



Logic and Accuracy Testing: A Fifty-State Review

Josiah Walker^(✉), Nakul Bajaj, Braden L. Crimmins, and J. Alex Halderman

University of Michigan, Ann Arbor, USA
{jhwalker,nbajaj,bradenlc,jhalderm}@umich.edu

Abstract. Pre-election logic and accuracy (L&A) testing is a process in which election officials validate the behavior of voting equipment by casting a known set of test ballots and confirming the expected results. Ideally, such testing can serve to detect certain forms of human error or fraud and help bolster voter confidence. We present the first detailed analysis of L&A testing practices across the United States. We find that while all states require L&A testing before every election, their implementations vary dramatically in scope, transparency, and rigorousness. We summarize each state’s requirements and score them according to uniform criteria. We also highlight best practices and flag opportunities for improvement, in hopes of encouraging broader adoption of more effective L&A processes.

1 Introduction

The vast majority of votes in the United States are counted mechanically, either by optical scanners that read paper ballots or by direct-recording electronic (DRE) voting machines [70]. To validate that these tabulation devices are configured and functioning correctly, jurisdictions perform a procedure called “logic and accuracy testing” (“L&A testing”) shortly before each election. It typically involves casting a “test deck”—a set of ballots with known votes—on each machine, then printing the results and ensuring the tally is as expected. Any deviation is a potential indicator that the election equipment has misbehaved.

While more sophisticated mechanisms such as risk-limiting audits [37], and end-to-end verification [15] can reliably detect and recover from both errors and attacks after the fact, they are not yet widely applied in the U.S. Even if they were, L&A testing would remain useful for heading off some sources of error before they affected results. Ideally, L&A testing can protect against certain kinds of malfunction, configuration error, and fraud as well as strengthen voter confidence, but its effectiveness depends on many details of how the testing is performed. In the U.S., L&A testing requirements—like most aspects of election procedure and the selection of voting equipment—are determined by individual states, resulting in a diversity of practices with widely varying utility.

Unfortunately, this heterogeneity means that many states diverge negatively from the norm and makes it difficult to offer the national public any blanket

assurances about the degree of protection that L&A testing affords. Moreover, many states do not publish detailed L&A procedures, leaving voters with little ability to assess the effectiveness of their own states’ rules, let alone whether any tests they observe comply with them. Yet this decentralized regulatory environment has also allowed a variety of positive L&A testing procedures to evolve, and there are abundant opportunities for the exchange of best practices.

This paper provides the first comparative analysis of L&A testing requirements across the fifty states. To determine how each state performs L&A testing, we conducted an extensive review of available documentation and reached out to election officials in every state. We then assessed and scored each state’s policy using criteria designed to reflect its functional effectiveness and suitability as a basis for voter confidence. The results provide a detailed understanding of how states’ procedures differ and how well they approach an ideal model of what L&A testing can achieve. Our analysis reveals that several important L&A criteria are absent in many or most states’ rules, yet we also highlight specific examples of policies that could serve as models for broader dissemination. We hope this work will encourage the adoption of more effective L&A testing requirements across the United States and help promote policies that better inspire public trust.

2 Background

2.1 L&A Testing Goals

L&A testing was first introduced in the early 1900s for lever-style voting machines [63], which contained a mechanical counter for each candidate. The counters were susceptible to becoming jammed due to physical failure or tampering, so tests were designed to establish that each counter would advance when voted.

Modern DRE voting machines and ballot scanners can suffer from analogous problems—miscalibrated touch-screens or dirty scanner heads can prevent votes in specific ballot positions from being recorded [31]—but they also have more complex failure modes that call for different forms of testing. These devices must be provisioned with an “election definition” that specifies the ballot layout and rules. If the election definition is wrong—for instance, the order or position of voting targets do not match the ballots a scanner will read—votes may be miscounted.

