Introduction to the Measure of American Elections
Barry C. Burden and Charles Stewart III

How good are American elections?
Where would one start in answering this question?
Whenever this question is posed, it is common to answer it from the position of deeply held beliefs, but rarely from the position of a systematic analysis of facts. These beliefs might arise from partisanship: a good election is one that my favored candidate wins. These beliefs might be chauvinistic: a good election is one run according to the rules of my community.
Rarely are these beliefs rooted in hard facts.
When facts intervene, they rarely are presented in a systematic fashion. Opinions about levels of voter fraud might be attributable to a viral YouTube video. Concerns about the effects of a new voter identification law might be informed by a reporter’s interview with an activist who is eager to share stories about how voters she has talked with will be disenfranchised on Election Day. Satisfaction with a new electronic voting machine may be illustrated by a picture of a smiling citizen coming out of the precinct with an “I Voted” sticker stuck to her lapel. Disdain about the ability of local governments to run elections might follow from a newspaper article detailing yet another season of long lines when waiting to vote in Florida (or South Carolina or Maryland or …). At its worst, this approach is evaluation by anecdote.
Consider instead how the question about the quality of American elections would be framed if first we asked about other policy domains: “How good are America’s prisons?” or “How good are America’s schools?” or “How good is America’s health care system?” Some people surely would respond based on fact-free beliefs; others would respond with a random story about the experience that one’s cousin had with one of these institutions. However, it would not be difficult to discover that in
2007 (the most recent year for which data are available), 15.5 percent of all parolees were reincarcerated, that Connecticut had the highest reincarceration rate (29.9 percent), and that Maine had the lowest (0 percent). Nor would it be difficult to find out that Alaska’s fourth graders ranked last among the fifty states in the reading portion of the 2011 National Assessment of Educational Progress and that Massachusetts ranked first; that the gap between girls and boys was greatest in Hawaii and smallest in Texas; and that the gap between whites and blacks was greatest in Connecticut and smallest in North Dakota. A brief Internet search would reveal that in 2008, the infant mortality rate among the fifty states ranged from 3.87 per 100,000 live births in New Hampshire to 9.95 per 100,000 in Mississippi (Mathews and MacDorman 2012, 6).

In other words, an obvious way to begin addressing questions about the state of public policy in these other important areas would be to draw on a large body of data about the performance of these institutions and policy systems.

None of the statistics just referenced is the be-all and end-all of the question about how well the prison systems, schools, and health care systems work in the states. The point is that in each of these policy domains, significant effort is poured into defining measures of policy input and output consistently across states, multiple measures of system performance are regularly reported through a federal agency, and entire professions have grown up to analyze these data. Despite the fact that answers to policy questions about criminal justice, education, and health care are legitimately informed by ideology and deeply held beliefs, even committed ideologues typically ground their appeals in statistics when they argue about policy; some will even be convinced they are wrong if the facts are against them. The data provide a common starting point.

This returns us to the original question: How good are American elections? If an American wanted to argue this question based on facts, he would most likely go to the turnout statistics and discover that in 2012, 58.2 percent of eligible Americans voted, ranging from 44.2 percent in Hawaii to 75.7 percent in Minnesota. Compared with other countries, the U.S. turnout was in the lower half of 112 countries.

---

with presidential elections in recent years – ranked seventy-fourth (right between the Democratic Republic of Congo and Romania), above the European nations of Poland, Serbia, Portugal, Slovakia, Ireland, Austria, Macedonia, and Lithuania, but below the nations of Moldova, the Russian Federation, Montenegro, Finland, Ukraine, Iceland, France, Kazakhstan, Armenia, Cyprus, Tajikistan, Belarus, Turkmenistan, and Uzbekistan.\(^4\) An American with a bit more perseverance might also discover that when asked in a national survey in 2012 whether they were confident that their votes were counted as cast, 63 percent stated they were “very confident,” ranging from 80 percent in Vermont to 54 percent in Washington.\(^5\)

An American who was interested in understanding in a more nuanced sense how well elections are run in this country would not find much in the way of defining consistent measures of election administration input and output across states. Nor would he or she find much effort at reporting such statistics even within states, much less a profession devoted to the proposition that elections would function better if we understood systematically the facts associated with election administration. Indeed, this American would find that states differ in how they even define critical aspects of election administration, including such fundamental measures as turnout. Sometimes even a state’s chief election officer lacks the authority to require local election boards to report to him or her basic statistics about election administration beyond the bare facts of the election returns themselves. This American would also discover that the National Association of Secretaries of State (NASS), which is the organization of top state constitutional officers who most often are ultimately responsible for the conduct of elections, has led a campaign to abolish the U.S. Election Assistance Commission (EAC), the only federal agency that gathers statistics about election administration nationwide.

In other words, while there are scientific professions devoted to the study of corrections, education, public health, transportation, and many other critical functions of state and local government, there is no scientific profession devoted to the study of election administration.

This book is part of an effort to change that.


In particular, the chapters in this volume are devoted to the study of ten areas of election administration through the lens of hard data and social science. The topics cover the waterfront in the field of election administration and policy, ranging from the registration of voters to the counting of votes. None of these chapters is the final word in any of these areas. They are something more important: in many cases, they are the first word in starting a conversation about the systematic analysis of election administration and policy in America.

**FIRST THINGS FIRST:**

**ESTABLISHING THE VALIDITY AND RELIABILITY OF AMERICAN ELECTION ADMINISTRATION**

In the areas of public policy that aspire to be data driven, great attention is paid to the validity and reliability of the key measures employed in the field.

*Validity* may be defined as the degree to which a measure actually describes the underlying concept it claims to measure, rather than something else. In other words, it refers to how well an observable quantity describes an unobservable theoretical construct. For instance, we cannot observe someone’s intelligence directly, but we can observe how well she or he performs on an IQ test. A robust literature and scholarly debate has grown up around the question of how valid IQ tests are as an indicator of (unobservable) intelligence.  

An example of the application of the concept of validity to elections is in the area of absentee ballots. For example, a valid measure of absentee ballot usage should track actual usage, even if it sometimes misses the mark somewhat. Similarly, the measure of line lengths is highly valid if it approximates the actual waiting times, rather than being consistently too high or too low.

There are many methods used to establish validity, including such things as relating a measure to other variables known to affect or be affected by it. This is often an interactive process of moving back and forth between measures and their actual scores (Adcock and Collier 2001).

---

6. Within political science, the classic statement on validity and reliability is Carmines and Zeller (1979).

7. This sort of refinement is discussed in some of the chapters that follow. Here we mention two examples. The absentee ballot rejection rate is an important measure that could be computed...
In contrast, *reliability* refers to the degree to which a measuring procedure yields the same results when the procedure is repeated. A highly reliable measure produces similar results when applied multiple times in the same setting. For example, a reliable measure of a state’s absentee ballot usage will yield a similar conclusion even if different staff members provide the data. Likewise, a reliable measure of polling place lines based on surveys of voters will show consistent times if the survey is repeated. Although low reliability does not affect accuracy of a measure, it does affect its precision. That is, while the answer might be right on average, the high degree of variability makes it difficult to discern the signal amid the noise. One way to increase reliability is to bring more data to the table by adding observations or combining measures into a summary index.

In psychometrics a great deal of attention is paid to whether respondents give the same answers when given the exact same battery of questions on different days. Similarly, a diagnostic medical test that gives the same results when repeated on the same individual in rapid succession is said to be reliable.

Measures of validity and reliability are generally measures of *degree* and not absolutes. It is well understood that reliable measures may fluctuate from one moment to the next, as when the results of an aptitude test vary depending on whether a student is well rested, under stress, or hungry. Thus, the standard is rarely perfection when it comes to assessing validity and reliability, but the closer we can come to perfection, the better.

On a scale where 1.0 indicates perfect reliability, most of the state measures used in the chapters that follow correlate between the 2008 and 2010 elections at levels between .7 and .9. Where the correlations are lower, it is sometimes a sign of poor reliability stemming from a small sample size. Aggregating over multiple elections will remedy this problem. In other cases, it is a sign that what the measure is capturing has changed between the presidential election and the midterm election. Several of the Uniformed and Overseas Citizens Absentee Voting Act (UOCAVA) and absentee ballot measures are of this variety.

as a share of absentee ballots cast or of all ballots cast. Which is used depends on both theory about the denominator of interest and also empirical information about which offers greater discrimination across states. Voter confidence is also of keen interest. Survey questions often ask whether the respondent is “very confident,” “somewhat confident,” “not too confident,” or “not at all confident.” Where to “cut” these four categories to create state-by-state percentages is partially a result of trial and error to find the most meaningful division.
In some fields, the reliability and validity of performance measures are based on a deep body of scientific research, stretching across decades and thousands of researchers, supported by billions of dollars of basic research. Consider, for instance, a study to assess the quality of medical care received by Medicare recipients in each of the fifty states, which was published in the *Journal of the American Medical Association* (JAMA) in 2000 (Jencks et al. 2000). To assess the quality of medical care received by Medicare patients, this study examined thousands of randomly chosen patient records, drawn from patients in every state in the nation. The basic question asked of these records was whether the patient in question was treated for his or her conditions in a way that reflected the consensus of appropriate care in a particular field. The overall index of patient care was based on rating twenty-four indicators of patient care such as these; the conditions covered by these indicators affected 85 percent of Medicare beneficiaries.

