Public Opinion Quarterly, Vol. 69, No. 4, Winter 2005, pp. 508–529

BALLOT DESIGN AND UNRECORDED VOTES ON PAPER-BASED BALLOTS

DAVID C. KIMBALL
MARTHA KROPF

Abstract The 2000 presidential election focused attention on the problem of unrecorded votes, in which a person casts a ballot but fails to record a valid vote for a particular contest. Although much recent research has evaluated voting technologies and their effects on unrecorded votes, there has been little research on the effects of ballot design. We argue that the same theories used to design and evaluate self-administered surveys can be used to analyze ballot features. We collect and code paper-based ballots used in the 2002 general election from 250 counties in 5 states. We code the ballots in terms of several graphic design elements, including the content and location of ballot instructions and the layout of candidate names and office titles. Our analysis suggests that several ballot features are associated with unrecorded votes (both overvotes and undervotes) in the gubernatorial contests. We also find that ballot design features exacerbate the racial disparity in unrecorded votes. Ballot design can be an important factor in determining whether voters are able to cast a ballot accurately, which can influence the legitimacy of elections.

The 2000 presidential election and the Florida recount controversy illuminated the phenomenon of unrecorded votes (in which some voters come to polling places but fail to cast a valid vote for a particular contest). Roughly 2 million voters (almost 1 out of every 50 to cast a ballot) failed to cast a valid vote for president in the 2000 election (Caltech/MIT Voting Technology Project 2001). The Florida imbroglio prompted a new wave of research on election administration and a flurry of election reform laws in Congress and

DAVID C. KIMBALL is an associate professor of political science at the University of Missouri–St. Louis. MARTHA KROPF is an associate professor of political science at the University of Missouri–Kansas City. The authors give special thanks to Amanda Baker, Lindsay Battles, and Susan Mason for research assistance. We also thank Alexandra Cooper, Jay Dow, Peter Miller, John Petrocik, the anonymous reviewers, and faculty and students in the Department of Political Science at the University of Missouri–Columbia for comments on earlier drafts. This research was supported by the University of Missouri Research Board. Address correspondence to David Kimball; e-mail: dkimball@umsl.edu.

doi:10.1093/poq/nfi054

state legislatures, reflecting a concern that real and perceived obstacles to voting may undermine the legitimacy of elections in the United States.

A large part of the election reform effort has been devoted to replacing outdated voting equipment, particularly the punch card ballots that were so problematic during the Florida recount. Hundreds of counties have replaced older voting technologies with optical scan methods or electronic voting machines since the 2000 election. Similarly, much recent research has evaluated voting technologies and their effects on unrecorded votes.

In contrast, aside from studies of the "butterfly ballot" used in Palm Beach County, Florida, in 2000 (Dillman 2002; Kimball, Owens, and Keeney 2004; Sinclair et al. 2000; Wand et al. 2001), there has been scant research on the effects of ballot design on unrecorded votes. As a result, researchers and election officials know little about whether ballot features (such as the location and readability of voting instructions, the graphic layout of offices and candidate names) might confuse voters. However, some election officials have begun consulting with graphic design experts on ballot layout (Kamin 2004; Omandam 2002), and scholars have begun noting potentially confusing ballot features (Dillman 2002; Niemi and Herrnson 2003).

This article examines several ballot features and their impact on unrecorded votes using theories and concepts from several disciplines. We rely most heavily on studies of questionnaire design, especially those that address item nonresponse. We collected and coded paper-based ballots (hand-counted paper and optical scan) used in 250 counties in 5 states with gubernatorial contests during the 2002 general election. We find that several ballot features are associated with higher rates of unrecorded votes. We also find an interaction between the racial composition of counties and ballot design features; the elevated rate of unrecorded votes associated with confusing ballots tends to fall disproportionately on counties with high concentrations of African-American voters.

Previous Literature Examining Unrecorded Votes

Unrecorded votes are defined commonly as the difference between total turnout and the number of valid votes cast in a certain contest. They occur as the result of undervotes (intentional or unintentional failures to record a vote) or overvotes (where voters select too many candidates, thus invalidating the vote). Theories to explain the incidence of unrecorded votes can be separated into at least three perspectives. First, scholars argue that some unrecorded votes are intentional for reasons such as voter fatigue, lack of a desirable choice, or low information about a contest (e.g., Bullock and Dunn 1996; Kimball, Owens, and Keeney 2004; Wattenberg et al. 2000).

Second, some researchers argue that accidental undervotes and overvotes occur due to faulty equipment or confusing ballot design. For example, Votomatic punch card ballots tend to produce more unrecorded votes than other voting methods (Alvarez, Sinclair, and Wilson 2003; Ansolabehere and Stewart 2005; Knack and Kropf 2003a). Furthermore, error prevention and correction mechanisms (e.g., precinct counters for optical scan ballots) tend to reduce unrecorded votes (Bullock and Hood 2002; Kimball, Owens, and Keeney 2004; Knack and Kropf 2003a; Nichols and Strizek 1995; Tomz and Van Houweling 2003). From studying ballots, it appears that the practice of listing candidates for the same office in multiple columns or on multiple pages produces more unrecorded votes (Darcy and Schneider 1989; Herron and Sekhon 2003; Jewett 2001; Kimball, Owens, and Keeney 2004; Sinclair et al. 2000). Finally, a straight-party option (where voters can make a single mark to cast a vote for all candidates from one party) tends to reduce the frequency of unrecorded votes (Kimball, Owens, and Keeney 2004).

A third research perspective focuses on equal protection issues, analyzing the relationship between unrecorded votes and demographic factors. For example, several studies find more unrecorded votes in precincts and counties with large populations of minority, low-income, less-educated, or elderly voters (e.g., Herron and Sekhon 2003; Knack and Kropf 2003a; Sinclair and Alvarez 2004; Tomz and Van Houweling 2003). Furthermore, the association between socioeconomic measures and unrecorded votes is weaker in places that use either equipment (e.g., error correction mechanisms) or ballot features (e.g., a straight-party option) that facilitate completion of a valid ballot (Kimball, Owens, and Keeney 2004; Knack and Kropf 2003a; Tomz and Van Houweling 2003). By the same token, the elevated rate of unrecorded votes associated with confusing ballots and voting technology tends to fall disproportionately on precincts and counties with high concentrations of poor, elderly, or minority voters (e.g., Darcy and Schneider 1989; Herron and Sekhon 2003; Kimball, Owens, and Keeney 2004; Knack and Kropf 2003a; Nichols 1998).

Overall, scholars have focused more on the impact of voting technology than ballot design. Recent exceptions include Dillman's (2002) critique of the "butterfly ballot" from Palm Beach County and a study by Niemi and Herrnson (2003) that identifies several ballot features in different states that may be confusing. We argue that many ballot features remain unexamined. Moreover, to our knowledge, little published research compares ballots that are actually used in different places in order to see which ballot features, if any, correlate with high levels of unrecorded votes.