Problems with election definitions caused by human error are surprisingly common. They contributed to the publication of incorrect initial election results in Northampton County, Pennsylvania, in 2019 [14], Antrim County, Michigan, in 2020 [28], and DeKalb County, Georgia, in 2022 [23]. In these documented cases the errors were fortunately detected, but only after the results were announced. They likely could have been prevented in the first place by sufficient L&A testing.

L&A testing can also serve a role in election security. Research has long recognized that L&A testing *cannot* reliably defeat an adversary who manages to execute malware on voting machines, because the malware could detect when it was under test and only begin cheating during the election itself (see, e.g., [24]).

However, L&A testing can potentially thwart more limited attackers who manage to tamper with election definitions or configuration settings. For example, although there is no evidence that the instances of error described above were caused by fraud, attackers could cause similar election definition problems deliberately in an attempt to alter results. This would likely require far less sophistication than creating vote-stealing malware. Moreover, there is growing concern about threats posed by dishonest election insiders, who routinely have the access necessary to perform such an attack [17].

Beyond providing these protections, L&A testing also frequently serves a role in enhancing public confidence in elections. Most states conduct at least part of their L&A testing during a public ceremony, where interested political party representatives, candidates, news media, and residents can observe the process and sometimes even participate by marking test ballots. Some jurisdictions also provide live or recorded video of their testing ceremonies online. These public tests can help build trust by allowing voters to meet their local officials, observe their level of diligence, and become more familiar with election processes. Additionally, public observers have the potential to make testing stronger, by providing an independent check that the required tests were completed and performed correctly. At least in principle, public observation could also help thwart attempts by dishonest officials to subvert L&A testing by skipping tests or ignoring errors.

2.2 U.S. Elections

L&A testing fills a role that is best understood with a view towards the broader context of election administration in the jurisdictions where it is practiced. In the U.S., many subjects are put to the voters, frequently all at once, and a single ballot might include contests ranging from the national presidency and congress to the state governor, legislature, and judges to the local mayor, city council, sheriff, and school board [13]. This means elections tend to involve many contests—typically around 20, although some jurisdictions have occasionally had nearly 100 [76]. There may also be several ballot variants within a single polling place to accommodate candidates from different sets of districts. These features make tallying by hand impracticable in many areas. As a result, nearly all jurisdictions rely on electronic tabulation equipment, today most commonly in the form of computerized ballot scanners [70]. Ensuring that these machines are properly configured and functioning on election day is the key motivation for L&A testing.

Election administration in the U.S. is largely the province of state and local governments. Although the Constitution gives Congress the power to override state law regarding the “manner of holding Elections for Senators and Representatives,” this authority has been applied only sparingly, for instance to establish accessibility requirements and enforce civil rights [30, 73]. Each state legislature establishes its own election laws, and the state executive (typically the secretary of state) promulgates more detailed regulations and procedures. In practice, election administration powers are exercised primarily by local jurisdictions, such

as counties or cities and townships, where local officials (often elected officials called “clerks”) are responsible for conducting elections [45].

Because of this structure, there is little standardization of election practices across the states, and L&A testing is no exception. Testing processes (and the ceremonies that accompany them) vary substantially between and within states. As we show, these variations have significant effects, both with respect to error-detection effectiveness and procedural transparency and intelligibility. Pessimistically, one can view this broad local discretion as a way for some jurisdictions to use lax practices with little accountability. We note, however, that it also grants many clerks the power to depart *upwards* from their states’ mandatory procedures, achieving stronger protections than the law requires. This provides an opportunity for improved practices to see early and rapid adoption.

2.3 Related Work

Although L&A testing itself has so far received little research attention, there is extensive literature analyzing other aspects of election mechanics across states and countries, with the goal of informing policymaking and spreading best practices. For instance, past work has examined state practices and their impacts regarding post-election audits [68], voter registration list maintenance [10], voter identification requirements [16], online voter registration [79], election observation laws [27], the availability of universal vote-by-mail [67]. A far larger body of research exists comparing state practices in fields other than elections.