The validity of the measures chosen by the researchers in the JAMA study had already been established in the medical community, based on hundreds of studies of the relationship between the care patients received and how well they fared after their treatments. For instance, these previous studies had established that one valid measure of the quality of care of Medicare recipients was what percentage of those admitted to hospitals because of acute myocardial infarction (better known as a heart attack) later received prescriptions for aspirin upon discharge.

The reliability of such an indicator would be more difficult to measure, because it is based on the assessment of patient records. Although there may be general guidelines for care of patients, there also may be well-established contraindications for applying that regimen to a particular patient – such as if the patient is known to be violently allergic to the drug of choice. Here, judgment must be applied to the coding of patient records in the making of the index. This is where reliability comes in.

Two physicians, and certainly two research assistants, reading the same patient medical records might come to different conclusions about whether a particular patient is a candidate to receive treatment consistent with the conventional standards of care. If they each read the same record and come to different conclusions, we would infer that this method of coding whether patients received good hospital care based on their records is unreliable. Conversely, if the coders frequently agreed, we would regard this measure as more reliable.
The reliabilities of the measures used in the JAMA study were assessed precisely according to this logic. Reliability was established by testing to see whether two independent coders came to the same conclusions about whether to include or exclude a particular patient record from the study. Part of the study’s report included measures of inter-rater reliability for each of the indicators.

The effort just described to compare the quality of hospital care received by participants in the Medicare program across the fifty states rested on decades of medical research that consumes billions of dollars in taxes and private foundation support each year. The study of election administration barely has thousands of dollars devoted to it each year, so the reliability and validity of any measures of election administration quality that we consider must be viewed as provisional.

To continue with the medical analogy, election administration is similar to the earliest days of public health studies, when the only measure of health care quality was the mortality rate – the number of deaths divided by the number of people. This was a crude measure, but it was powerful. It could establish, for instance, that modern sewer systems improved the health of city dwellers, leading to greater attention on public works projects aimed at improving the quality of life of everyone and, ultimately, promoting economic growth.

As the raw mortality rate was used increasingly to compare health care outcomes across geographical units, hospitals, and demographic populations, it became clear that the validity of the crude mortality rate could be improved by adjusting based on the risk of dying, which might be quite independent of the quality of health care (or other public health factors). For instance, the mortality rate of a city with a high fraction of elderly residents would be higher than that of a city with few elderly residents, even if the sanitation facilities were equivalent in the two places. Thus, as the field of public health has advanced, the mortality measure has been improved by adjusting for risk factors such as age. In doing so, the new risk-adjusted mortality statistics have become even more valid measures of the underlying health of a given population.

Election administration is still in the “raw mortality rate” phase of measuring outcomes in its policy domain. Few measures of performance within election administration are widely understood and accepted.

Indeed, it could be argued that there is only one widely known measure of performance in this area: the turnout rate, which shares many of the advantages and disadvantages of the raw mortality rate. On the plus side, the turnout rate is intuitively understood and easily calculated.
Most of the country understood, for instance, that it was bad when only 25 percent of Mississippi’s voting-age population (VAP) turned out to vote in the 1960 presidential election. The country also understood that when the turnout rate in Mississippi rose to the 60 percent level in 2012, this was a valid indication that the state’s electoral system had improved over the intervening half-century.

On the negative side, the raw turnout rate does not take into account the “luck” or “skill” that might be involved. For instance, what do we make of the fact that the turnout rates of Hawaii and Texas in the 2008 presidential election were 44 percent and 50 percent, respectively, despite levels of educational attainment — a factor that strongly predicts whether an individual will vote — being much higher in Hawaii (90 percent high school graduates) than in Texas (81 percent high school graduates)? Would we consider, on a risk-adjusted basis, that turnout in Texas was actually much higher than in Hawaii, because efforts to get voters to the polls in the Lone Star State must battle against lower educational levels than those in Hawaii? Or what about the comparison of Minnesota and Mississippi? In raw turnout terms, it is no contest. The Minnesota turnout rate was 75 percent in 2008, whereas Mississippi’s was 61 percent — a fourteen-point deficit for Mississippi. But if we account for different levels of educational attainment using a simple technique that relies on linear regression, the tables are turned, with Minnesota suffering a five-point turnout deficit in “risk-adjusted turnout” in comparison with Mississippi.

As far as we are aware, no report of turnout rates that compares the states has ever reported turnout rates adjusted for “risk factors” such as education so that the effects of policy or other systemic factors on turnout can be better understood. The best studies to this point are multivariate analyses that attempt to control for “luck” factors, such as the demographics of the state’s electorate or efforts by campaigns to turn out voters. At least turnout levels of the states are periodically reported to the public and discussed. The same cannot be said of a

---

8  Educational attainment levels are taken from the 2011 American Community Survey (three-year series).
9  The method of risk adjusting here is based on a simple regression method, in which we regress the turnout rate of each state on the percentage of residents twenty-five years of age and up with a high school diploma and then generate the residuals of the regression. In this case, the Minnesota residual is +5.8 percentage points, and the Mississippi residual is +10.4 percentage points. The regression equation in this case was Turnout Rate = –84.8 + 1.68 H.S. Graduate Rate, with standard errors of 15.9 and 0.18, respectively. The R-squared statistic was .63, with 51 observations. The regression was weighted by turnout in each state.
long list of other measures of election administration that could easily be constructed from official sources, such as the rejection rate of provisional ballots or the nonreturn rates of absentee ballots.

The ideal measure is one that is both reliable and valid. That is, it has a high level of precision and a high level of accuracy. It is possible for a measure to do well on one dimension but poorly on the other, or poorly on both. A digital watch is highly precise, but it can be inaccurate if set to the wrong time; likewise, a mechanical watch without a second hand might be highly accurate while being far from precise. A key purpose of this volume is to evaluate a variety of election-related measures on these two dimensions. It is important to note that for the purposes of evaluating election performance, an invalid measure is not necessarily useless. If a measure of provisional ballot usage, for example, systematically underestimated the actual use across the states, it would be less valid but could still be used effectively to rank order states. In contrast, a valid measure with low reliability would produce problematic rankings because of the excessive noise in the indicator, but it could still be used to describe how states perform on average.

WHERE ARE THE DATA?

SOURCES OF DATA FOR THE ASSESSMENT OF AMERICAN ELECTIONS

This volume begins the process of assessing how elections are conducted in America by identifying a manageable set of candidate indicators, subjecting them to scrutiny, and examining them for what they tell us about elections in America. If we agree that data-driven scrutiny of American elections is to be commended, we need data. Luckily there are plenty of data out there, oftentimes hiding in plain view. The remaining chapters of this book rely on these data sources to quite a detailed level. Here we introduce the reader to the most important.

There are four major sources of data available for the assessment of American elections:

1. The Voting and Registration Supplement of the Current Population Survey (CPS), conducted by the U.S. Census Bureau.

A more detailed examination of many of these data sources may be found in Pew Center on the States (2012).
2. The Election Administration and Voting Survey (EAVS), conducted by the U.S. EAC.
3. State and local election board records.
4. Academic and commercial survey research.

*The Voting and Registration Supplement*

Ever since the 1960s, the U.S. Census Bureau has conducted a survey every other November about voting in the most recent federal general election, as a supplement to the Census Bureau’s monthly CPS. The primary purpose of the CPS is to help determine the unemployment rate and other economic statistics at the state level. The biennial survey that studies voting is called the Voting and Registration Supplement (VRS). The main CPS contains a treasure trove of information about the participants in the survey, including information about their ethnicity, education, housing, income, and (since 2010) disability status. Thus, it is possible to study the relationship between important demographic characteristics of American adults and the likelihood they will be registered and vote, and even the means used to register and to vote. The overall survey sample is quite large. In 2010, for instance, nearly 80,000 respondents were asked whether they voted in the most recent federal election, ranging from 578 in New Mexico to 5,862 in California.

Compared with other survey research that focuses on election behavior, the VRS is actually quite limited in what it asks about voting – some would say “focused.” It does not ask respondents whom they voted for. (This being a survey conducted by the federal government, it is easy to understand why.) Nor does it ask other questions that political scientists studying participation might be interested in, such as interest in politics, knowledge of politics, or stances on important issues. It essentially only asks respondents whether they voted and if they are registered; if they report they did not vote or were not registered, they are asked why not. Those who say they voted are asked what mode they used to vote (in person on Election Day, in person before Election Day, or via the mail). Finally, they are asked how long they have lived at their current address and how they registered to vote.

---

11 This description of items is based on the 2010 VRS. While the items change infrequently, there is some fluctuation in questions from time to time. For instance, after the controversy that arose in 1980 over Jimmy Carter conceding defeat in the 1980 presidential election before the polls had closed on the West Coast, the VRS asked respondents that year what time of day they voted.
Despite the relatively limited range of questions about political participation asked in the VRS, it is the most important data set available for the study of who votes and who is registered. Some of the most authoritative studies of voting turnout have relied heavily on the VRS, beginning with the classic book *Who Votes?*, by Raymond Wolfinger and Steven Rosenstone (1980).