Methodology

We examine counties using paper-based ballots in five states from the 2002 elections. Election administration is decentralized in the United States, typically delegated to counties or townships. This produces variation in voting methods and ballots even within the same states. Thus, the unit of analysis for our study is the county.[1] We collected ballots from the November 2002 election from counties in five states: Florida, Illinois, Iowa, Kansas, and Tennessee.[2] These states were chosen because they use a variety of voting methods and ballots, and they featured a competitive gubernatorial race. We mailed surveys to each county or city election official in these states to ask for information, including a copy of the ballot, the type of voting equipment used, the number of ballots cast in the county, the number of votes cast for each candidate, and the number of overvotes and undervotes cast in the county. We coded the paper-based ballots in terms of several graphic design elements, as described in a later section.

We focus on paper-based ballots for several reasons. First, paper-based ballots most resemble the paper-and-pencil questionnaires covered in the survey research literature on nonresponse. Other voting methods, including electronic machines, have a very different user interface than written questionnaires and paper-based ballots. Second, paper-based ballots are the most commonly used voting method in the United States.[3] Third, the current debate about the security of electronic voting machines has prompted many counties to purchase optical scan equipment instead. Finally, even counties that use electronic voting machines on Election Day use optical scan ballots for absentee voting, which is rapidly growing.

There are 476 counties in the 5 states chosen for this study. However, we were only interested in the 261 counties from these states that used paper-based ballots in the 2002 general election. We obtained voting data and ballots for 250 of the 261 target counties.[4] Roughly 4.3 million ballots were cast during the November 2002 election in the 250 counties in our sample. Twenty counties in our sample used hand-counted paper ballots, 127 counties used centrally counted optical scan ballots, and 103 counties used precinct-counted optical scan ballots.[5]

To measure the frequency of unrecorded votes for governor in each county, we calculated the difference between the total number of ballots cast and the

1. Some Illinois election districts within counties (Aurora, Bloomington, Chicago, Danville, East St. Louis, Galesburg, Peoria, and Rockford) use different equipment and ballots, so they are considered as separate cases.
2. We asked election officials to submit to us one ballot that was most representative of the county or was from the largest precinct. One might argue that using data from only five states limits the generalizability of this study. Based on 2000 census figures, our sample is quite similar to the rest of the country in terms of the percentage of African-American residents (13.4 percent in our sample versus 12.6 percent in the rest of the country), the percentage of Hispanic residents (10.4 percent versus 12.9 percent), the percentage of citizens over the age of 65 (14.4 percent versus 12.0 percent), the percentage of adults with a high school degree (80.8 percent versus 80.4 percent), and median household income ($37,126 versus $39,699).
3. Nationwide, 39 percent of ballots in the 2002 general election were cast on optical scan ballots and 1 percent were cast on hand-counted paper ballots.
4. There are missing data for three counties from Illinois, seven counties from Kansas, and one county from Tennessee.
5. We discovered some counties that have precinct-count optical scan equipment but do not use the error correction feature. These counties were placed in the central-count category.

number of votes cast for governor and use it as our dependent variable in the analyses that follow. The distribution of unrecorded votes across counties is skewed, with outliers at the high end. In our sample of 250 counties, there were 42,059 unrecorded votes for governor. Unrecorded vote percentages for gubernatorial contests ranged from 0.2 percent to 6.9 percent, with a mean of 1.8 percent. We conducted further analyses on 133 counties in our sample that provided data on overvotes (selecting votes for more than one candidate) and undervotes (recording no votes for any candidate). In this smaller sample of counties, there were 3,303 overvotes for governor and 24,639 undervotes for governor. Overvotes for governor ranged from 0 percent to 1.97 percent of ballots cast, with a mean of 0.17 percent, while undervotes ranged from 0.1 percent to 4.77 percent of ballots cast, with a mean of 1.20 percent.

Information Processing and Ballot Design

To our knowledge, there are very few studies that draw on other disciplines to evaluate ballot design and usability (Callegaro and Peytcheva 2003; Roth 1994). We believe there is a parallel with the survey methodology literature, which borrows theories and concepts from psychology, graphic design, usability research, and optometry to examine ways in which the visual layout and content of a questionnaire influence survey answers, as well as item and survey nonresponse (for example, see Christian and Dillman 2004; Couper et al. 2000; De Vaus 2002; Dillman 2000; Dillman, Sinclair, and Clark 1993; Gower and Dibbs 1989; Jenkins and Dillman 1997; Schuman and Presser 1981; Tourangeau, Couper, and Conrad 2004; Zukerberg and Lee 1997). We argue that unrecorded votes are the ballot equivalent of item nonresponse in a questionnaire. In both cases, voters (or respondents) fail to record a selection from the list of choices on the ballot (or questionnaire). Ballots and self-administered questionnaires have many features in common, especially when voting involves making written marks on a paper ballot.

As in self-administered questionnaires, the voter must process verbal, symbolic, and graphical language when completing a ballot (Christian and Dillman 2004; Dillman 2002; Jenkins and Dillman 1995, 1997). The verbal language on a ballot includes the written instructions, as well as text that lists offices and candidate choices. The graphical and symbolic language on a ballot includes shading, font size, party symbols, the spatial layout of choices, and other design features. The visual information on a printed form should be organized so that all readers follow a consistent and efficient path to process the information on each page.

In reviewing the questionnaire design literature, we identify several criteria to evaluate the degree to which ballots simplify or complicate the voting process. Some ballot features are specific to the instructions, such as their location and readability. Other features are specific to the layout of the candidates, such as where the voter marks the ballot, shading and bolding of candidate names and office titles, and clutter around candidates' names.[6]

BALLOT INSTRUCTIONS

Location of Instructions. When considering the entire ballot, one first confronts the fact that voters usually look first at the upper left-hand corner of the ballot, as is typical in Western culture (Dillman 2000, p. 113; Dillman 2002; Jenkins and Dillman 1997). This is supported by Roth's study of voting (1994, p. 59). In addition, work in survey research indicates that respondents may not always read the directions (Gower and Dibbs 1989). Thus, some conclude that instructions should appear just before the response task to which they apply (Christian and Dillman 2004; Dillman 2000, pp. 98–99; Zukerberg and Lee 1997). The theory here is based on Gestalt psychology's Law of Proximity: placing items close together encourages readers to view them as related (Dillman 2000, p. 107; Wallschlaeger and Busic-Snyder 1992). Finally, when instructions are placed in the top left corner of the ballot, they appear as shorter lines of text, which are easier to read and comprehend than long lines spread over an entire page (Dillman 2000, p. 129; Long et al. 1996, p. 93).
We code the location of voting instructions for each of the ballots in this study. In 56 percent of the ballots in our sample, voting instructions are in the top left corner, just above the first contest (see the example ballot in figure 1). In all but two of the remaining counties, instructions are spread across the entire width of the top of the ballot (see the example ballot in figure 2).[7] This feature varies within each state in our sample, except Illinois, where all counties listed voting instructions across the top of the ballot. For this analysis, we create a dummy variable to identify ballots with instructions in the top left corner. We expect that unrecorded votes are less common in counties where voting instructions are located in the upper left corner of the ballot.

Readability. Questionnaire design research also indicates that instructions should be easy to read. Words and sentences should be short and simple, written in an active, affirmative style (Sanders and McCormick 1993, p. 110; Zukerberg and Lee 1997). In general, readability describes the ease of processing the information content of written words. To measure the readability of each ballot's voting instructions, we type the instructions into Microsoft Word and compute Flesch-Kincaid Grade Level scores, which indicate the grade level needed to

6. One relatively well known ballot principle is that candidates for the same office should be listed in a single column, unlike the "butterfly ballot" used in Palm Beach County, Florida, in the 2000 presidential election (Dillman 2002). None of the ballots in our sample listed candidates for governor in multiple columns, so we do not examine this feature. By and large, election officials have learned from the case of the butterfly ballot.

