**Analysis of Overvotes in the 2016 California Senate Primary Election**

David C. Kimball, University of Missouri-St. Louis

Martha Kropf, University of North Carolina-Charlotte

June 2017

The 2016 Senate primary contest in California was the first prominent statewide office to be filled under the state’s recently adopted “top-two” voting rules without an incumbent running. Under the top-two rules all candidates, regardless of political party, appear together on the primary ballot and the top two vote-getters advance to the general election. After U.S. Senator Barbara Boxer announced her retirement in 2015, the subsequent election in 2016 to fill her seat attracted 34 candidates. This is the largest number of candidates to appear on the ballot for the same office in California since the 2003 gubernatorial recall election. Fitting 34 candidates for one office on the primary election ballot posed a serious ballot design challenge for election officials in each of the state’s 58 counties. Different counties took different approaches in placing the Senate candidates on the 2016 primary ballot. This report examines ballot design and voting equipment used in California’s 2016 primary election and its potential impact on voting errors.

There is evidence that design features created confusion for some voters in the 2016 California Senate primary election. We examine the number of overvotes in the Senate primary contest as a measure of voting errors.[[1]](#footnote-1) An overvote is when a voter casts a vote for more than the allowable number of candidates in a contest. In the California primary voters were allowed to vote for just one candidate per contest so overvotes occurred when voters mistakenly voted for multiple candidates in a race. Overvotes can be a serious problem because the voter’s choices are disqualified in any race where he casts an overvote. In the 2016 California primary election the Senate race shared the ballot with other contests, including the hard-fought presidential primaries for the Democratic and Republican parties. There were approximately 8.5 million voters in the 2016 California primary election. The Republican and Democratic presidential primaries produced a combined 5,833 overvotes, less than one-tenth of one percent of ballots cast.[[2]](#footnote-2) This overvote rate is fairly typical for American elections.[[3]](#footnote-3) In contrast, there were at least 235,821 overvotes in the California Senate primary, over 2.8% of ballots case and more than 40 times as many overvotes as in the presidential contests. At the county level, the overvote rate in the Senate primary ranges from a low of 0% to a high of 14.8% (Lake County). The number of overvotes was substantially smaller than the vote margin between the second place and third place finishers in the Senate primary. Thus, overvotes did not affect the outcome of the top-two Senate primary in California in 2016. Nevertheless, when nearly one quarter of a million voters make a detectable error in a single contest it merits further investigation about the sources of those voting errors.

We identify one ballot design feature and one voting equipment feature that contributed in roughly equal measure to the large number of overvotes in the 2016 California Senate primary election. The crucial ballot design feature that contributes to overvotes is when candidates for a single office are listed in multiple columns on the ballot. When candidates for the same office are listed in one column voters see that they are grouped together for the same voting task. However, when candidates for the same office are listed in multiple columns voters tend to get confused and mistakenly assume that a new column of candidates indicates a new voting task. An infamous example of the multi-column format is the “butterfly ballot” used in Palm Beach County, Florida in the 2000 presidential election. Prior studies show that listing candidates for the same office in more than one column generates notably higher rates of overvotes and spoiled ballots.[[4]](#footnote-4) In the 2016 California primary 33 of the state’s 58 counties listed the Senate candidates in more than one column.[[5]](#footnote-5) The remaining counties listed the 34 Senate candidates in a single column on one ballot page. The overvote rate in the Senate primary was 3.6% of ballots cast in counties using the multiple column format. In counties that listed the Senate candidates in one column on a single page the overvote rate was 0.8%.

The critical voting equipment feature involves whether ballots are tabulated at polling places or at the central county election office. Some counties use vote-by-mail or optical scan ballots in which all ballots are counted at a central location. Other counties use electronic voting machines or optical scan ballots that are tabulated at polling places. One advantage of the polling place tabulators is that those machines are designed to prevent overvotes or alert voters when there is an overvote on their ballot. This error-correction mechanism helps reduce overvotes and other types of voter confusion.[[6]](#footnote-6) In the 2016 California Senate primary election 24 counties counted ballots at a central location.[[7]](#footnote-7) The overvote rate in the Senate primary was 4.1% of ballots cast in counties using a central tabulation system. In counties that tabulated ballots at polling places the overvote rate was 1.1%.