Because the VRS has such an enormous sample size and is paired with a larger survey that asks a number of sociodemographic questions, it is a good tool for helping us understand individual demographic circumstances that might get in the way of citizens participating in elections. As an example, consider the question of voting by the disabled, a topic covered by Lisa Schur and Douglas Kruse in Chapter 8. When a respondent to the VRS states that she or he did not vote in the most recent election, a follow-up question asks why. One of the response categories is “permanent illness or disability.” Because the CPS also asks respondents if they have one of six different disabilities, it is possible to gauge the number of eligible voters who fail to vote because of illness and disability. In 2010, this amounted to 40 percent of all disabled voters—or a total of more than 200,000 people. Furthermore, we also learn that among the disabled respondents reporting that they *did* vote, 27 percent said they voted by mail, compared with 17 percent of those who did not report a disability.

The VRS has shortcomings. It relies on respondents to recall their past behavior, and we know that recall about voting can be imprecise and inaccurate. Indeed, we know that people tend to overreport whether they voted because of a social desirability bias. Unfortunately, the VRS has never undertaken a “voter validation” study, in which researchers verify the vote history of respondents, so the precise nature of the overreporting bias is unknown in this study. The VRS does significantly overreport voting rates, although less than in other surveys. For instance, in 2012, 62 percent of VRS respondents said they had voted in the most recent election, when in fact only 58 percent of the voting eligible population had voted in that year; the corresponding rates were 54 percent and 41 percent in 2010. Thus, overall participation rates

---

12 The disabilities that are asked about are difficulty dressing or bathing, deafness or serious hearing problems, blindness, difficulty doing errands, difficulty walking, and difficulty remembering or making decisions.

13 Unless otherwise noted, the turnout rates used in this book are taken from the United States Election Project Web site (http://elections.gmu.edu/index.html), maintained at George Mason University by Michael McDonald.
will be elevated when we use the VRS. Still, the VRS is an invaluable resource for the comparison of demographic groups or geographic areas, where the overreporting problems presumably wash out.

*The Election Administration and Voting Survey*

The EAVS was begun by the EAC in 2004 as the only national census of information related to election administration in the United States. A biennial project, the EAVS collects information from approximately 4,600 local jurisdictions that conduct elections in the United States – mostly counties but also municipalities in New England and the upper Midwest, where counties play a secondary (or no) role in conducting elections.\(^{14}\)

The EAVS arises out of the EAC’s mandate to produce periodic studies of election administration issues. The EAVS is actually an amalgam of three surveys. One section deals with voter registration statistics and is used to produce the EAC’s biennial report to Congress under the National Voter Registration Act (NVRA) of 1993. Another section focuses on ballots sent out and returned from overseas and military voters, for the production of the biennial report mandated under the UOCAVA. Finally, sections on domestic civilian absentee ballots, election administration, provisional ballots, and Election Day activities fill out the rest of the survey. The complete survey contains more than 400 separate items, though most items only pertain to a subset of local jurisdictions.

Without the EAVS, we would not know how many voters are registered nationwide; how many registration forms were processed (or rejected) in a two-year period; how many absentee ballots were sent to voters, returned, and accepted for counting (or rejected); how many polling places there were on Election Day and how many people staffed them; how many provisional ballots were handed out and processed (including how many rejected); and which types of voting technologies were being used by local governments.

Indeed, the volume of information in the EAVS has overwhelmed the election administration community, which has not leaped to the task of using the data contained in it. The surprising lack of eagerness among

\(^{14}\) By focusing on counties as the jurisdictions of interest, the EAVS collects data from approximately 4,600 administrative units. If one instead focuses on municipalities rather than counties in states such as Michigan and Wisconsin, where election administration is often in the hands of city, village, or town officials, the number of jurisdictions is closer to 10,000.
the election policy community to use EAVS data in their research is no doubt attributable to the survey’s rocky start in 2004 and continued struggles with data gathering in the election administration domain, some of which are discussed in the following. Still, as many of the chapters in this volume demonstrate, the EAVS often is the only source of data we have to study important questions of election policy.

Consider, for instance, the simple controversy that arose in the 2012 general election in Fulton County (Atlanta), Georgia, when 11,000 provisional ballots – 2.8 percent of all ballots cast – remained on the table at the end of Election Day.15 Was this unusual or not? Without comparative information about the usage rates of provisional ballots in other counties throughout Georgia or throughout the country, it was impossible to know. With data from other counties, it was possible to put Fulton County’s experience in some perspective.

At the time the Fulton County controversy arose, it was too early in the vote-counting process to make good comparisons using 2012 election data, so the best one could do was compare Fulton County with other counties of similar size in the previous presidential election in 2008. Using the EAVS, we see that among the twenty-three counties with between 400,000 and 600,000 voters in 2008, only four had provisional ballot usage rates above 2.8 percent. Two of these counties were in Ohio (Franklin and Hamilton) and two were in California (Contra Costa and Sacramento), both states with very high provisional ballot usage rates overall. In 2008, California’s provisional ballot usage rate was 5.8 percent, while Ohio’s was 3.6 percent. In contrast, Georgia’s usage rate was 0.4 percent.

Only by reference to the EAVS was it possible to see that the number of provisional ballots issued in Fulton County in 2012 was unusually high, both in comparison to counties nationwide of a similar size and in comparison to the number of provisional ballots one could expect in Georgia as a whole.

If the EAVS is the only national census of election-administration-related statistics, what accounts for its underutilization in policy discussions, aside from the fact that it is not well known? Three things undermine the quality of the EAVS as a data source – or at least introduce challenges that must be kept in mind when using it to draw general

---

conclusions about elections in the United States. The first challenge is the high rate of missing data, especially from the earliest years of its administration. The missing value rates from 2004 are so high that the data from the first year of the EAVS are almost impossible to use for most applications. Things got much better in 2006 and 2008. According to Pew’s *Election Administration by the Numbers* report, among the most important data contained in the EAVS, 28 percent was missing in 2006, dropping to 15 percent in 2008. Conducting a similar calculation for 2010 and 2012 lowers the missing data rate further to 6 percent.

The presence of missing data from the EAVS also raises a shortcoming in the reports issued by the EAC using the EAVS data—the EAC makes no effort to impute missing data in those cases where a jurisdiction has not reported the required information. Furthermore, in calculating rates, the EAC sometimes uses a numerator that is based on data from one set of jurisdictions and a denominator that is based on data from another set. As a result, there is a significant absence of basic quantities reported by the EAVS and a significant number of rates that are inaccurately calculated.

Consider, for instance, the important question of how many absentee ballots are rejected because they are received after the deadline for counting. In 2008, the EAC reported that almost 75,000 absentee ballots were rejected for this reason, amounting to 18.4 percent of all rejected absentee ballots. However, eight states either did not report the number of late-returned absentee ballots at all or reported a value of precisely zero. Of the forty-three states and the District of Columbia that reported some count of the number of late-returned absentee, in fifteen states fewer counties reported the number of late-returned absentee ballots than reported the total number of absentee ballots returned; in two states, more counties reported the number of late-returned absentee ballots than reported how many absentee ballots were returned overall. In all, only twenty-three states were able to get every county to report both the number of absentee ballots rejected and the number rejected for missing the deadline. Among the states in which all jurisdictions reported both the number of rejected absentee ballots and the number rejected because they were late, the rejection rate for being late was an aggregate 36.3 percent—twice the rate reported by the EAC.

---


17 As further demonstration of the improving nature of the EAVS data, the EAC’s report on the 2012 election estimated that 32.1 percent of absentee ballots were rejected for being late, a figure much closer the aggregate number from 2008.
A second challenge facing the use of the EAVS is the lack of a common vocabulary across states when they refer to basic features of voting. A good example is the term early voting, which was used in some form by twenty-three states in the 2012 EAVS. At the same time, Kansas calls it “in-person advanced voting,” Ohio calls it “absentee-in person,” and North Carolina calls it “one-stop absentee voting.”

The EAC’s 2012 summary of what states mean when they label someone an “active voter” yields a dizzying array of definitions, which range from “a voter who is not inactive” (Alabama and Florida) to “registered to vote” (most states) to “legal resident of the state” (West Virginia) to simply “not defined by statute” (six states).

A final challenge in using the EAVS to assess the quality of elections in America pertains to the fact that in many states, the chief election officer does not have the authority to mandate that local officials report to them the data necessary to fill out the EAVS survey instrument, or even to comply with national election laws, such as the NVRA. The uneasy tug-of-war that exists between state and local election officials (LEOs) is one of the little-known dramas in the management of elections across the United States. Most state election officials would probably like to receive better information from their local units about how they are performing. However, lacking the ability to sanction local units for not reporting information to the state, there is often little a secretary of state can do if a local election board simply refuses to report information beyond what is required to certify elections.

Despite these challenges, the EAVS remains the most important national data set documenting the activities of local election units in federal elections. States are slowly but surely adapting to the survey instrument. The data are more complete with each passing election. The mean state completion rate rose from 86.1 percent in 2008 to 94.2 percent in 2012. As long as the dislike among the secretaries of state for the EAC does not spill over into the demise of the EAVS, it will only improve its usefulness as time moves forward.

State and Local Election Boards

If there is one thing that citizens know about elections, it is that they are used to choose political leaders. Political scientists know elections for
another reason: they are the source of a lot of data. Data are a natural by-product of elections, of course, because elections involve counting up the number of marks made on paper (literally or metaphorically) and then figuring out which candidate or ballot proposition received the most marks for a given contest.

