surveys is not exactly comparable to either of these metrics. According to Professor Michael McDonald at George Mason University, actual VEP turnout rose from 60% in 2004 to 62% in 2008. http://elections.gmu.edu/voter_turnout.htm

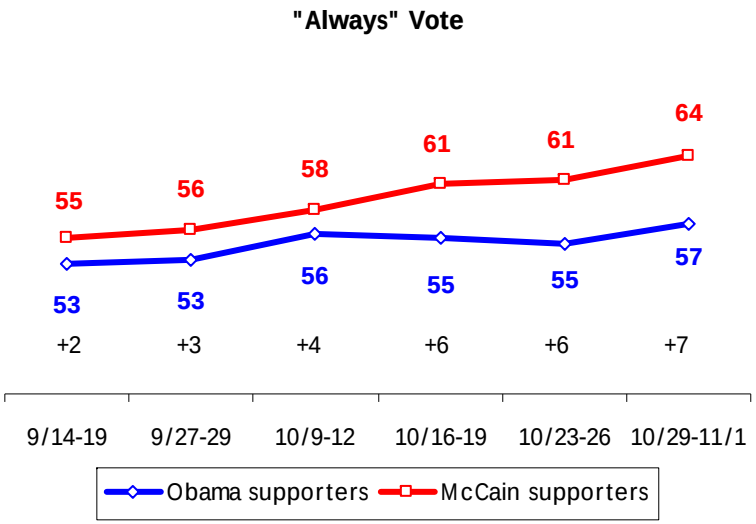
Demographic Profile of Registered and Likely Voters

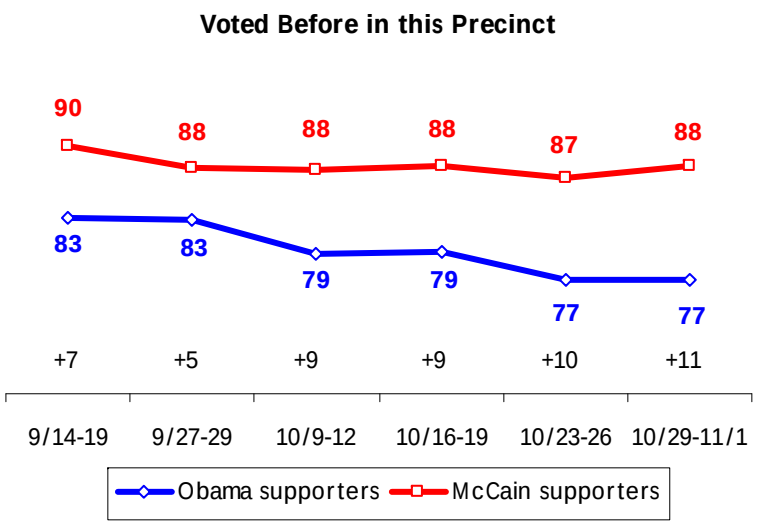
	Mid-Sept %	Late Sept %	Early Oct %	Mid-Oct %	Late Oct %	Election Weekend %
Registered Voters						
Men	48	46	46	47	47	48
Women	52	54	54	53	53	52
	100	100	100	100	100	100
18-29	16	16	16	19	18	19
30-49	37	38	38	36	36	35
50-64	27	28	26	27	27	26
65+	19	17	18	17	17	17
White	78	79	75	78	76	76
Black	11	12	13	12	12	12
Republican	31	29	30	29	25	29
Democrat	36	37	37	38	40	38
Independent	28	30	29	28	30	28
Likely voters						
Men	48	45	45	46	46	48
Women	52	55	55	54	54	52
	100	100	100	100	100	100
18-29	14	14	15	17	16	15
30-49	36	36	36	35	35	36
50-64	29	30	28	28	29	29
65+	20	18	19	19	18	18
White	79	79	77	79	76	78
Black	11	12	13	12	13	11
Republican	33	30	33	30	28	31
Democrat	37	38	39	39	40	37
Independent	26	28	26	27	28	26