The multi-column ballot format and the central counting system tend to work together to produce relatively high rates of overvotes in an election. If the multi-column ballot format is a high wire inviting the danger of overvotes then the error correction mechanism inherent in polling place tabulation acts as a safety net to help voters catch and correct overvotes. However, in a county using central count tabulation (no safety net) the multi-column ballot format is likely to produce a dramatic increase in overvotes. As it happens, the two election features appear to be independently distributed among California counties in the 2016 primary election. In the 14 counties that list the Senate candidates in a single column and tabulate ballots at polling places the overvote rate is a relatively low 0.7% of total ballots cast (see Figure 1). In the 10 counties listing the Senate candidates in a single column while tabulating ballots at a central location the overvote rate is still fairly low at 0.9%. In the 16 counties that place candidates in multiple columns but also use the safety net of polling place tabulation the overvote rate in the Senate contest creeps higher to 1.4%. Finally, in the 13 counties with the multi-column ballot format and central ballot tabulation the overvote rate jumps to 4.9% of ballots cast. Nine of the ten counties with the highest overvote rates in the 2016 Senate primary (all above 2%) employed this combination of listing candidates in multiple columns while counting ballots at a central location.

**Figure 1**

**Overvote Rates in 2016 California Senate Primary by Ballot Format and Counting System**

****

Other factors likely influence the frequency of overvotes in elections with many candidates running for the same office. For example, counties probably vary in terms of their efforts to educate voters before the election about the voting process, and in how thoroughly they train poll workers to assist voters who may face difficulty. We are unable to measure these election administration practices at the county level but we can compare the 2016 primary to a previous election that is somewhat comparable. As noted above, the 2003 gubernatorial recall election was the previous statewide election in California with an unusually large number of candidates running for the same office (135 candidates). We don’t have overvote data from the 2003 recall election but we do have data on the residual vote rate, the difference between the number of ballots cast and the number of valid votes cast for candidates in the replacement contest. In the 2003 recall election the county residual vote rate ranged from 1.8% to 12.2%. We find that a county’s residual vote rate in the 2003 gubernatorial recall election is a strong predictor of overvotes in the 2016 Senate primary, even after controlling for the multi-column ballot format and a county’s tabulation system. Each one percentage point increase in the 2003 residual vote rate is associated with a half percentage point increase in the 2016 overvote rate (see Table 1 below). In addition to column format and the tabulation system there are other features of county ballot design, voting equipment, and election administration that seem to consistently influence the ability of voters to properly cast their ballots.

We did examine a few other ballot design features that appear to be unrelated to the frequency of overvotes cast in the 2016 California Senate primary election. One involves the manner in which voters mark the ballot. Most counties ask voters to darken an oval or square next to the name of their preferred candidate. However, in 17 counties voters draw a line to connect an arrow pointing to their chosen candidate. There is some evidence indicating that voters tend to be more confused by the connect-the-arrow format than by other ballot marking styles.[[8]](#footnote-8) However, in the 2016 California Senate primary election the overvote rate is actually higher in counties that ask voters to fill in a circle or square (3.2%) than in counties that have voters connect an arrow (1.8%). Once we account for a county’s prior voting behavior and the election features described above then connect-the-arrow format is unrelated to the overvote rate in the Senate primary (see Table 1 below).

In a contest with many candidates it is important for voters to know that they can only vote for one candidate. In the Senate primary each county included text saying “vote for one” or “vote for no more than one candidate” at the beginning of the Senate portion of the ballot. Most counties placed that language at the top center or top left of the Senate section on the ballot. However, 23 counties right-justified the “vote for one” language, a location where some voters might miss that instruction. In any case, we did not find higher overvote rates in the Senate contest in counties that placed the “vote for one” instruction on the right side of the ballot.

Finally, voting by mail (VBM) is an important development in California when considering overvotes. A small number of counties have shifted to holding elections entirely by mail, while the remaining counties let voters choose whether to vote at a precinct or by mail. In the 2016 California primary election 59% of ballots were cast by mail. Since all VBM ballots are centrally counted, VBM does not include an error detection feature. When we account for the percentage of centrally counted ballots in each county (rather than the use of central versus precinct tabulation) we find very similar results as what is reported in Table 1. The highest overvote rates in the Senate primary are concentrated in counties that (1) used a multi-column ballot format and (2) counted all ballots centrally.