Before the 1890s, little election data was shared widely with the public. Newspapers were the primary vehicles for communicating election returns; state governments rarely disseminated accounts of election returns for even statewide office, broken down by town or county. This all changed in the 1880s and 1890s, as a series of reforms, most notably the Australian ballot, swept the American states. The key feature of the Australian, or secret, ballot is that the ballot papers are printed by the government, not by the political parties. So the government would know which candidates to put on the ballot, it was necessary that access to the ballot be regulated and, in turn, for the activity of political parties to be regulated. As a result of the interest in regulating elections more directly, a related reform in most states was the requirement that the state also begin publishing election returns and related election statistics – such as the number of eligible voters or poll taxpayers – broken down at a finer level of disaggregation. States also needed to determine who was eligible to vote. Around this time official voter registration standards became more widely established.

With such statistics regularly reported, it became possible for average citizens to monitor the conduct of elections in their states and across states. To this day, many state libraries stockpile detailed election reports from other states. Stuffing ballot boxes became more obvious, as it was now easy to compare the total ballots cast for an office and the number of eligible voters in a precinct. Falsifying returns became harder, as it became easier to compare election returns in a precinct across time.

Fast-forwarding to the present, state and local election departments continue to spew forth incredibly detailed election reports after each election. With the spread of recordkeeping technologies such as desktop computers and spreadsheet programs, it has become easy to not only gather election reports quickly, but also to disseminate reports worldwide. States such as Idaho – small in population and far from the nation’s industrial centers – post spreadsheets online that provide detailed election returns and turnout statistics for each of their precincts. Many large states with greater capacity, such as North Carolina, provide public sites for the download of similar data for anyone around the world who would use it.
The explosion of the public reporting of election-related statistics by states and localities was begun over a century ago for the purpose of allowing the public to monitor the conduct of elections and to defer fraud in election administration, under the premise that many eyes make for cleaner elections. Nowadays, much of these data remain unused, except by political parties, candidates, and consultants who are looking to find pockets of potential voters to exploit. Yet the older performance-related use of these data is still very much a possibility.

One notable way that data generated by states and localities during elections can be used to assess how well elections are conducted is through the calculation of the “residual vote rate.” The residual vote rate is the centerpiece of Chapter 9, by Charles Stewart, which examines the performance of election machines.

One way to define the residual vote rate for a particular contest is the following:

\[
\text{Residual vote rate} = \frac{\text{Overvotes} + \text{Undervotes}}{\text{Turnout}},
\]

where the denominator is based on total turnout in the jurisdiction. An over-vote occurs when a person casts too many votes for an office; an under-vote occurs when a person does not cast the number of votes allowed for an office. Most often, the residual vote rate is calculated for the presidency, at the state level, but in principle it could be calculated for any electoral contest at any unit of aggregation.

As discussed in Chapter 9, the residual vote rate was pioneered following the 2000 presidential election as a way of identifying places where problems with voting machines might lead to confusion or outright malfunctioning. Through the use of the residual vote rate, it was possible to document the degree of improvement brought about by retiring old voting machines and replacing them with new ones in recent years.

The residual vote rate is currently the most frequently used measure that comes from official election sources, but one could imagine other uses as well. For instance, it is possible to use significant discontinuities in the support for political parties from one election to the next to spot places where there may be partisan tampering with election returns or voting machines (Myagkov, Ordeshook, and Shakin 2009).

Election returns are useful for studying the quality of election administration only if the data are reliable and reported. The residual
vote rate, for instance, cannot be calculated if an electoral unit does not report how many people turn out to vote, which remains the case in three states. Nor is the residual vote rate helpful if an electoral unit counts write-in votes but does not publish them. More broadly, electoral data are the most useful if they are reported at fine levels of disaggregation—hopefully at the precinct level—and if they are reported for all modes of voting (Election Day, absentee, or early voting). The state that is the model of electoral data availability is North Carolina, which provides highly disaggregated election returns for all state elections on a relatively easy-to-use Web site.

*Academic and Commercial Survey Research*

Finally, academic and commercial public opinion surveys occasionally contain questions that gauge directly the experience of voters with the apparatus of elections. Election administration per se has rarely been the subject of public opinion surveys, although there has been greater interest in the topic since the 2000 presidential election. Since that election, for instance, public opinion surveys have begun asking respondents whether they are confident their votes will be counted (or were counted) fairly in the election. Analysis of such questions is the focus of Paul Gronke’s analysis in Chapter 10 about whether Americans regard elections as being run fairly.

The most extensive public opinion project focusing on election administration is the Survey of the Performance of American Elections (SPAE), which was conducted on a nationwide basis in 2008 and again in 2012, but was also conducted in two odd-year gubernatorial elections and in the 2008 “Super Tuesday” presidential primary. Using the SPAE, it is possible to get a more representative view of how Americans view their own elections experiences than press accounts of the experience provide. The overall survey results are often quite at odds with the view of Election Day often conveyed by the press, where disorder and chaos are rampant. In 2012, for instance, despite the fact that there were reports of long lines at many polling places and continued distrust about the performance of electronic voting machines, 78 percent of respondents said their polling place was run “very well” (the top response category) and did not see any problems when they went to vote. Less than 2 percent of respondents stated they saw either minor or major problems that interfered with the ability of people to vote.
The Measure of American Elections

The SPAE is nonetheless a useful tool for identifying some problems that arise when people vote, such as the presence of long lines. This provides answers that are not as uniformly positive as the verdict on how well polling places are run. In 2012, 13 percent of American waited in line for more than half an hour to vote, for instance. This metric about long lines is revisited in the chapters that follow, both in Chapter 7, by Robert Stein and Greg Vonnahme, and in Chapter 9, by Stewart.

Summary: Data Used in Measuring American Elections

As this overview of data sources indicates, if we want to provide a portrait of election administration issues in the United States, we have a lot to work with. Because we strove in this volume to make the analysis from chapter to chapter as comparable as possible, we focus on data measured at the state level, although we could push the analysis down a further level and examine counties or even precincts.

HOW TO EVALUATE STATES

Although the science of assessing election performance is new, we are able to draw on massive literatures on measurement theory to evaluate the data considered in this book. Scholarship on measurement provides valuable guidance on how to judge the quality of individual metrics and how best to pool those items into scales and indexes to determine how well they tap into underlying dimensions. We will not devote space in this chapter to reviewing decades of research on measurement, but it serves as an informative backdrop for what follows. More recent work in the social sciences provides a specific framework for how to aggregate disparate measures into a common dimension of evaluation, a central issue for those trying to measure election performance.

Most of the chapters in this book analyze data on the states. This focus is in line with Gerken’s vision of a “democracy index,” but it also recognizes the fundamental autonomy that the U.S. Constitution bestows on state election practices and reflects where the most abundant data are being collected. It also provides a common denominator for connecting measures across chapters. State data are also more

---

20 For example, see Gerring (2012), King, Keohane, and Verba (1994), and Zeller and Carmines (1980).
likely to be collected over multiple years and generally have higher levels of reliability. Yet our state focus should not preclude collection of measures at a lower level of analysis. Many states exhibit tremendous variation across counties and municipalities within their jurisdictions. The stark differences between large urban centers and less populated rural areas within a state are often dramatic. Indeed, several chapters make brief excursions to examine data at a lower level of analysis before returning to the states.

We should note that it is generally preferable to draw data from disparate places rather than a single source. Any one source of information may have idiosyncrasies owing to its methods or purposes. Multiple measures drawn from the same source will often appear to go together because of common error variance rather than true connections among variables. Creating a portfolio of indicators with differing origins helps balance against the possibility of a systematic bias tainting the collective assessment.

In providing a data-driven assessment of election administration and policy in the United States, we can choose to set our sights broadly or narrowly. We can assess the current state of American elections as a whole, such as when a college calculates the cumulative grade point average for a graduating senior, or we can assess individual aspects of elections at a fine level, such as when a college professor assigns a grade to a student’s paper. The chapters that follow take a middle ground, by analyzing a limited set of measures associated with particular aspects of American elections, such as absentee voting and provisional ballots. Here we offer some thoughts about assessing elections at a more general level.

Given the fledging nature of election performance data, it might be tempting to gather all the measures one can find and then add them up to produce an overall assessment of the states. This has been done in some other policy areas by a variety of advocacy groups. But this approach violates best practices in the social sciences. Rather than falling prone to a haphazard search for measures, we follow scholarly practice to determine whether a global evaluation is even possible. We begin with the idea of election performance, then distill this notion into four concepts, and finally collect and evaluate a series of measures thought to tap these concepts. Only then do we consider the possibility...
of combining the measures into an aggregate metric using various methodological techniques. This approach avoids the concerns raised by Alvarez, Atkeson, and Hall (2012) about jumping immediately to an overall index that fails to provide useful diagnostic information to election administrators seeking improvements in performance. Regardless of how much aggregation occurs in the production of a broad index, we continue to emphasize the value of individual measures.

We adopt Munck and Verkuilen’s (2002) analogy of measurement as parts of a tree. They describe the broad idea as the “trunk” of the tree and its major conceptual parts as “branches.” The specific measures of these concepts are “leaves.” Compared with branches, leaves are more numerous and more tangible. That is, as one shifts from the trunk to branches to leaves, the level of abstraction decreases and the number of components increases, just as a tree becomes more differentiated as one moves from the ground upward to the tips of its branches. Once the leaves are selected and scored, they can be reaggregated, or “rolled up,” to generate higher-level metrics.