Despite the abundance of this work, we are the first (to our knowledge) to examine states’ L&A testing practices in detail. A 2018 state-by-state report by the Center for American Progress [58] considered L&A testing among several other aspects of election security preparedness; however, it primarily focused on the narrow question of whether states required all equipment to be tested. To build upon this research, we consider many other policy choices that influence the effectiveness of L&A requirements and procedures.

3 Methodology

3.1 Data Collection

To gather information on states’ practices, we began by collecting official documentation where publicly available, relying primarily on state legal codes, state election websites, and Internet search engines. If we could not locate sufficient information, we attempted to contact the state via email or by phone to supplement our understanding or ask for clarifications. We directed these inquiries to the state elections division’s main contact point, as identified on its website.

State responses varied. While some states provided line by line answers to each of our questions, it was common for states to indicate that our criteria were more specific than what state resources dictated, pointing us instead to the same statutes and documentation we had already examined, providing us

with additional documentation that was still unresponsive to our questions, or replying in paragraphs that partially addressed some questions while completely disregarding others. In cases where we could not find evidence to support that a state satisfied certain criteria and the state did not provide supporting evidence upon request, we did not award the state any points for those criteria.

Upon finalizing our summary of each state's practices, we contacted officials again to provide an opportunity for them to complete or correct our understanding. Over the course of nine months, we communicated with all 50 states and received at least some feedback on our summaries from all but seven states—Iowa, New Jersey, New York, Rhode Island, Tennessee, Vermont, and Wisconsin. Our data and analysis are current as of July 2022.

3.2 Evaluation Criteria

To uniformly assess and compare states practices, we applied the following criteria and scoring methodology, which reflect attributes we consider important for maximizing the benefits of L&A testing in terms of accuracy and voter confidence. These criteria are non-exhaustive, but we believe they are sufficiently comprehensive to evaluate state procedures relative to one another. (Additional desirable testing properties are discussed in Sect. 5.) Note that our assessments do not necessarily reflect practice in each of a state's subdivisions, since local officials sometimes have authority to exceed state guidelines. To keep the analysis tractable, we instead focus on the *baseline* established by statewide requirements.

We developed two categories of criteria: *procedural criteria*, which encompass the existence of procedures, the scope of testing, and transparency; and *functional criteria*, which reflect whether the testing could reliably detect various kinds of errors and attacks. To facilitate quantitative comparisons, we assigned point values to each criterion, such that each category is worth a total of 10 points and the weights of specific items reflect our assessment of their relative importance.

Procedural Criteria

Rules and Transparency (5 points). To provide the strongest basis for trust, testing should meet or exceed published requirements and be conducted in public.

RT1 (1.5 pts): Procedures are specified in a detailed public document.

This captures the threshold matter of whether states have published L&A requirements. Detailed or step-by-step guidelines received full credit, and general laws or policies received half credit.

RT2 (1.0 pts): The document is readily available, e.g., via the state's website.

Making L&A procedures easily available helps inform the public and enables observers to assess tests they witness.¹

¹ Even when procedures are public documents, they are not always readily accessible. One state, Delaware, instructed us that we would need to find a resident to file a Freedom of Information Act request before their procedures would be provided.

RT3 (1.5 pts): Some testing is open to the public, candidates/parties, journalists.

This tracks the potential for public L&A ceremonies to strengthen confidence.

RT4 (1.0 pts): Local jurisdictions have latitude to exceed baseline requirements.

Scope of Testing (5 points). A comprehensive approach to testing covers every ballot design across all the voting machines or scanners where they can be used.

ST1 (2.0 pts): All voting machines/scanners must be tested before each election.

ST2 (1.0 pts): All devices must be tested *at a public event* before each election.

ST3 (2.0 pts): All devices must be tested with every applicable ballot design.

Failing to test all machines or all ballot styles risks that localized problems will go undetected, so each was assigned a substantial 2 points. One additional point was provided if all testing is public, to reflect transparency interests.