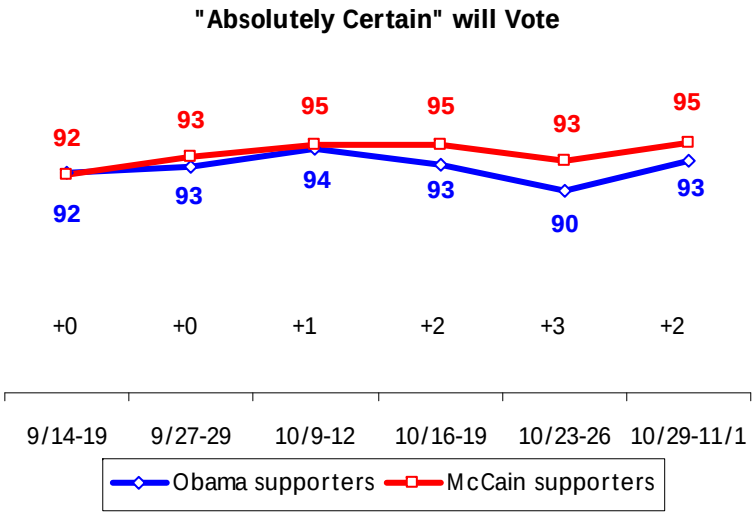
Components of the Likely Voter Scale, by Each Candidate's Supporters