Munck and Verkuilen apply this approach to the operationalization of “democracy” at the national level. This is the tree trunk. They then apply Dahl’s (1971) distinction between “contestation” and “participation” as the two major concept branches. Each concept then implies several measures, or leaves. For example, the level of “contestation” is evaluated with measures of freedom of the press and the right to form political parties.

We decompose the broad idea of election administration performance into four concepts. These are Gerken’s (2009a) three dimensions of registering, voting, and counting, as well as a fourth dimension, data transparency. This framework implies that a state’s election performance is exemplary if it scores well in registering those who want and are eligible to vote, facilitating voting by those who are eligible and wish to do so, tabulating ballots accurately, and providing thorough public access to data and other election information. A poorly performing state would fare badly in each domain. The more difficult cases occur when one wishes to assign an evaluation to a state that does well in some domains but less well in others. In some settings we go further and divide each of these concepts according to whether they emphasize voter access or ballot security, the two concerns most commonly highlighted in policy discussions about election reform.

After gathering measures for each of these concepts, it is crucial to assess their reliability and validity. This is the task of the chapters that
follow. What they do not attempt is a complete aggregation of measures (although the authors do investigate connections among measures within a topic area). Aggregation is a more difficult process than is commonly assumed. As we have noted, simply adding up the measures to create a single index can be problematic. It does not reflect the distinctions between categories, it overlooks the relationships among measures, and it forces all the measures to be equally important. Working downward to the leaves and then back up to the idea permits the researcher to test whether an undimensional factor underlies the measures, rather than beginning with the assumption that a single “democracy index” metric would adequately reflect the full range of election-related things that states do.

A Look at Two States

To illustrate how the approach works in practice, we compare two midwestern states, Indiana and Minnesota. Leaving aside questions of reliability and validity for individual chapters, what do measures say about how these states perform in terms of registering, voting, counting, and data access? Can these leaves be reaggregated into a reasonable summary indicator for each state?

Before examining the data, we need to highlight the issue of missing data. If one wishes to examine the raw values of the measures, the data largely speak for themselves. In contrast, if one wishes to rank or score states, missing data create a methodological fork in the road. We could assume that the data are missing completely at random, in which case the states with data could be ranked or assigned to percentiles. The percentile approach is common in other domains, such as educational testing. A downside of having at most fifty-one units (fifty states plus the District of Columbia) is that it is impossible for a state to be in the first percentile no matter how well it performs. If the analyst wants to penalize states for missing data, they could be ranked below all those with data. But in addition to artificially compressing the range of scores, the problem of ranking all states with missing data on a particular indicator below all states without missing data is that this strategy imports the data transparency concept into every substantive measure of election performance. This muddies the interpretation of an analysis of states that seeks to understand how well they perform particular election-related tasks. To avoid imposing any assumptions about missing data now, we simply present the
raw values for each measure and the state’s rank among those states reporting data.

Table 1.1 reports nineteen measures of election performance, divided into the four categories. Before discussing how Indiana and Minnesota stack up on all nineteen of these measures, we should say a word about what they are and how they were constructed.

Between 2010 and 2012, the Election Initiatives project of the Pew Center on the States conducted a series of meetings with an advisory group, composed of two dozen election officials and academics from across the country, to explore the possibility of releasing an Elections Performance Index (EPI) building off the ideas first articulated by Gerken (2009a). To make these deliberations tangible, the advisory group discussed a long list of measures – nearly fifty in all – that could potentially be included in such a performance index. The nineteen measures reported in Table 1.1 were discussed at an intermediate stage of the advisory group’s deliberations. They are greater in number than the indicators that eventually were reflected in Pew’s EPI, which was released in early 2013. However, the measures in Table 1.1 all provide interesting insights into election administration and policy at the state level, even in cases where they were not included in Pew’s 2012 EPI.

What do the data say about Indiana and Minnesota? Starting with registration, it appears that Minnesota is performing better than Indiana. It ranks first out of all states in terms of the registration rate. Indiana’s registration rate, which is almost nine points lower, puts it forty-second among the states. The gap is somewhat narrower when it comes to the rejection rate for absentee ballots, although Minnesota’s tiny value of 0.3 percent puts it ahead again. Here is where data completeness affects a substantive measure; only forty-four states provided this information to the EAC. Lastly, Minnesota also comes out ahead when looking at the percentage of people who report being deterred from voting because of registration problems. Minnesota is well above the median state on this measure, while Indiana is just below it.


22 The measures are calculated as in the 2012 EPI. For EAVS-based data, a state is only considered to have an entry on a particular measure if it is reported by at least 85 percent of its local jurisdictions (weighted by voter registration). Contributors to this volume frequently use a different standard for what is considered “complete” and may compute rates such as absentee ballot rejections based on different denominators. A more detailed discussion of the measures is included in the Appendix to the book.
## Table 1.1. Comparing two states in 2012

<table>
<thead>
<tr>
<th></th>
<th>Indiana</th>
<th></th>
<th>Minnesota</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank</td>
<td>Value</td>
<td>Rank</td>
<td>Value</td>
</tr>
<tr>
<td><strong>Registering</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registration</td>
<td>42/50</td>
<td>78.4%</td>
<td>8/50</td>
<td>87.7%</td>
</tr>
<tr>
<td>New registrations rejected</td>
<td>28/28</td>
<td>71.7%</td>
<td>2/51</td>
<td>0.01%</td>
</tr>
<tr>
<td>Provisional ballots cast</td>
<td>15/48</td>
<td>.2%</td>
<td>1/45</td>
<td>0.0%</td>
</tr>
<tr>
<td>Provisional ballots rejected</td>
<td>14/40</td>
<td>.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registration a reason for not voting</td>
<td>28/51</td>
<td>6.5%</td>
<td>2/51</td>
<td>1.2%</td>
</tr>
<tr>
<td>Online registration available</td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Voting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnout</td>
<td>42/51</td>
<td>55.1%</td>
<td>1/51</td>
<td>75.7%</td>
</tr>
<tr>
<td>Absentee ballot rejections</td>
<td>39/44</td>
<td>.5%</td>
<td>35/44</td>
<td>.3%</td>
</tr>
<tr>
<td>Absentee ballot nonreturns</td>
<td>2/45</td>
<td>7%</td>
<td>11/45</td>
<td>4.3%</td>
</tr>
<tr>
<td>UOCAVA rejections</td>
<td>40/46</td>
<td>20.6%</td>
<td>34/46</td>
<td>7.6%</td>
</tr>
<tr>
<td>UOCAVA nonreturns</td>
<td>3/47</td>
<td>13.5%</td>
<td>15/47</td>
<td>21.1%</td>
</tr>
<tr>
<td>Disability a reason for not voting</td>
<td>10/51</td>
<td>17.9%</td>
<td>24/51</td>
<td>14.7%</td>
</tr>
<tr>
<td>Polling place a reason for not voting</td>
<td>35/49</td>
<td>3.4%</td>
<td>3/49</td>
<td>7%</td>
</tr>
<tr>
<td>Time waiting to vote</td>
<td>41/51</td>
<td>15 min.</td>
<td>14/51</td>
<td>6 min.</td>
</tr>
<tr>
<td><strong>Counting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postelection audits</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Residual vote rate</td>
<td>39/48</td>
<td>1.5%</td>
<td>6/48</td>
<td>0.5%</td>
</tr>
<tr>
<td>Confident vote was counted</td>
<td>19/51</td>
<td>91.9%</td>
<td>20/51</td>
<td>90.5%</td>
</tr>
<tr>
<td><strong>Data access</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EAVS completeness</td>
<td>23/51</td>
<td>99.5%</td>
<td>1/51(^b)</td>
<td>100%</td>
</tr>
<tr>
<td>Five online lookup tools</td>
<td>1/51(^c)</td>
<td>5</td>
<td>1/51(^d)</td>
<td>4(^e)</td>
</tr>
</tbody>
</table>

Notes:
- Data are not available for that measure.
- Four states are tied at this value.
- Fourteen states tied at this value.
- Eight states tied at this value.
- Seventeen states tied at this value.
- Minnesota does not have provisional ballots. Therefore, whether the state has a feature on its Web site allowing a voter to check on the status of his or her provisional ballot is irrelevant. Minnesota is ranked at the top of this measure because it has 100 percent of the lookup tools relevant to it.
- States are not ranked on the dichotomous postelection audits and online registration measures.
- The denominator of each ranking fraction is the number of states for which data are available.
- “Registration Rate” is based on citizens in the CPS.
- “Registration a Reason for Not Voting” is the average over the 2004, 2008, and 2012 elections.
In terms of voting, Minnesota also beats Indiana on most measures. Most obviously, the 2012 turnout rate was almost twenty-one points higher in the Gopher State, ranking it first in the country. Yet this higher turnout rate did not result in longer lines, which averaged a third of the time it took to vote in the Hoosier State. Smaller percentages of people in Minnesota reported that they failed to vote because of a disability or something problematic about the polling place. Although Indiana reports a lower nonreturn rate among absentee and UOCAVA ballots, Minnesota rejected smaller shares of those ballots that were returned.