7. In one of the two remaining counties, no instructions appear on the ballot, and in the other county instructions appear in the middle of the ballot below some of the contests. Examples of other ballots that meet or violate each of the features described in this paper are available online at http://www.umsl.edu/~kimballd/ballots.htm (accessed September 26, 2005).

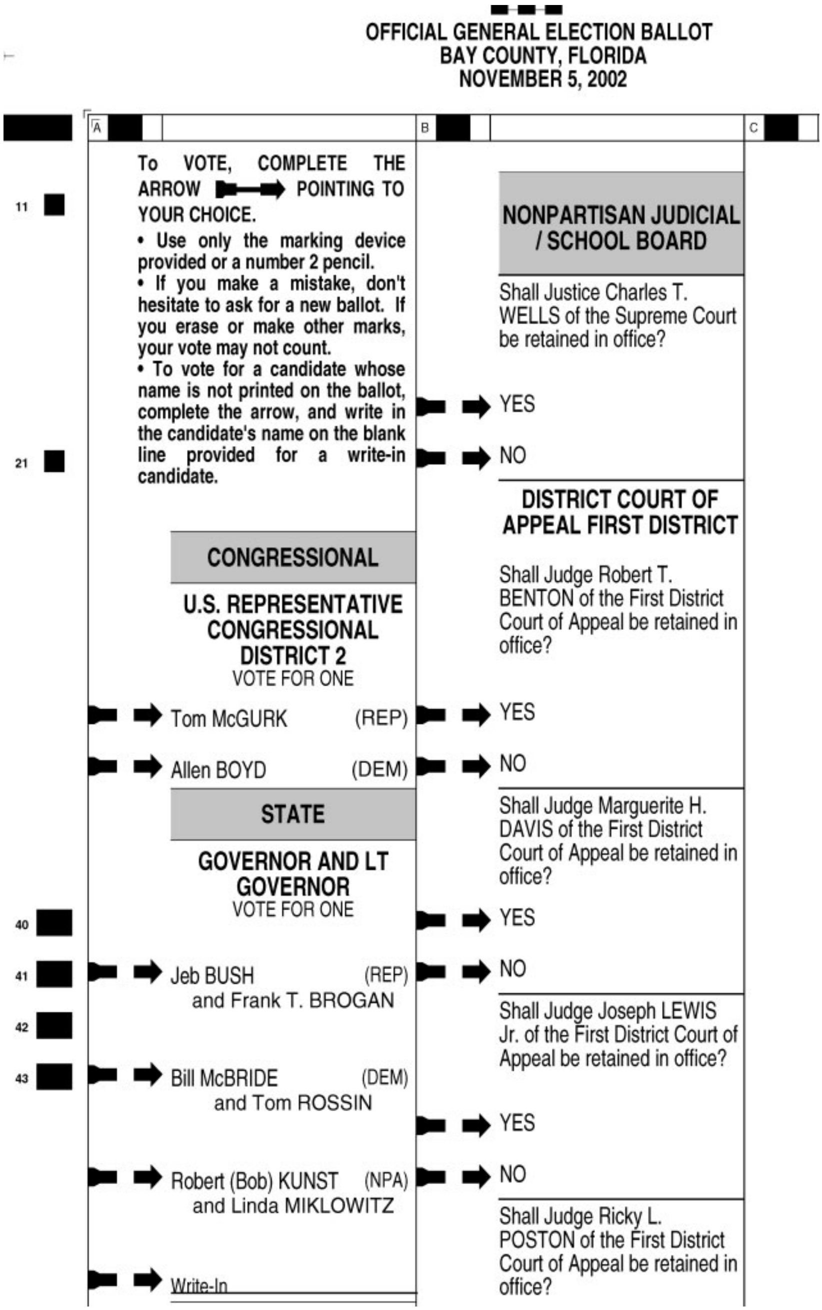

Figure 1. Portion of a ballot with mostly positive features.

NOTICE
If you tear, deface, or make a mistake and wrongfully mark any ballot, you must return it to the election board and receive a new ballot or set of ballots.
To vote for a person whose name is printed on the ballot darken the oval at the left of the person's name. To vote for a person whose name is not printed on the ballot, write the person's name in the blank space, if any is provided, and darken the oval to the left.
TO VOTE, DARKEN THE OVAL NEXT TO YOUR CHOICE, LIKE THIS:

AVISO
Si usted rompe, altera o hace un error y equivocadamente marca cualquier papeleta, usted tiene que retornarla a la junta de elecciones y recibir una nueva papeleta o un grupo de papeletas.
Para votar por una persona(s) a quien su nombre esta impreso en la papeleta obscurezca el ovalo a la izquierda del nombre de la persona. Para votar por una persona(s) a quien su nombre no esta impreso en la papeleta, escriba el nombre de la persona(s) en el espacio en blanco, si hay alguno provisto, y obscurezca el ovalo a la izquierda. **PARA VOTAR, LLENE COMPLETAMENTE EL OVALO JUNTO A SU SELECCION.**

NATIONAL OFFICES
OFICINAS NACIONALES

FOR UNITED STATES SENATOR
PARA SENADOR DE LOS ESTADOS UNIDOS
VOTE FOR ONE / VOTE POR UNO

PAT ROBERTS, Dodge City, Republican
STEVEN A. ROSILE, Wichita, Libertarian
GEORGE COOK, Mission, Reform

FOR UNITED STATES REPRESENTATIVE 1ST DISTRICT
PARA REPRESENTANTE DE LOS ESTADOS UNIDOS 1ST DISTRITO
VOTE FOR ONE / VOTE POR UNO

JACK WARNER, Wright, Libertarian
JERRY MORAN, Hays, Republican

STATE OFFICES
OFICINAS ESTATALES

To vote for persons for governor and lieutenant governor whose names are printed on the ballot darken the oval at the left of the names of the persons running together for such offices.

Para votar por personas para gobernador y lugarteniente del gobernador a quienes sus nombres están impresos en la papeleta obscurezca el óvalo a la izquierda de los nombres de las personas corriendo juntos para esos cargos

FOR GOVERNOR AND LIEUTENANT GOVERNOR
PARA GOBERNADOR Y LUGARTENIENTE DEL GOBERNADOR
VOTE FOR ONE PAIR / VOTE POR UN PAR

TED PETTIBONE, St. Marys, Reform
MIKE WILSON, Salina, Reform

KATHLEEN SEBELIUS, Topeka, Democratic
JOHN MOORE, Wichita, Democratic

TIM SHALLENBURGER, Baxter Springs, Republican
DAVID LINDSTROM, Overland Park, Republican

STATE OFFICES
OFICINAS ESTATALES

FOR SECRETARY OF STATE
PARA SECRETARIO DEL ESTADO
VOTE FOR ONE / VOTE POR UNO

CHARLES ST-GEORGE, Wheaton, Reform
RON THORNBURGH, Topeka, Republican
DAVID HALEY, Kansas City, Democratic

FOR ATTORNEY GENERAL
PARA PROCURADOR GENERAL
VOTE FOR ONE / VOTE POR UNO

CHRIS BIGGS, Junction City, Democratic
PHILL KLINE, Shawnee, Republican

Figure 2. Portion of a ballot with mostly negative features.

understand the text.[8] Higher scores indicate documents that are harder to read. The use and validity of the Flesch-Kincaid scores are supported by other studies (Heilke, Joslyn, and Aguado 2003; Sanders and McCormick 1993; Tefki 1987).