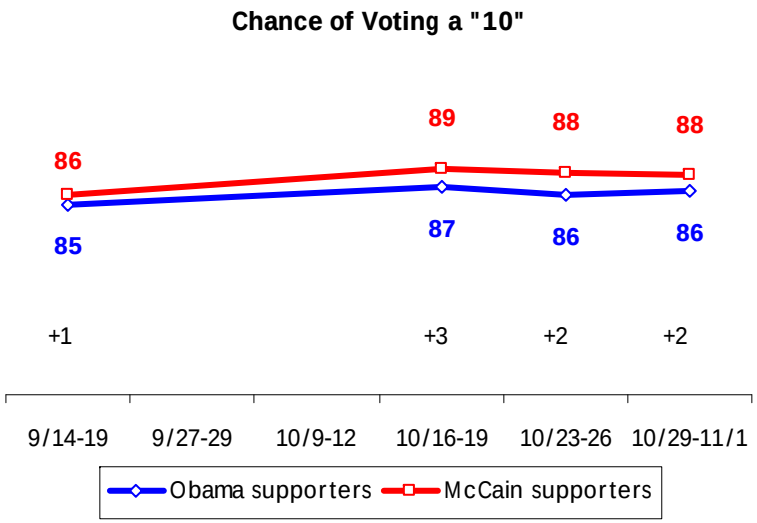


Appendix: Understanding Likely Voters

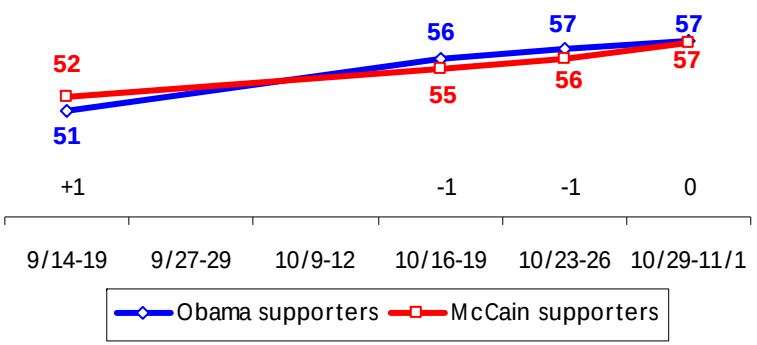




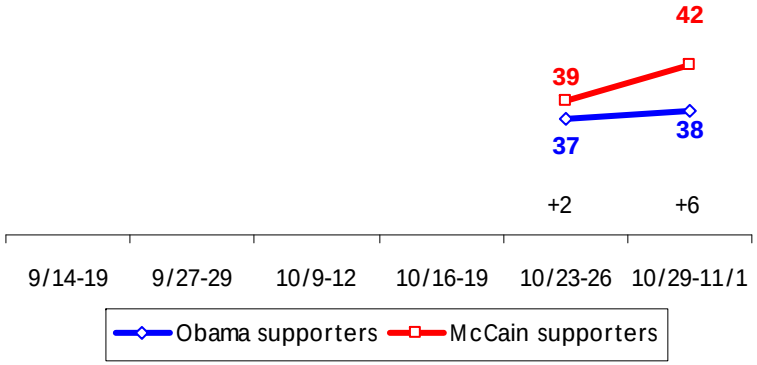




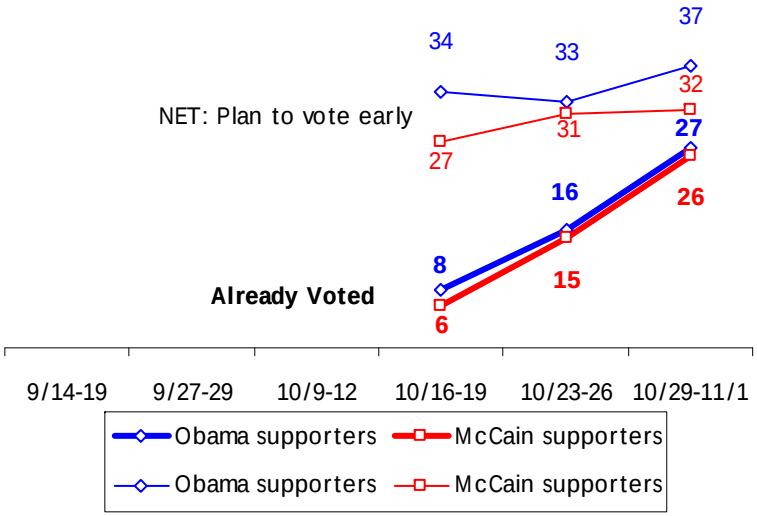
Following Campaign News "Very Closely"



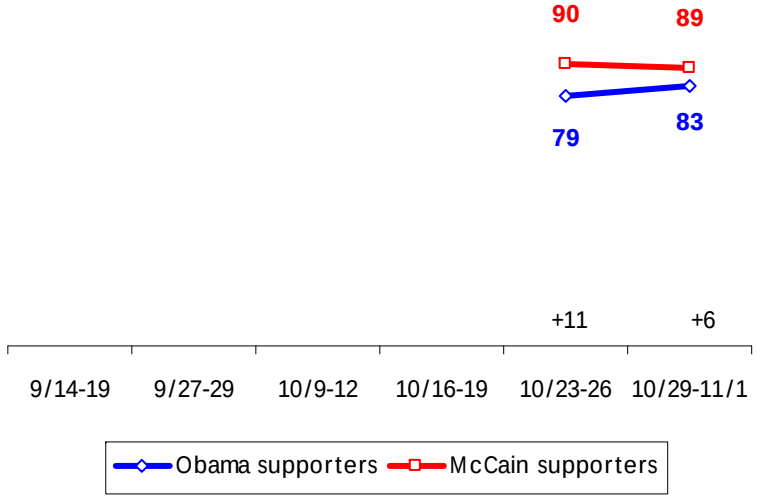
"A Great Deal" of Interest in Politics

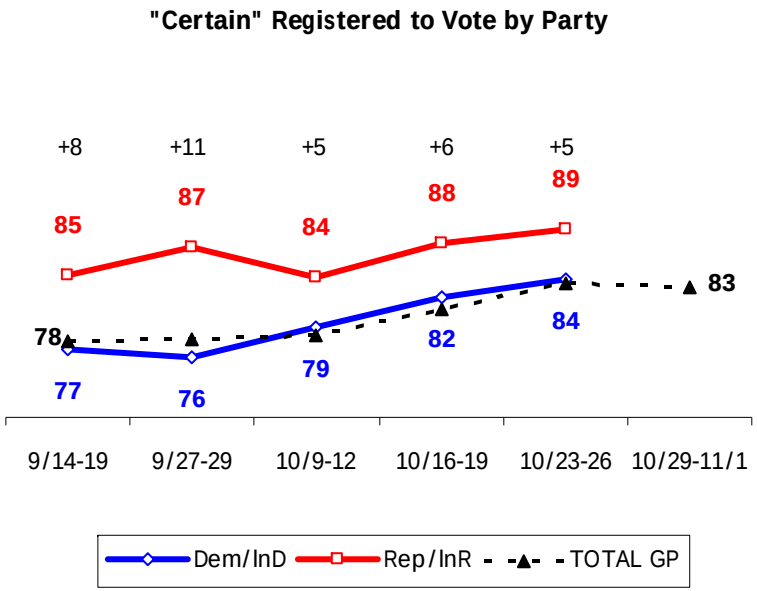


Early Voting



Know Where People Go (Or Voting Early)




"Certain" Registered to Vote by Party
+8
+11
+5
+6
+5
85
87
84
88
89
78
77
76
79
82
84
83
9/14-19
9/27-29
10/9-12
10/16-19
10/23-26
10/29-11/1
Dem/InD
Rep/InR
TOTAL GP