For the counting category, Minnesota performs better on all three measures. The state conducts postelection audits, whereas Indiana does not. In terms of outcomes, Minnesota’s residual vote rate is one of the best in the country at 0.5 percent. Indiana’s is one of the worst at 1.5 percent. Voters in both states are quite confident that their ballots were counted as they intended, although Indiana ranks slightly higher.

Finally, the states are similar in terms of data completeness in the EAVS, with Indiana falling just short of 100 percent completion. Both states provide all the online lookup tools that are assessed and relevant to the state.23

The impression that any reasonable observer would take from these indicators is that Minnesota is performing better than Indiana, and in some cases by a substantial margin. It is true that Indiana is ahead in a few categories, so the measures are not unanimous in their verdict. But the collective judgment of the metrics strongly suggests that Minnesota is in a better position. If the “leaves” in the table were aggregated up to the four larger categories of registering, voting, counting, and data transparency, Minnesota should be ranked more highly in each. If the states were students in a class, it would be natural to assign Minnesota a better letter grade.

How to Aggregate

Not all state comparisons will be as clear-cut as the difference between Indiana and Minnesota. There are likely to be states that perform much better in one category than another. This makes the task of aggregating more difficult and the decisions made along the way more consequen-

---

23 Minnesota does not have a provisional ballot lookup tool on its state Web site. However, Minnesota is exempt from provisional ballot requirements of HAVA because it has Election Day registration (EDR).
Burden and Stewart

tial. Smart analysts will disagree about the placement of some states in any global evaluation. So how does one proceed?

The most agnostic approach is to report the raw, disaggregated measures as they are. Every state’s score on every indicator is listed. This approach is most transparent and imposes the fewest assumptions. It allows each state to have a unique “profile” just as each sound has a unique wavelength and amplitude, or an image has a fingerprint in the form of a color profile. By faithfully reflecting each of the various performance metrics, this approach also provides election administrators with task-specific feedback that will help them make surgical improvements to the most problematic aspects in their state. Officials in Indiana, for example, will learn that waiting times and registration problems deserve real attention, whereas absentee rejection rates are lesser concerns.

At the same time, this wide-open approach comes with liabilities. It does not apply any judgment about the quality of performance. Thinking of this task like evaluating a student, this approach reports the grade on each assignment without trying to render a final grade that summarizes the student’s overall performance throughout the course. Because it throws a bundle of scores at the states, it might be unwieldy. By not providing a more global assessment for each state or ranking them in any way, it fails to encourage accountability and competition as Gerken’s democracy index proposes to do. State officials might be able to avoid culpability for poor performance by directing attention toward a small number of measures on which they fare well. Reporting all the data also implicitly suggests that each measure is equally important. Ironically, this is one of the complaints about a summary index.

The solution to this messiness is to impose an aggregation rule, that is, to roll up the many leaves into a smaller set of concept branches. This necessarily involves the trade-off of adding assumptions about how measures should flock together and the relative weight that each should get. We believe that any aggregation exercise should also be accompanied by the raw data so that scholars and policy makers may benefit from learning about performance in specific domains and evaluate how the states would rank if a different aggregation rule were used.

The simplest way to combine individual measures of election performance into a small number of aggregated measures is simply to calculate the average value of each individual measure for each state. Although highly intuitive, this approach immediately comes up against the problem that the individual indicators are measured using different
scales. Consider, for instance, the registration-related measures in Table 1.1. Five of the six measures are percentages, while the sixth, whether a state allows online voter registration, is a dichotomous variable equal to 1 if the state allows this, 0 otherwise. Also note that each of the five percentage measures vary across different ranges. The registration rate, for instance, varies from a low of 70.9 percent (Hawaii) to a high of 92.4 percent (Minnesota). On the other hand, the percentage of new registrations rejected ranges from 0 percent (New Hampshire) to 71.7 percent (Indiana). The average of these two percentages produces nonsense; even more nonsensical would be taking an average of all six measures grouped together under “registering” in Table 1.1.

Faced with the prospect of combining disparate measures into a single index, scholars generally “normalize” the measures in an attempt to get them all on the same scale. Without relying on higher statistical techniques, such as the one discussed later in the book, three normalization methods are typically used in situations like this. First, one can simply do the aggregation based on the rank ordering, rather than on the raw values. The best-performing state on a particular measure is assigned a value of 1, the second-best-performing state assigned a value of 2, and so on, until the last state is assigned a value of 51 (counting the District of Columbia as a state). The intermediate scale is then an average of all these rankings.

Second, one can transform the raw measures into the percentage of the range, so that the lowest-ranking observation is set to 0, the highest-ranking observation is set to 1, and every other observation is assigned a value between 0 and 1, depending on where in the interval that observation lands. Third, one can transform the raw measures by first converting them to z-scores and then taking the average. A z-score first subtracts the raw score for a state from the average of all states, then divides this by the standard deviation across all states. This provides a handy normalization for comparing across measures.

A more sophisticated form of aggregation is to apply a data reduction technique in which the measures “speak” about how they go together. Exploratory factor analysis is a prime example of this. The exploratory version of factor analysis does not impose any constraints or other theoretical knowledge, instead asking the data to reveal what latent structure

---

24 More formally, if \( x \) is the raw score for a state, \( x_{\text{low}} \) is the minimum value among all states, \( x_{\text{high}} \) is the maximum value among all states, the transformed value is defined as \( (x - x_{\text{low}})/(x_{\text{high}} - x_{\text{low}}) \).
exists. The method uses correlations among variables to identify dimensions underlying the observed data. This has the benefit of generating empirical weights or “loadings” that indicate how important each variable is to a factor. It also provides direct evidence on the dimensionality of the latent data structure, as indicated by the eigenvalues, which indicate how much variance each factor explains.

In contrast, confirmatory factor analysis uses the researcher’s theoretical knowledge to guide the analysis. The researcher posits which factors organize the data and then estimates them in a more structured fashion. As in the exploratory version, the analysis itself provides evidence for how well items in a factor cohere and what weight or loading each should get. Ansolabehere, Rodden, and Snyder (2006) apply confirmatory factor analysis in their study of political polarization in the electorate. They use several surveys, such as the American National Election Study (ANES), to find questions about issues and then code them as either “economic” or “moral.” Returning to our tree analogy, one might say that ideology is the “trunk,” economic and moral issues are the “branches,” and the individual survey questions are the “leaves.” For example, the economic issues scale from the ANES includes items asking about spending on various programs, trade-offs between the environment and jobs, and feelings toward unions. Factor analysis of these items produces only one major factor (eigenvalue). The items all have similar weights, with loadings ranging from a low of .35 to a high of .67. More heterogeneous loadings or a greater number of factors would have produced a more complex interpretation of the data.

Collapsing leaves in this way is theoretically satisfying but does require making assertions about the importance of measures and how they should cohere. This is easier to justify the better developed the theory. In election administration, the state of theory is rapidly improving, although it remains uneven. Indeed, the purpose of this volume is to advance the science of measuring election performance. This exercise necessarily involves the interplay of evidence, experience, knowledge, and theory.

Applying a scaling approach to our 2012 data reveals that scale reliability is intermediate for the registering branch (α = .62) and the voting branch (α = .54) and much lower for the counting and data transparency branches (α = .36 and .35, respectively). The overall scale built

The measure of scale reliability we use here is Cronbach’s alpha. In performing this scaling exercise, we have to deal with two practical issues. The first is the fact that many of the EAVS-produced scales are highly skewed. To account for the skew, we perform the scaling
from the aggregation of these four branches has an overall reliability coefficient (Cronbach’s $\alpha$) of .40. The overall scale reliability is actually greater ($\alpha = .68$) if we skip the intermediate “branch” step and instead aggregate all the leaves in one step into an overall score. Over-time reliability among 2008, 2010, and 2012 is quite high within each of the four concepts and for the overall scale, regardless of whether it is computed in one step or two.

Based on this initial examination of data from 2008, 2010, and 2012, the measures work together to render a fairly consistent verdict about state election performance. However, the measures of reliability cited earlier are, at best, weaker than one expects in highly developed areas of the social sciences, such as psychometrics. There are good reasons to believe that the coherence among measures will increase as data quality improves and data access and transparency practices in the states catch up to recent expectations about data availability. Still, there will be cases where states’ rankings do not tell a unified story. A state might fare quite well in terms of registration but poorly in terms of counting. If these cases are common, one might refrain from emphasizing overall scores and instead focus on the four branches. An even less aggressive approach would rely only on leaves, which we believe should always be available regardless of the amount of data reduction taking place.

Nonetheless, there are good reasons to generate global evaluations, if we are cautious in interpreting them. Even the student whose scores on individual assignments and exams are uneven will still receive a final grade. That grade will be the instructor’s best summary of the performance over the year, including assumptions about how much each score should contribute to the final judgment. The student will learn from the many marks through the year where he or she excels and where improvements are necessary. By receiving a final grade, it is possible to compare across students, for students to compare their performance across subjects, and to track performance as students move from one grade level to the next. The grade point average, which may be considered a sort of “meta” evaluation, goes a step further.