In our sample the grade level scores for ballot instructions range from fourth grade to twelfth grade. The mean reading score for our sample is 8.1, suggesting that the average ballot in our sample required an eighth-grade education to understand the instructions. The instructions on the ballot in figure 1 are written at a seventh-grade level, while the ballot instructions in figure 2 are written at the twelfth-grade level. We expect that unrecorded votes are more common in counties with high reading level scores for voting instructions.

Spoiled Ballot Instructions. We also examine whether ballot instructions include a warning about the consequences of spoiling a ballot, as well as directions for correcting ballot errors. We examine this feature because of our culturally based practice of trying to correct bad marks when we make a mistake (Dillman 2002). On optical scan ballots, this often produces an overvote. The Help America Vote Act of 2002 (HAVA) requires local election officials to implement more vigorous voter education programs. In particular, HAVA requires jurisdictions with paper ballots to include instructions on the effect of multiple votes for a single office and how to correct a spoiled ballot (HAVA 2002, title III, section 301.a.1.B).

Roughly 10 percent of the ballots we examined did not contain any warning or other information about what voters should do if they spoil their ballots. The remaining 90 percent of ballots told voters they could turn in their ballots to an election judge or poll worker and obtain a new ballot if they made a mistake. However, only a subset of those ballots (22 percent of the sample) included a warning that votes will not count if the ballot is spoiled. We create a three-category variable to measure the content of instructions regarding spoiled ballots. It is coded “2” for ballot instructions that warn about the consequences of spoiled ballots *and* how to correct them (as in the ballot in figure 1); “1” for ballot instructions that only state how to correct a spoiled ballot (as in the ballot in figure 2); and “0” for ballot instructions that make no mention of spoiled ballots. We expect that unrecorded votes are less common in counties where ballot instructions discuss the consequences of spoiled ballots and how to correct them.

LAYOUT OF OFFICES AND CANDIDATE NAMES

The Use of Shading and Bolding. When considering the layout of offices and candidate names on a ballot, we hypothesize that the use of shading and boldface fonts can help guide people through the voting process. Kahneman

8. The Flesch-Kincaid Grade Level scores are based on the lengths of words and sentences. The formula for computing the score is $(.39 \times ASL) + (11.8 \times ASW) - 15.59$, where ASL is the average number of words per sentence and ASW is the average number of syllables per word. In Microsoft Word, the “Spelling and Grammar” feature in the “Tools” menu computes Flesch-Kincaid scores for a document.

(1973) argues that people, while reading text, focus on physically informative "high contrast areas," a characteristic that can be used to a questionnaire designer's advantage (Dillman 2000, p. 106; Jenkins and Dillman 1997, p. 11). The appropriate response choices (e.g., the list of candidates) can be identified more quickly if they are differentiated from other questions and instructions on a ballot (Redline and Dillman 2001). Shading certain sections of text can encourage the appropriate grouping of information and guide the respondent from one task to the next on the ballot (Design for Democracy 2002; Dillman 2000; Dillman, Sinclair, and Clark 1993; Omandam 2002). Also, boldface text may be used to highlight the questions or office sections on which a voter is working, while non-bolded print may be used for the candidates, to make the candidates different from the office names and highlight the response task (Dillman 2000, p. 118).

Design features such as shading and bolded text are generally left to local election officials, so there is substantial variation in the use of these ballot features within each state in our sample. Most ballots in our sample (63 percent) do not have any shading (as in figure 2). The remaining ballots employ shading to highlight each office or groups of offices, such as federal offices or state offices (as in figure 1). We create a dummy variable to identify ballots that use shading to draw attention to different offices. We expect fewer unrecorded votes in counties with ballots that use shading in this manner.

Furthermore, most ballots in our sample (61 percent) do not use boldface text or shading to highlight candidates for office or make the candidate names stand out from the office for which they are running (as in figure 2). Again, we create a dummy variable to identify ballots where candidate names are differentiated from office names in terms of boldfaced text or shading (as in figure 1). We expect unrecorded votes to be less common in counties that use shading or bolded text to differentiate candidate names from office titles.

Finding the Correct Box or Oval to Mark. One possible reason for high levels of unrecorded votes may be that voters are marking the wrong box for their chosen candidate. Aside from the simple positioning of the choices, survey methodology research indicates that there should be no ambiguity about which box or circle corresponds to each candidate, in keeping with the Law of Proximity. The "relative closeness" of items to each other can cause respondents to see the items as a related group (Roth 1994, p. 63; see also Jenkins and Dillman 1997). Ballots often divide a page into two or three columns of offices and candidate names, so confusion may arise if spots for marking a vote appear on both sides of a candidate's name. The ballots in our sample are coded as to whether there is a circle, arrow, or box on both sides of the candidate names (as in the contests for Congress and governor in figure 1). Even though most ballots have the response location and the candidate names left-justified, approximately 22 percent of ballots in our sample have marking areas on both sides of the gubernatorial candidates' names, leaving some confusion about where to mark a vote in those races. We expect unrecorded votes to be more common in counties where ballots have spots for marking a vote on both sides of candidate names.

Clutter around the Candidates. Election officials or state legislatures sometimes require that ballots include extra detail (such as a candidate's occupation or hometown). However, this may be a misguided effort to help voters. Survey researchers advise against putting any extraneous text near the response options on a questionnaire (Babbie 1990; Dillman 2000). Niemi and Herrnson (2003) apply that criterion to ballots, arguing that some ballots include too much information, beyond a candidate's name and party affiliation. In addition, listing candidates for governor and lieutenant governor together (in states where they run as a team) can add clutter to the ballot. The names of lieutenant governor candidates should be indented and there should be extra spacing between teams of candidates to avoid crowding the names of the candidates for governor. We attempt to measure clutter by creating a dummy variable to identify ballots that violate these principles. About 64 percent of ballots in our sample have excessive clutter around the candidate names for governor (as in the ballot in figure 2). In some cases, clutter is a result of state law. For example, for statewide contests in Kansas, ballots are required to list a candidate's city of residence (Kansas Statute No. 25-613). We hypothesize that more clutter around candidate names on the ballot will result in higher rates of unrecorded votes.

OVERALL INDEX OF BALLOT FEATURES

Finally, for each county in our sample we compute an overall index of the ballot features described above. Since each ballot feature except for the readability of ballot instructions is measured as a categorical variable, the index is a count of the number of simplifying and complicating features on a ballot. The continuous measure of the grade level of ballot instructions is recoded to a three-category measure of low (fourth grade to eighth grade), medium (ninth grade to eleventh grade), and high (twelfth grade and above) reading levels to incorporate it into the summary index.[9] In creating the index, we sum features hypothesized to simplify the voting process and subtract features hypothesized to make voting more difficult. None of the ballots in our sample was perfect on all indicators (which would be an index score of +5), and none of the ballots in our sample failed on all of the features (which would be an index score of –4). The ballot in figure 1 has a score of +4 in the ballot index, while the ballot in figure 2 has a score of –2 on the ballot index. The ballot index values in our data range from –3 to +4, with a mean of 1.1 and a standard deviation of 1.7.