These findings are relevant for the recently passed California Voter’s Choice Act. Under the terms of the Act all voters in participating counties will receive and ballot via mail 28 days before Election Day and will have the choice of returning the ballot by mail, returning the ballot to a drop-off location, or voting in person at a voting center. Fourteen counties are eligible to implement the law in the 2018 elections and the remaining counties can adopt the new voting procedures in 2020. Whether voting centers offer the error correction feature of precinct tabulation may be an important consideration if there are more contests with an unusually large number of candidates in future elections.

**Table 1**

**Predictors of Overvote Rates in 2016 California Senate Primary**

|  |  |
| --- | --- |
| **Independent Variable** | **OLS Coefficient**  **(standard error)** |
| Multi-Column Format | 1.40\*  (0.77) |
| Central Ballot Tabulation | 0.46  (0.88) |
| Multi-Column X Central Count | 2.25\*  (1.05) |
| Residual Vote Rate in 2003 Recall | 0.51\*  (0.09) |
| Connect-the-Arrow Format | -0.33  (0.56) |
| Constant | -3.54\*  (0.99) |
| Number of cases  Adjusted R2  Root MSE | 53  .70  1.65 |

Note: The dependent variable is the frequency of overvotes in the U.S. Senate contest (as a percentage of total ballots cast). Observations are weighted by the number of ballots cast in each county. Cell entries are OLS coefficients (standard errors in parentheses).

\*p<.05 (one-tailed)

1. Data on overvotes comes from the following report: Davit Avagyan and Philip Muller, “CA June 2016 Over-Vote Analysis,” Your Voter Guide, August 17, 2016 (<https://www.scribd.com/document/322175665/California-Over-Voting-Report>). Overvote data are not available from Mendocino, Napa, Plumas, Trinity and Tulare counties. Thus, we have overvote data from 53 of the 58 counties in California. [↑](#footnote-ref-1)
2. The California Republican presidential primary featured 5 candidates and the Democratic presidential primary featured 7 candidates. [↑](#footnote-ref-2)
3. Martha Kropf and David C. Kimball, *Helping America Vote: The Limits of Election Reform* (New York: Routledge, 2012), p. 86. [↑](#footnote-ref-3)
4. As far as we can tell, the problem with multi-column ballot layouts was first reported by Aubrey Jewett, “Explaining Variation in Ballot Invalidation among Florida Counties in the 2000 Election” (paper presented at the annual meeting of the American Political Science Association, San Francisco, August 2001. Also, see Jonathan N. Wand et al., “The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida,” *American Political Science Review* 95(2001):793-810; Alan Agresti and Brett Presnell, “Misvotes, Undervotes, and Overvotes: The 2000 Presidential Election in Florida,” *Statistical Science* 17:436-440; Kropf and Kimball, *Helping America Vote*, chapter 5. [↑](#footnote-ref-4)
5. We thank the Center for Civic Design for assistance in collecting and coding ballots from the primary election. We examined sample ballots from each county in California to code the multiple column format and other ballot features described in this report. Three counties listed Senate candidates in a single column but on two separate ballot pages. We code these three counties as using a multi-column format under the principle that some voters may mistakenly assume that the column on the second page indicates a new voting task. [↑](#footnote-ref-5)
6. See Kropf and Kimball, *Helping America Vote*, chapter 3. [↑](#footnote-ref-6)
7. We gathered information on counties in California using central tabulation from a report by the Secretary of State: “Voting Systems Used by Counties: June 7, 2016 Presidential Primary Election” (<http://elections.cdn.sos.ca.gov/sov/2016-primary/24-vot-sys-by-counties.pdf>). [↑](#footnote-ref-7)
8. Charles S. Bullock, III and M. V. Hood, III, “One Person – No Vote; One Vote; Two Votes: Voting Methods, Ballot Types, and Undervote Frequency in the 2000 Presidential Election, *Social Science Quarterly* 83:981-993; Kropf and Kimball, *Helping America Vote*, pp. 83-87. [↑](#footnote-ref-8)