In the end, this book emphasizes the leaves over the branches, or even the trunk. The main reason is that each of the chapters focuses on an aspect of election administration and policy that can be thought of as a self-contained whole. If a state wishes to improve its standing by first converting the raw scores to ranks. The second issue is missing data. To address this issue, we impute the average rank on a measure to replace missing values.
in how its election administration is regarded, the most likely successful strategy will be one that is focused on particular areas that need improvement, rather than scattershot across all areas. In this early phase of building a data-driven science of election administration, a lot is to be gained by shining a spotlight on different aspects of election administration, one at a time.

OVERVIEW OF THE BOOK’S CHAPTERS

Each of the chapters in this volume evaluates measures of an aspect of contemporary election administration. Chapters largely fall within one of the four broad categories of registering, voting, counting, and data transparency. These categories make intuitive sense and Gerken has provided a sound rationale for the first three being distinct. They also capture the essential steps in the workflow of election administration. There are sure to be cases where indicators cross these boundaries, reflecting the fact that performance in one domain often has consequences for another domain. For example, more generous use of absentee and UOCAVA ballots (voting) tends to result in higher rejection rates for those ballots (counting). In addition to assessing the reliability and validity of specific measures, the authors get to the meaning of those measures. They reveal the ways in which measures result from such things as a state’s demographic profile or election laws. The authors show how multiple measures may be used in tandem to tell a coherent story about the electoral ecosystem in a state or locality. Here we review them according to whether they focus on registering, voting, counting, or data transparency.

Chapter 2 (“Registration and Voting: A View from the Top”), by Barry C. Burden, launches the empirical portion of this book by examining the most fundamental statistics in the study of election administration and policy making: registration and turnout rates. This is an appropriate starting point, not only because registration is a precondition to voting (except in North Dakota) and the turnout rate is such an easily grasped measure of the health of a democracy, but also because there is so much more to these measures than first appears.

To study registration and turnout rates, we first need to agree to how to measure these quantities. Without digging deeply into the manner in which states report registration and turnout data, academics, policy makers, and courts have too easily taken at face value the official reports issued by state and local election departments. This is especially
problematic in studying registration. In the management of the states’

evoter registration rolls, America’s dynamic population encounters a

combination of bureaucratic inertia, antiquated administrative proce-
dures, and provisions of the NVRA that conspires to produce registra-
tion lists that are bloated by the presence of deadwood. On top of this,

states define voter registration in different ways, and the science of esti-
mating the number of eligible voters – which forms the denominator

when we calculate the registration rate – is still evolving. Thus, compar-
ing a 75 percent registration rate in one state with an 80 percent rate in

another can be an exercise of comparing apples to turkey basters.

Burden’s analysis alerts the reader to nettlesome measurement issues

attendant even in the most basic indicator of electoral performance. He

makes a strong counterintuitive argument that the most valid and

reliable measure of the registration rate in the United States is based on

survey research, not on official registration records. Having settled on

a firm strategy for estimating how many eligible Americans are actu-

dally registered to vote, and how many vote, Burden provides insightful

analysis into the possibilities and limitations for using policy choice to

increase turnout and registration. In particular, demographic charac-

teristics of a state and its policy choices have little effect on registration

rates but a considerable effect on turnout.

In Chapter 3 (“Voter Registration: The Process and Quality of Lists”),

Stephen Ansolabehere and Eitan Hersh pick up on the registration list

themes initiated by Burden and engage in a thorough exploration of

state voter registration lists with the assistance of a unique resource – an

aggregated nationwide list of voting-age Americans constructed by the

political consulting firm Catalist. Their chapter provides a nationwide

view of the accuracy of information contained in voter lists. It is a novel

approach to the study of voter registration lists and is one of the first

academic studies of voter registration that takes advantage of the “big

data” revolution in American campaigns (Issenberg 2012). Beyond tak-
ing advantage of a remarkable data set, Ansolabehere and Hersh pro-

vide a useful taxonomy of registration lists based on their functions (to

map voters onto precincts, authenticate voters at the polls, audit elec-
tion results, and prevent in-person voter fraud) and a series of measures

that help to illustrate how well the states’ lists may be suited to these

tasks. Finally, they make a strong case for paying attention to the com-

plexities that attend the management of registration rolls and caution

against generalizing across functions in assessing how well a particular

state’s voter registration list is suited to the tasks at hand.
In Chapter 4 (“Provisional Ballots”), Michael J. Hanmer and Paul S. Herrnson address an issue that has vaulted to prominence since the 2000 election: provisional ballots. A provisional ballot is issued to a person who tries to vote but is not listed on registration lists or whose eligibility is otherwise questioned by poll workers. The provisional ballots must be then verified after the election when a voter returns to show identification. Although mandated at the federal level by the Help America Vote Act (HAVA), implementation has varied significantly across the states. A high usage rate, for example, has the benefit of providing a fail-safe option for would-be voters and preventing fraud, yet it also indicates problems with the registration files provided to poll workers. They argue that the usage and acceptance rates be considered jointly; a rise in use generally translates to less frequent acceptance. They also recommend that data be collected on the reasons why provisional ballots were issued and why some were not counted.

Despite the nuances of provisional ballot activities, state-by-state data from the EAVS show that a few states are out of line with national norms. Whereas most states have usage rates below 1 percent, a handful of states give provisional ballots to between 4 percent and 6 percent of voters, and a few go even higher. Perhaps most concerning are states that do not conform to the general positive relationship between usage and acceptance. Both usage and acceptance rates are quite reliable over time, suggesting that state-specific factors are behind them.

Hanmer and Herrnson conclude by showing that key factors that drive these patterns are whether states allow a provisional ballot to be counted if cast in the wrong precinct and whether those who request but do not cast absentee ballots must vote provisionally. These two provisions account for more than a percentage point difference in usage rates and twenty- to thirty-point differences in acceptance rates. In addition, states that used provisional ballots before HAVA mandated that all states use them issued more and accepted more for counting. This suggests that election officials in states that adopted the provisional ballot mechanisms because of the HAVA mandate are responding somewhat haltingly to this policy change.

We learn from Christopher B. Mann’s Chapter 5 (“Mail Ballots in the United States: Policy and Administrative Challenges”) that all absentee voting is not created equal. He argues that measures of absentee ballot usage are highly reliable and valid, but that rejection rates and the reasons for these rejections show surprisingly high volatility across the 2008, 2010, and 2012 elections. These fluctuations can only
be understood in the context of which policy “regime” produces them. The four regimes Mann identifies are (1) absentee voting by excuse, (2) voting by mail without excuse, (3) permanent voting by mail, and (4) postal voting. These policy differences allow us to understand state differences: Wisconsin had twice the percentage of mail ballots as in Minnesota because Wisconsin belongs to the second regime and Minnesota belongs to the first. Useful comparisons for election administrators are thus among states within the same regime. One would not expect the universal use of mail voting in Oregon to hold in Pennsylvania, where voting is largely constrained to take place at traditional polling places. Acknowledging the regimes also helps make sense of disparate statistics. For example, the nonreturn rate for absentee ballots in 2010 was a mere 0.5 percent in Idaho but 30.8 percent in Utah. Once we know that Utah allows voters to put themselves on a permanent absentee ballot list, but Idaho does not, it is easy to see how this difference has come about: many people who register to vote in presidential elections continue to receive ballots in midterm elections in Utah without returning them.

While they also vote using absentee ballots, overseas voters who are covered by the UOCAVA present a distinct set of challenges to election administration and policy, which are explored in Thad Hall’s Chapter 6 (“Voting from Abroad: Evaluating UOCAVA Voting”). Overseas military voters played a bit part in the 2000 election drama in Florida, but they have taken on a bigger role in national policy making since then. Indeed, aside from HAVA, the most important election reform at the national level since the 2000 election was the Military and Overseas Voting Enfranchisement (MOVE) Act of 2009, which was aimed at streamlining the process of getting ballots delivered to overseas voters in time for them to be returned and counted. Hall assesses the degree to which states and localities vary in how they handle requests for UOCAVA ballots and in how successful they are in getting those ballots counted. He documents the continuing difficulties that overseas voters have in returning their ballots, even in the wake of the MOVE Act. He also demonstrates that states can take steps to increase the successful participation of overseas voters even further, by liberalizing their ballot transit rules and (perhaps) liberalizing their registration facility rules.

In Chapter 7 (“Polling Place Practices and the Voting Experience”), Robert M. Stein and Greg Vonnahme report on research that establishes the relationship between where and when voters cast their ballots, on the one hand, and how satisfied they are with the experience,
on the other hand. The past decade and a half has witnessed a revolution in the modalities of American voting. Rather than an election day we have an election period. This development has not gone completely unnoticed – witness the attention to early voting in the 2008 and 2012 elections as auguries for who would win the election – but the subject of whether early voting and easier access to absentee ballots leads the electorate to be more confident in the outcome has been largely overlooked in the rush to make voting more convenient. Furthermore, as Stein and Vonnahme make clear, once a state has decided to go all in with early voting, there are better and worse ways to do it, if the goals are to improve the voter experience and voter confidence.

In expanding their prior research, which had been focused on just two states, Stein and Vonnahme show that voter experiences are more positive in states in which voters have greater opportunities to vote before Election Day and when they have opportunities to vote in places they might frequent in their everyday lives, such as shopping centers. Voters who avail themselves of voting early do report a better experience than those who show up on Election Day, and they are willing to trade off waiting longer to do so in return for the greater flexibility in choosing the most convenient day on which to vote.