As a preliminary step, we examined bivariate relationships between unrecorded votes and each ballot feature described above (presented in an online appendix to this article). Each of the bivariate relationships is in the expected

9. The low reading level is coded as "0," medium is coded "1," and high is coded "2."

Table 1. Frequency of Ballot Features in Sample Counties by State

Ballot Feature	Florida	Illinois	Iowa	Kansas	Tennessee	Total
Instructions in top left corner of ballot	33%	0%	77%	57%	90%	56%
Mean grade level of instructions (SD)	6.7 (0.6)	8.5 (0.3)	6.9 (1.2)	10.2 (1.1)	6.0 (1.1)	8.2 (1.9)
Instructions mention how to correct a spoiled ballot	100%	100%	72%	98%	80%	90%
Instructions warn about the consequences of a spoiled ballot	98%	0%	0%	0%	30%	22%
Shading to identify different offices	44%	67%	71%	2%	10%	36%
Bolded text to differentiate offices from candidate names	87%	60%	14%	31%	50%	40%
Possible confusion in marking governor votes	23%	33%	8%	34%	10%	22%
Clutter around candidate names	12%	20%	72%	100%	0%	64%
Mean index of ballot features (SD)	3.3 (0.7)	0.8 (1.1)	1.4 (1.0)	–0.5 (0.9)	2.4 (1.1)	1.1 (1.7)
Number of counties in sample	52	15	78	95	10	250

direction: unrecorded vote rates are higher when a confusing ballot feature is present and lower when a simplifying feature is present. Each ballot feature except for the location of instructions has a statistically significant bivariate association with unrecorded votes. In the multivariate analyses that follow, we use the index of ballot features as the primary independent variable. We believe that each of the ballot features described above derive from the same basic design principles. In a factor analysis, each item loaded on the main factor in the direction we hypothesized: positive for simplifying ballot features, negative for complicating ballot features. In addition, using the ballot index instead of separate measures for each ballot feature substantially reduces multicollinearity. We expect the summary index to be negatively correlated with unrecorded votes.[10]

Table 1 summarizes the frequency of the ballot features described above in each state of our study. There is a great deal of variation across states and within states. With one exception, there are few cases where 100 percent or 0 percent of a state's counties share the same ballot feature. The one exception is the case of spoiled ballot instructions. Almost all Florida counties warn voters of the consequences of a spoiled ballot, while no counties in Illinois, Iowa, or Kansas

10. Cronbach's alpha (a measure of scale reliability) for our index of ballot features is .56.

warn of the consequences of a spoiled ballot. In addition, almost all counties in Florida, Illinois, and Kansas include instructions for correcting a spoiled ballot. We believe that the variation in ballot features is due to the fact that most ballot design and printing decisions are made by local election officials who are not trained in survey or graphic design methods. The next section describes a multivariate analysis to more rigorously test our ballot design hypotheses.

Multivariate Analyses

To assess the impact of ballot features, we estimate a model of unrecorded votes in the 2002 gubernatorial elections for the counties in our sample. The model includes the ballot index, as well as several control measures as explanatory variables. For controls, our multivariate analysis includes two other ballot characteristics that are expected to affect unrecorded votes. One variable measures whether the ballots in our sample contain a straight-party option. The straight-party feature is not a matter of discretion for county election officials, for it is either required or banned by state law. Iowa is the only state in our sample that requires a straight-party option on the ballot. We expect the option to simplify the voting task and thereby reduce unrecorded votes.

For optical scan voting systems, there are two ways to mark a ballot. Voters either darken an oval (see the ballot in figure 2) or connect the point and stump of an arrow next to the name of their preferred candidate (see the ballot in figure 1). The connect-the-arrow format tends to produce a higher number of unrecorded votes because this is a task with which voters are not generally familiar (Bullock and Hood 2002). Thus, the ballot in figure 1, which we describe as having several positive features, also has a significant demerit (the connect-the-arrow format). Our multivariate model includes a dummy variable for counties with a connect-the-arrow format since we expect more voting errors on these ballots.

Our multivariate analysis includes two voting technology measures. We include a dummy variable for counties using precinct-count optical scan systems, since they have a feature that allows voters to detect and correct mistakes, reducing unrecorded votes. We include a second dummy variable for hand-counted paper ballots, leaving central-count optical scan ballots as the excluded category. At least one study suggests that hand-counted paper ballots yield relatively low rates of unrecorded votes (Ansolabehere and Stewart 2005).

Finally, the model includes several demographic variables that often correlate with unrecorded votes. These control variables include the percentage of a county's residents who are African-American, the percentage who are over the age of 65, the percentage of adults with a high school degree, and the natural log of the county's population, all obtained from the U.S. Census Bureau. Based on previous studies, we expect unrecorded votes to be positively correlated with the size of the African-American and elderly populations and negatively correlated with the percentage of high school graduates. As for population, previous studies indicate that the smaller the county, the larger the number of unrecorded votes (Brady et al. 2001; Kimball, Owens, and Keeney 2004; Knack and Kropf 2003a, p. 887). This finding may be due to election administration—larger urban counties tend to have more professional operations than smaller rural counties. Thus, we expect the natural log of the population to be negatively correlated with unrecorded votes.[11]

The dependent variable is the number of ballots cast in each county that fail to record a valid vote for governor. There are concerns that least squares regression does not adequately capture the data generating process for unrecorded votes. First, the number of unrecorded votes has a lower bound of zero, but least squares regression models do not constrain the expected value to be greater than or equal to zero. Second, counts of unrecorded votes have a strongly skewed distribution: most observations cluster slightly above zero with a long tail extending in the positive direction.

As a result, we estimate a negative binomial regression model more appropriate where the dependent variable is a count.[12] The negative binomial model is used when there is "overdispersion" in count data (Long 1997), meaning that there is higher than expected variation in the dependent variable. Overdispersion can occur in count data if events are clustered in particular locations or time periods. For example, an unrecorded vote in a particular precinct or county may be linked to more unrecorded votes in the same place (because of faulty equipment in that location, for example). The negative binomial regression model estimates an extra parameter (alpha) to test for overdispersion.

We also modify the generic negative binomial regression model. Since the number of voters in each county varies dramatically, we weight each county by the number of ballots cast, which has the effect of counting each voter equally. In other words, it prevents the results from being overly influenced by the many small counties that have relatively few voters. It is also possible that observations within the same state are not independent, due to unobserved state-level factors (such as state administration or political culture). Thus, we estimate standard errors that are corrected for the clustering of counties within a state (Williams 2000).[13]

11. Two potential control variables are the number of candidates for governor on the ballot and the ballot placement of the governor's contest (as a test of the ballot fatigue hypothesis). Too few candidates may indicate too few choices for voters, but too many candidates may cause confusion for voters. We do not include either measure in this study because there is little variance on these measures for the states in this study. Adding either control produces an insignificant coefficient and does not change the other results.

12. Bullock and Hood (2002) and Sinclair and Alvarez (2004) also estimate negative binomial models to examine unrecorded votes.