Functional Criteria

In each of three sets of functional criteria, we assess a simple form of the protection (with a small point value) and a more rigorous form (with a large point value).

Basic Protections (4 points). To guard against common errors, tests should cover every voting target and ensure detection of transpositions.

BP1 (1.0 pts): All choices receive at least one valid vote during testing.

BP2 (3.0 pts): No two choices in a contest receive the same number of votes.

The first test minimally detects whether each candidate has some functioning voting target. The second further ensures the detection of transposed targets within a contest, which can result from misconfigured election definitions.

Overvote Protection (2 points). Testing should exercise overvote detection and, ideally, confirm that the overvote threshold in each contest is set correctly.

OP1 (0.5 pts): At least one overvoted ballot is cast during testing.

OP2 (1.5 pts): For each contest c , a test deck includes a ballot with n_c selections and one with $n_c + 1$ selections, where n_c is the permitted number of selections.

An overvote occurs when the voter selects more than the permitted number of candidates, rendering the selections invalid. The first practice minimally detects that the machine is configured to reject overvotes, while the second tests that the allowed number of selections is set correctly for each contest.

Nondeterministic Testing (4 points). For stronger protection against deliberate errors, attackers should be unable to predict how the test deck is marked.

ND1 (1.0 pts): Public observers are allowed to arbitrarily mark and cast ballots.

ND2 (3.0 pts): Some ballots must be marked using a source of randomness.

Attackers who can predict the test deck potentially can tamper with the election definition such that errors will not be visible during testing. If the public can contribute test ballots, this introduces uncertainty for the attacker, while requiring random ballots allows for more rigorous probabilistic detection.

State	Procedural Criteria							
	RT1	RT2	RT3	RT4	ST1	ST2	ST3	∑
	1.5	1.0	1.5	1.0	2.0	1.0	2.0	(/10)
MT	●	●	●	●	●	●	●	10.00
OH	●	●	●	●	●	●	●	10.00
PA	●	●	●	●	●	●	●	10.00
WA	●	●	●	●	●	●	●	10.00
CO	◐	●	●	●	●	●	●	9.25
CT	◐	●	●	●	●	●	●	9.25
MA	◐	●	●	●	●	●	●	9.25
MO	◐	●	●	●	●	●	●	9.25
NV	◐	●	●	●	●	●	●	9.25
SD	◐	●	●	●	●	●	●	9.25
UT	◐	●	●	●	●	●	●	9.25
VA	●	●	◐	●	●	●	●	9.25
WV	◐	●	●	●	●	●	●	9.25
WI	◐	●	●	●	●	●	●	9.25
WY	◐	●	●	●	●	●	●	9.25
MI	●	●	●	●	●	○	●	9.00
MN	●	●	●	●	●	○	●	9.00
NH	●	●	●	◐	●	●	●	9.00
VT	●	●	●	◐	●	●	●	9.00
ID	◐	◐	●	●	●	●	●	8.75
AK	●	◐	●	◐	●	●	●	8.50
DE	●	◐	●	◐	●	●	●	8.50
GA	●	◐	●	◐	●	●	●	8.50
NC	●	◐	●	◐	●	●	●	8.50
AL	◐	●	●	●	●	○	●	8.25
ND	◐	●	●	◐	●	●	●	8.25
AZ	●	●	●	●	●	●	◐	8.00
IL	●	◐	●	●	●	●	○	7.50
FL	●	●	●	●	●	○	○	7.00
IN	●	●	●	●	○	○	●	7.00
SC	●	◐	●	◐	●	●	◐	6.50
IA	◐	●	●	◐	●	●	◐	6.25
KS	◐	●	●	◐	●	●	◐	6.25
MS	◐	●	●	◐	●	●	◐	6.25
RI	◐	●	●	◐	●	●	◐	6.25
TX	◐	●	●	●	○	○	●	6.25
AR	●	●	●	◐	●	◐	◐	6.00
NE	●	◐	◐	◐	●	◐	◐	6.00
KY	◐	●	○	◐	●	○	●	5.75
MD	◐	●	○	◐	●	○	●	5.75
NY	◐	●	◐	◐	●	◐	●	5.75
CA	◐	●	◐	●	●	○	◐	5.50
LA	◐	●	◐	◐	●	●	◐	5.50
NM	◐	●	◐	○	●	●	○	5.50
HI	●	◐	○	●	●	○	○	5.00
OR	◐	●	◐	◐	●	○	◐	4.50
ME	◐	●	●	◐	◐	◐	◐	3.25
NJ	◐	●	●	◐	◐	◐	◐	3.25
OK	●	◐	◐	◐	◐	◐	◐	2.75
TN	◐	●	◐	◐	○	○	◐	2.50