Lisa Schur and Douglas Kruse provide an insightful look into the experiences of disabled voters in Chapter 8 (“Disability and Election Policies and Practices”). Frequently overlooked in standard histories of the HAVA is the priority placed on increasing access to the ballot among disabled voters, particularly through HAVA’s requirement that all precincts have voting systems that are “accessible for individuals with disabilities, including nonvisual accessibility for the blind and visually impaired, in a manner that provides the same opportunity for access and participation (including privacy and independence) as for other voters” (HAVA Sec. 301(a)(3)(A)). Despite mandates for accessibility, Schur and Kruse demonstrate that turnout rates for disabled voters continue to be below those for nondisabled voters, especially among those with cognitive impairments and difficulty going outside alone. They provide compelling evidence that turnout among the disabled is increased when absentee balloting is more readily available to all voters, indicating that vote-by-mail systems can help reduce the turnout gap between people with and without disabilities.

A key measure of vote tabulation is the residual vote rate, examined in Stewart’s Chapter 9 (“The Performance of Election Machines and the Decline of Residual Votes in the United States”). Developed by
researchers after the 2000 election, the residual vote rate is the percentage of ballots cast that are not counted. In 2012, the nationwide residual vote rate was about 1 percent, suggesting that 1 out of every 100 votes was lost. Stewart shows that the residual vote rate is a valid measure of technological problems, such as poor ballot design and flawed counting procedures. This is clear from analyses showing higher rates for punch card machines and absentee ballots compared with lower rates for optical scan machines, where voters are often given the opportunity to correct over-votes and under-votes. The shift of states away from punch cards and lever machines toward optical scan and direct recording electronic (DRE) voting machines has contributed to a declining residual vote rate over time. Yet in the 2012 election, the rate varied significantly across states, from trivial levels in DC, Nevada, and Wisconsin to rates above 2 percent in states including Kansas and Missouri, all the way up to 3.5 percent in Nebraska.

Voting technology is behind much of this variation. Voting times also correlate with technologies, as the residual vote rate might suggest, with paper ballots being faster and touch screens taking longer. Shifts from one technology to another create more problems as election staff and voters adapt to them. By drawing on more granular data from Florida and Massachusetts, Stewart demonstrates that even within a state, the types of equipment and preparation in setting them up affect the residual vote rate. Stewart concludes by pointing to what appears to be a practical lower bound on the residual vote rate of about 0.5 percent, thus providing a reasonable standard for judging residual vote rates generally.

In his analysis in Chapter 10 (“Voter Confidence as a Metric of Election Performance”), Paul Gronke takes on an important metric that has emerged to assess elections across the country – answers to the survey question asking whether people are confident their votes were counted as they intended. A survey question such as this could serve as a convenient summary statistic of the electoral process, much like the “presidential approval” question helps summarize citizens’ assessments of the president, but only if it can be deemed reliable and valid as discussed in this chapter. Gronke casts a skeptical eye on this question, but he ends up acknowledging its usefulness for guiding policy makers toward improving elections in the eyes of voters.

The usefulness of the idea of voter confidence is undermined somewhat by the tendency of voters to approach the question through the lens of loser’s remorse or winner’s triumph. Voters for winning candidates are
more confident their vote was counted as cast than voters for losing candidates. At the same time, Gronke demonstrates that there are lessons for the LEO who wishes to improve the confidence of voters in the administration of elections. For instance, voters who encounter competent poll workers and machines that work, and who wait in short lines, have a better feeling about their vote being counted than voters who face surly poll workers, broken machines, and long lines. Election officials cannot (and should not) be in the business of ensuring that every voter votes for a winner; however, they can be in the business of improving election administration, which has demonstrable effects on boosting confidence among the electorate in the legitimacy of the outcome.

In Chapter 11 (“Election Data Transparency”), Lonna Rae Atkeson examines the rapidly changing environments of accessibility and transparency. She contends that an open election system boosts public confidence and provides valuable information for a range of stakeholders. Unlike other domains of election administration, where policies are a matter of debate, there is little disagreement that more information is a good thing. Indeed, one goal behind the creation of the EAC was to promote sharing of data. States have improved their reporting to the point where more than 90 percent of the most important items in the EAVS was complete in the 2012 survey. Yet the validity of these data is open to question; it is often ambiguous whether a missing entry means that a state did not collect it or could not calculate it, or if the item is not applicable. And when the full EAVS is evaluated, completion rates drop to between 70 percent and 86 percent, depending on the section. Overall completeness in 2012 ranged from the District of Columbia’s nearly perfect record to New York, which provided only a bit above 40 percent of the data requested in the EAVS.

The other component of Atkeson’s analysis is the availability of online information for the public. Drawing on the Pew study of state election Web sites and an original study of Internet searches, she finds that states with higher levels of EAVS data completeness also made it easier to find voter information using Google searches. Importantly, she discovers that neither completeness nor online information is a product of state characteristics or demographics. Rather, both appear to be the result of “time and resources” that state officials devote to their online presences. Yet she finds modest correlations among her usability and search measures, the availability of online tools, and data completeness. These are areas where improvements are happening most quickly and dramatically.
CONCLUSION: COMMON THEMES

Each chapter in this volume offers a unique portrayal of a different aspect of election performance, yet when looking across the chapters this volume, several common themes emerge.

First, and most positively, the quality of election data has now developed to a point where it is possible to conduct systematic evaluation of election performance. While the research and policy communities should continue to press for more and better data, recent advances allow us to understand what is happening – and why – in a way that was unimaginable just a decade ago. Having more complete data on a wider range of measures makes evident how much the states vary and why. As noted at the outset of this chapter, this permits valuable insights into which state and local innovations are successful at improving elections. Although a state will often perform relatively better on some measures than others, it is safe to conclude generally that some states are doing better than others across a range of measures. States that succeed in one area tend to thrive in others areas.

Second, the measures often confirm folk wisdom about particular election and states, but the data also turn up some surprises. The regularities are reassuring. For example, changes in several indicators between the 2008 presidential election and the 2010 midterm elections could have been predicted in advance. In the more intense presidential campaign environment, absentee ballots are more likely to be returned and turnout increased. Simultaneously, more voters have trouble finding their polling places and a large share of absentee ballots are rejected. These swings reflect well-known cycles in these two rather different electoral environments.

But the data also reveal the unexpected. Common sense would expect more people to say their voting was hindered by disability or illness in states where health outcomes are poorer. Indeed, Alabama, Kentucky, and West Virginia rank near the bottom in terms of state health outcomes and display some of the highest disability and illness problems in voting. What is surprising is that Connecticut and Massachusetts have similarly large percentages of voters citing disability and illness problems, yet these states have two of the best health outcomes in the country. Clearly more is at work than simple demographics, such as the accessibility of older public buildings in the Northeast. Likewise, it is

26 See www.americashealthrankings.org.
not surprising that “battleground” states, such as Florida and Virginia, saw longer lines at the polls in 2012. But the intensity of the campaign is not the full story. The swing states of Iowa and Wisconsin both had average line lengths of less than ten minutes even though uncompetitive states such as Maryland, the District of Columbia, and South Carolina averaged more than twenty minutes.

Third, a state’s data do not necessarily tell a single story about election performance. An individual state will sometimes vary in how it does at the four broad categories of analysis: registering, voting, counting, and data transparency. The same appears to be true when it comes to other dimension of our analysis: security versus access. This book project has made clear that there are more and better measures of voter access than of ballot security. Although indicators of provisional ballot usage might be considered on both sides of the ledger, in general there is better data on how voters interact with the system than on how the system protects its integrity. We hope that this book will help spur more effort on filling out the needs for security measures as it pushes for more comprehensive election performance data generally. The substantial variation in outcomes the authors observe across the states demonstrates that no state need be locked in subpar performance. The fact that other states – even those in similar circumstances – are able to produce higher registration rates, fewer rejected ballots, shorter lines, fewer people deterred by disabilities, and lower residual vote rates should encourage state and local officials to continue pursuing improvements.

Fourth, election performance is often, but not always, resistant to change. State indicators are rather sluggish; they display a great deal of inertia in terms of evolution over time. For example, state voter turnout in 1980 is correlated with turnout in 2008 at .56. States that used provisional voting before HAVA required all states to use them have the highest usage levels of all the states. This suggests that ingrained political cultures, demographics, and folkways have a constraining effect on election performance. However, the slow pace of change also makes it easier to spot more abrupt turning points in election performance. As long as the campaign environment does not vary significantly between two elections, when an indicator shifts in a state from one election to the next, it may often be attributed quite clearly to changes in election administration because other background factors were roughly held

---

27 This implies that knowing turnout in 1980 only explains 31 percent of the variance in 2008 turnout (.56 × .56), leaving the remaining 69 percent to demographic changes, electoral politics, and election administration.
Thus, when a state experiences longer lines or higher absentee rejection rates, it is not hard to detect and diagnose.

Thus, taken together, the chapters in this volume demonstrate that it is possible to conduct a thorough, detailed analysis of the administration of elections in the United States using a varied collection of data and techniques that is not difficult to assemble. It is our hope that the works in this volume will seed an enhanced interest in improving elections in the United States in a way that arises out of a common understanding of what needs our immediate attention as a nation.