13. We use the *nbreg* command in Stata version 8 to estimate the model. The corrected standard errors are specified by the "cluster" option. The total number of ballots cast in a county is used to standardize the counts with the "exposure" option. This simply accounts for the fact that there are more opportunities for unrecorded votes in larger counties than in smaller counties. Long and Freese (2001, pp. 241–50) describe the negative binomial model and the exposure concept. Some might prefer using ordinary least squares (OLS) regression to model the unrecorded vote rate. OLS regression analyses (available from the authors) produce similar results to the ones reported here.

The results of our negative binomial regression analysis are presented in table 2. The first column provides the results of a baseline model that includes the index of ballot features, voting technology, and demographic measures as independent variables.[14] The second model in table 2 includes an interaction between the index of ballot features and the percentage of African-American residents in a county. To assess the substantive impact of each independent variable, negative binomial coefficients do not yield straightforward interpretation. As a result, we use the model estimates to compute the percent change in the expected number of unrecorded votes given a change in a particular variable, while holding the other variables constant. These "percent change" calculations are included in table 2.[15]

In general, our analyses suggest that ballot design features affect the rate of unrecorded votes in the way we hypothesize. The substantive impact of the ballot index is strong as well. Based on the results in model 1, moving from one standard deviation below the mean to one standard deviation above the mean of the ballot index reduces the expected number of unrecorded votes by 58 percent, holding other factors constant. Compared to the percent change calculations for all other independent variables in the model, the ballot index has the strongest substantive impact on unrecorded votes.

Considering other ballot features, we find that a straight-party ballot option does not significantly reduce the frequency of unrecorded votes, while unrecorded votes are more common on ballots with a connect-the-arrow format. Holding all other factors constant, the connect-the-arrow format increases the expected number of unrecorded votes by 15 percent. The controls for voting technology and demographics tend to support the findings of past studies. We find that unrecorded vote rates are substantially lower in counties using precinct-count optical scan systems, which have an error correction mechanism not available in counties using hand-counted paper ballots or centrally counted optical scan ballots. At the same time, the frequency of unrecorded votes is no higher on hand-counted paper ballots than on centrally counted optical scan ballots.

We also find that unrecorded votes are more common in counties with large concentrations of African-American voters. We also find a significant interaction between race and the ballot index (model 2 in table 2), suggesting that well-designed ballots minimize the impact of race on unrecorded votes. To

14. Adding each ballot feature as a separate explanatory variable (instead of the ballot index) causes a high level of multicollinearity in the regression model and only slightly improves the goodness-of-fit.

15. For dummy variables, we calculate the expected percent change in unrecorded votes when switching from one category of the independent variable to the other category (e.g., when switching from a darken-the-oval ballot to a connect-the-arrow ballot). For other independent variables, including the ballot index, we compute the expected change in unrecorded votes when switching from one standard deviation below the mean to one standard deviation above the mean. We use the *listcoef* suite of commands developed for Stata by Long and Freese (2001, pp. 232–34) to calculate the percent change in expected values reported in tables 2 and 3.

Table 2. Multivariate Analyses of Unrecorded Votes in the 2002 Gubernatorial Elections

Explanatory Variable	Model 1		Model 2	
	Coefficient (SE)	Percent Change	Coefficient (SE)	Percent Change
Ballot Features				
Index of ballot features	–.21*** (.03)	–58%	–.15** (.05)	–44%
Straight-party option	–.11 (.11)	–11%	–.11 (.13)	–11%
Connect-the-arrow format	.14** (.05)	15%	.11* (.05)	11%
Voting Technology				
Error correction feature (in precinct-count optical scan)	–.38*** (.10)	–32%	–.38*** (.12)	–32%
Hand-counted paper ballot	.04 (.15)	5%	.14 (.13)	15%
Demographic Controls				
Percent African-American	.012* (.005)	26%	.021* (.012)	47%
Percent 65 or older	–.004 (.006)	–5%	–.009 (.008)	–10%
Percent with a high school degree	–.011*** (.001)	–12%	–.009* (.005)	–10%
County population (natural log)	–.20*** (.02)	–47%	–.21*** (.02)	–49%
Index of ballot features × percent African-American			–.005* (.004)	–29%
Constant	–.83* (.42)		–.93* (.51)	
Alpha	.10*** (.02)		.10*** (.02)	
Number of Cases	250		250	
Model χ^2	933.3***		1060.6***	

NOTE.—The dependent variable is the number of ballots cast that failed to record a valid vote for governor. Cell entries are negative binomial regression coefficients with standard errors in parentheses. Standard errors are corrected for nonindependence of errors within states. Observations (counties) are weighted by the number of ballots cast in the 2002 election. The "percent change" column reports the expected change in the number of unrecorded votes given a change in a particular variable, while holding the other variables constant. (See note 15 for more details.)

* $p < .1$, one-tailed.

** $p < .01$.

*** $p < .001$.

demonstrate the conditional relationship, we use the coefficient estimates and variance-covariance matrix from model 2 in table 2 to calculate the impact of race on unrecorded votes for different levels of the ballot index. The main effect of race reported in model 2 (coefficient of .021, standard error of .012) reflects the impact of race in a county with a ballot index score of 0 (slightly below average). The significant relationship between race and unrecorded votes is strongest in counties with poorly designed ballots (a ballot index score of –3), with a coefficient of .035 (standard error of .022). In contrast, the relationship between race and unrecorded votes is statistically insignificant in counties with high scores (+3 or above) on the ballot index, where the coefficient drops to .008 (standard error of .008). This result is consistent with other studies that indicate that voting procedures designed to help voters tend to minimize the relationship between race and unrecorded votes.

In addition, unrecorded votes are more common in less populated counties. Interestingly, the coefficient for county population gets smaller when the ballot index is added to the equation. In our sample the ballot index is strongly correlated with the natural log of county population ($r = .47$, $p < .001$). This suggests that higher rates of unrecorded votes occur in less populated counties partly because ballots tend to be more confusing in less populated counties. We also find that higher levels of education are related to a reduction of unrecorded votes. Finally, we find no relationship between the percentage of people over age 65 in a county and unrecorded votes.

EXAMINING OVERVOTES AND UNDERVOTES

We also examine the impact of ballot features on overvotes and undervotes. The mechanisms underlying overvotes and undervotes may be different. Overvotes are almost always unintentional, whereas undervotes may be intentional. Put differently, overvotes are almost always the result of voting errors, while undervotes may result from voting errors or the voter's intent. We collected complete data on overvotes and undervotes from 133 counties (slightly more than half of the counties in our sample). While our conclusions about the predictors of overvotes and undervotes are more tentative due to the smaller sample, the results support our ballot design hypotheses. We repeat similar models described above. The dependent variables are the number of ballots cast with overvotes or undervotes for governor, respectively.

Looking first at overvotes (the first column in table 3), the results indicate that overvotes are almost entirely a function of ballot features and voting technology, rather than demographics. Furthermore, our results suggest that the ballot index is a significant predictor of overvotes. Based on the model results, moving from one standard deviation below the mean to one standard deviation above the mean in the ballot index produces a 76 percent reduction in the expected number of overvotes. It is a goal of election officials to reduce the number of overvotes (which are the result of voting mistakes) to zero. To provide

Table 3. Multivariate Analyses of Overvotes and Undervotes in the 2002 Gubernatorial Elections

	Overvotes		Undervotes	
Explanatory Variable	Coefficient (SE)	Percent Change	Coefficient (SE)	Percent Change
Ballot Features				
Index of ballot features	–.34*** (.05)	–76%	–.15*** (.03)	–39%
Straight-party option	–.64** (.25)	–48%	.05 (.16)	6%
Connect-the-arrow format	1.64*** (.23)	413%	–.006 (.09)	–1%
Voting Technology				
Error correction feature (in precinct-count optical scan)	–1.50*** (.39)	–78%	–.26* (.13)	–23%
Demographic Controls				
Percent African-American	–.001 (.008)	–3%	.010*** (.002)	21%
Percent 65 or older	.010* (.006)	12%	–.005* (.004)	–6%
Percent with a high school degree	–.023 (.031)	–23%	–.005 (.004)	–5%
County population (natural log)	–.09 (.07)	–20%	–.24*** (.05)	–49%
Constant	–2.83 (2.30)		–1.25* (.67)	
Alpha	.56*** (.06)		.09*** (.02)	
Number of Cases	133		133	
Model χ^2	118.3***		581.8***	

NOTE.—The dependent variable is the number of ballots cast with overvotes or undervotes for governor, as indicated at the top of the column. Cell entries are negative binomial regression coefficients with standard errors in parentheses. Standard errors are corrected for nonindependence of observations within states. Observations (counties) are weighted by the number of ballots cast in the 2002 election. The "percent change" column reports the expected change in the number of overvotes or undervotes given a change in a particular variable, while holding the other variables constant. (See note 15 for more details.)

* $p < .1$, one-tailed.

** $p < .01$.

*** $p < .001$.

another view of the impact of ballot features on overvotes, we use the negative binomial estimates in the first column of table 3 to estimate the probability of zero overvotes for different types of ballots, holding other factors constant at median values. For poorly designed ballots (with low scores on the ballot index), the probability of producing zero overvotes is less than .1, while well-designed ballots (with high ballot index scores) increase the probability of zero overvotes to almost .3.

As expected, other ballot and voting technology features are also strongly associated with overvotes. The connect-the-arrow format increases the expected number of overvotes by a shocking 413 percent, controlling for other factors. This is related to the overdispersion in the distribution of overvotes across counties. Relatively high levels of overvotes are concentrated in a small number of counties, many of which use the connect-the-arrow format. We also find that overvotes are substantially less common in counties using the error correction feature in precinct-count optical scan ballots. Finally, a straight-party punch reduces the expected number of overvotes.

Our results also suggest that ballot features have a negative and statistically significant effect on undervotes (the second column in table 3). Furthermore, undervotes are less common in counties using precinct-count optical scan ballots. In comparing the two regression models in table 3, it appears that various ballot features and the precinct-count mechanism have a stronger impact on overvotes than on undervotes. This is consistent with the view that overvotes are voting mistakes caused by voting technology and procedures, while some undervotes are intentional and thus unaffected by voting technology, ballot design, and other election procedures. However, the fact that the ballot index and error correction features of precinct-count optical scan balloting are significant determinants of undervotes suggests that some undervotes are unintentional, consistent with studies that examine the incidence of intentional undervoting at the top of the ballot (Knack and Kropf 2003b).

For the demographic control variables, there are two findings worth noting. First, undervotes are more common in counties with higher concentrations of African-Americans. Second, less populated counties tend to have higher levels of undervotes. However, population is unrelated to overvote rates. It appears that the higher rates of unrecorded votes observed in less populated counties are due to undervotes but not overvotes. Many small rural counties are outside major media markets, where most campaigning in national and statewide contests occurs. Thus, higher rates of undervotes in small counties may be due lower levels of interest in top-of-the-ballot contests in those counties.

Conclusion

In general, we find that ballot format has a significant (both substantively and statistically) effect on unrecorded votes in the 2002 gubernatorial races we studied. This should not come as a surprise to scholars who have studied how the format of self-administered surveys affects nonresponse. These results suggest that ballot design can be an important factor in determining whether someone is able to cast a ballot accurately. While unrecorded votes did not affect the outcome of the contests we examined, the cumulative effect of poor ballot design can be substantial, especially where elections are close, as Florida was in 2000.

Ironically, researchers have spent much effort to determine survey features that decrease survey and item nonresponse, yet ballots in the United States receive much less attention. Before the 2000 election, ballot nonresponse was not a major concern for election officials. In the wake of the Florida controversy and subsequent legislation intended to reduce the frequency of unrecorded votes, election officials and researchers have spent more time analyzing sources of voting errors and pursuing election reforms. However, much of the reform effort has focused on upgrading voting technology. While new voting technology is likely to help, as indicated by these results, voting equipment is not the only source of voting error.

There are several implications of this work. First, we find ballot design effects in contests where one might not expect to find them—in five, highly competitive races for governor that appear at or near the top of the ballot. Voters are likely to be more interested in the governor's contest than other races on the ballot, and they are likely to pay closer attention to casting their votes in such a race. This suggests that ballot design effects may be even stronger in down-ballot contests that do not capture the interest of voters to the same degree.

Second, while this study only focuses on paper-based ballots, we believe a similar approach is needed to examine ballot design for electronic voting machines. Given the growing popularity of electronic machines in the United States, it is important that the interface and layout are easy for voters to comprehend. Researchers have already begun to apply similar theories of survey response to computer-based questionnaires (Couper et al. 2000; Tourangeau, Couper, and Conrad 2004) and electronic machines (Callegaro and Peytcheva 2003).

Third, purchasing new voting equipment can be expensive and difficult during the current retrenchment in state and local government budgets. In contrast, modifying ballot features to create a less confusing layout is relatively inexpensive. Buying new voting equipment without paying attention to ballot design may not produce the desired effect of reducing unrecorded votes (although the precinct-count optical scan system is one piece of equipment that reduces voting errors, even controlling for ballot design). Ballot design deserves closer inspection than it has received thus far, and election officials should consider their ballot design decisions carefully.

References

Alvarez, R. Michael, D. E. "Betsy" Sinclair, and Catherine Wilson. 2003. "Counting Ballots and the 2000 Election: What Went Wrong?" In *Rethinking the Vote*, ed. Ann Crigler, Marion Just, and Edward McCaffery, pp. 34–50. New York: Oxford University Press.

Ansolabehere, Stephen, and Charles Stewart III. 2005. "Residual Votes Attributable to Technology." *Journal of Politics* 67(2):365–89.

Babbie, Earl R. 1990. *Survey Research Methods*. Belmont, CA: Wadsworth.

Brady, Henry E., Justin Buchler, Matt Jarvis, and John McNulty. 2001. "Counting All The Votes: The Performance of Voting Technology in the United States." Survey Research Center and Institute of Governmental Studies, University of California, Berkeley. Available online at http://ucdata.berkeley.edu/new_web/countingallthevotes.pdf (accessed September 26, 2005).

Bullock, Charles S., III, and Richard E. Dunn. 1996. "Election Roll-Off: A Test of Three Explanations." *Urban Affairs Review* 32:71–86.

Bullock, Charles S., III, and M. V. Hood III. 2002. "One Person—No Vote; One Vote; Two Votes: Voting Methods, Ballot Types, and Undervote Frequency in the 2000 Presidential Election." *Social Science Quarterly* 83(4):981–93.

Callegaro, Mario, and Emilia Peytcheva. 2003. "Electronic Voting Machines: A Comparison of the Principles of Human Computer Interaction and Computer-Assisted Questionnaire Design." Paper presented at the annual meeting of the American Association for Public Opinion Research, Nashville, TN.

Caltech/MIT Voting Technology Project. 2001. "Voting: What Is, What Could Be." Available online at http://www.vote.caltech.edu/reports/2001report (accessed September 26, 2005).

Christian, Leah Melani, and Don A. Dillman. 2004. "The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions." *Public Opinion Quarterly* 68:57–80.

Couper, Mick P., Paul Beatty, Sue Ellen Hansen, Mark Lamias, and Theresa Marvin. 2000. "CAPI Design Recommendations." Report submitted to the Bureau of Labor Statistics. Survey Research Center, University of Michigan.

Darcy, R., and Anne Schneider. 1989. "Confusing Ballots, Roll-Off, and the Black Vote." *Western Political Quarterly* 42:347–64.

Design for Democracy. 2002. "Case Studies." Available online at http://www.electiondesign.org/case.html (accessed March 24, 2004).

De Vaus, David. 2002. *Surveys in Social Research*. 5th ed. London: Routledge.

Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley and Sons.

———. 2002. "Navigating the Rapids of Change: Some Observations on Survey Methodology in the Early Twenty-First Century." *Public Opinion Quarterly* 66(3):473–94.

Dillman, Don A., Michael D. Sinclair, and Jon R. Clark. 1993. "Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys." *Public Opinion Quarterly* 57:289–304.

Gower, Allen R., and Ruth Dibbs. 1989. "Cognitive Research: Designing a 'Respondent Friendly' Questionnaire for the 1991 Census." In *Proceedings from the Fifth Annual Census Bureau Research Conference*, pp. 257–66. Washington, DC: Census Bureau.

Heilke, Thomas, Mark R. Joslyn, and Alex Aguado. 2003. "The Changing Readability of Introductory Political Science Textbooks: A Case Study of Burns and Peltason, *Government by the People*." *PS: Political Science and Politics* 36:229–32.

Help America Vote Act (HAVA) of 2002. U.S. Public Law 252. 107 Cong., 2d sess., October 29.

Herron, Michael C., and Jasjeet S. Sekhon. 2003. "Overvoting and Representation: An Examination of Overvoted Presidential Ballots in Broward and Miami-Dade Counties." *Electoral Studies* 22:21–47.

Jenkins, Cleo R., and Don A. Dillman. 1995. "The Language of Self-Administered Questionnaires as Seen through the Eyes of the Respondents." *Statistical Policy Working Paper 23: New Directions in Statistical Methodology, Volume 3*, pp. 470–516. Washington, DC: U.S. Office of Management and Budget.

———. 1997. "Towards a Theory of Self-Administered Questionnaire Design." In *Survey Measurement and Process Quality*, ed. Lars Lyberg, Paul Biemer, Martin Collins, Edith deLeeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin, pp. 165–96. New York: Wiley-Interscience.

Jewett, Aubrey. 2001. "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.

Kahneman, Daniel. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kamin, Blair. 2004. "The (Design) Fix Is In: In a Stunning Turnabout, Chicago and Cook County Have Become National Leaders in Election Reform." *Chicago Tribune*, March 16.

Kimball, David C., Chris T. Owens, and Katherine M. Keeney. 2004. "Residual Votes and Political Representation." In *Counting Votes: Lessons from the 2000 Presidential Election in Florida*, ed. Robert P. Watson, pp. 135–50. Gainesville: University Press of Florida.

Knack, Stephen, and Martha Kropf. 2003a. "Invalidated Ballots in the 1996 Presidential Election: A County-Level Analysis." *Journal of Politics* 65(3):881–97.

———. 2003b. "Roll Off at the Top of the Ballot: Intentional Undervoting in American Presidential Elections." *Politics and Policy* 31(4):575–94.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.

Long, J. Scott, and Jeremy Freese. 2001. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press.

Long, William F., Ralph P. Garzia, Timothy Wingert, and Sylvia R. Garzia. 1996. "The Ergonomics of Reading." In *Vision and Reading*, ed. Ralph P. Garzia, pp. 71–110. St. Louis, MO: Mosby.

Nichols, Stephen M. 1998. "State Referendum Voting, Ballot Roll-off, and the Effect of New Electoral Technology." *State and Local Government Review* 30:106–17.

Nichols, Stephen M., and Gregory A. Strizek. 1995. "Electronic Voting Machines and Ballot Roll-Off." *American Politics Quarterly* 23:300–18.

Niemi, Richard G., and Paul S. Herrnson. 2003. "Beyond the Butterfly: The Complexity of U.S. Ballots." *Perspectives on Politics* 1(2):317–26.

Omandam, Pat. 2002. "Ballot May Confuse, Design Expert Says." *Honolulu Star-Bulletin*, September 19. Available online at http://starbulletin.com/2002/09/19/news/story3.html (accessed September 26, 2005).

Redline, Cleo D., and Don A. Dillman. 2001. "The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires." In *Survey Nonresponse*, ed. Robert Groves, Don Dillman, John L. Eltinge, and Roderick Little, pp. 179–96. New York: John Wiley and Sons.

Roth, Susan King. 1994. "The Unconsidered Ballot: How Design Affects Voting Behavior." *Visible Language* 28:48–67.

Sanders, Mark S., and Ernest J. McCormick. 1993. *Human Factors in Engineering and Design*. 7th ed. New York: McGraw-Hill.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* New York: Academic Press.

Sinclair, D. E. "Betsy," and R. Michael Alvarez. 2004. "Who Overvotes, Who Undervotes, Using Punchcards? Evidence from Los Angeles County." *Political Research Quarterly* 57:15–25.

Sinclair, Robert C., Melvin M. Mark, Sean E. Moore, Carrie A. Lavis, and Alexander S. Soldat. 2000. "Psychology: An Electoral Butterfly Effect." *Nature* 408:665–66.

Tefki, Chafai. 1987. "Readability Formulas: An Overview." *Journal of Documentation* 43:261–73.

Tomz, Michael, and Robert Van Houweling. 2003. "How Does Voting Equipment Affect the Racial Gap in Voided Ballots?" *American Journal of Political Science* 47(1):46–60.

Tourangeau, Roger, Mick P. Couper, and Frederick Conrad. 2004. "Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions." *Public Opinion Quarterly* 68:368–93.

Wallschlaeger, Charles, and Cynthia Busic-Snyder. 1992. *Basic Visual Concepts and Principles for Artists, Architects and Designers*. Dubuque, IA: William C. Brown Publishers.

Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael C. Herron, and Henry E. Brady. 2001. "The Butterfly Did it: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95:793–810.

Wattenberg, Martin P., Ian McAllister, and Anthony Salvanto. 2000. "How Voting is Like Taking an SAT Test: An Analysis of American Voter Rolloff." *American Politics Quarterly* 28:234–50.

Williams, Rick L. 2000. "A Note on Robust Variance Estimation for Cluster-Correlated Data." *Biometrics* 56:645–46.

Zukerberg, Andrew, and Meredith Lee. 1997. "Better Formatting for Lower Response Burden." Census Bureau Research Report Series. Available online at http://www.census.gov/srd/www/byyear.html (accessed June 30, 2003).