State	Functional Criteria						
	BP1	BP2	OP1	OP2	ND1	ND2	∑
	1.0	3.0	0.5	1.5	1.0	3.0	(/10)
CT	●	●	●	●	○	◐	7.50
AZ	●	◐	●	●	●	●	7.00
SD	●	●	●	●	●	○	7.00
IL	●	●	●	●	○	○	6.00
MO	●	●	●	●	○	○	6.00
NE	●	●	●	●	◐	◐	6.00
MI	●	●	●	○	●	○	5.50
UT	●	●	●	○	●	○	5.50
VT	●	●	◐	◐	◐	◐	5.50
WY	●	●	●	○	●	○	5.50
AR	●	●	●	◐	◐	◐	4.50
DE	●	●	●	○	○	○	4.50
FL	●	●	●	◐	◐	○	4.50
MN	●	●	●	○	◐	◐	4.50
MT	●	●	●	○	○	○	4.50
ND	●	●	●	○	○	○	4.50
OH	●	●	●	◐	◐	◐	4.50
WA	●	●	●	○	○	○	4.50
ID	●	○	●	●	●	○	4.00
NM	●	○	●	●	○	○	3.00
CO	●	○	●	○	●	○	2.50
IA	●	◐	●	◐	●	◐	2.50
AL	●	◐	●	◐	◐	◐	1.50
GA	●	◐	●	◐	◐	◐	1.50
IN	●	◐	●	◐	○	○	1.50
KS	●	◐	●	◐	◐	◐	1.50
KY	●	◐	●	◐	◐	◐	1.50
ME	●	◐	●	◐	◐	◐	1.50
NV	●	◐	●	◐	◐	◐	1.50
NH	●	◐	●	◐	◐	◐	1.50
NJ	●	◐	●	◐	○	◐	1.50
NY	●	◐	●	◐	◐	◐	1.50
PA	●	○	●	◐	◐	◐	1.50
TN	●	◐	●	◐	◐	◐	1.50
TX	●	○	●	○	○	○	1.50
VA	●	◐	●	◐	◐	◐	1.50
WV	●	◐	●	◐	◐	◐	1.50
WI	●	◐	●	◐	◐	○	1.50
HI	○	○	●	○	◐	○	1.00
MD	●	○	○	○	○	○	1.00
SC	●	◐	○	○	◐	◐	1.00
MA	○	○	●	○	○	○	0.50
MS	○	◐	●	◐	○	◐	0.50
NC	◐	◐	●	◐	○	◐	0.50
OK	◐	◐	●	◐	○	◐	0.50
AK	◐	◐	◐	◐	◐	◐	0.00
CA	○	○	○	○	○	○	0.00
LA	○	◐	◐	◐	○	◐	0.00
OR	◐	◐	◐	◐	○	◐	0.00
RI	◐	◐	◐	◐	◐	◐	0.00

●: Fully met ◐: Partly met ○: Not met ◑: State not responsive (scored as unmet)

4 Analysis

Our nationwide review of L&A procedures highlights significant variation among the testing practices of the fifty states, as illustrated by the maps in Fig. 2. The tables on page 7 summarize our findings and rank the states with respect to the procedural and functional criteria. We also provide a capsule summary of each state’s practices in Appendix A.

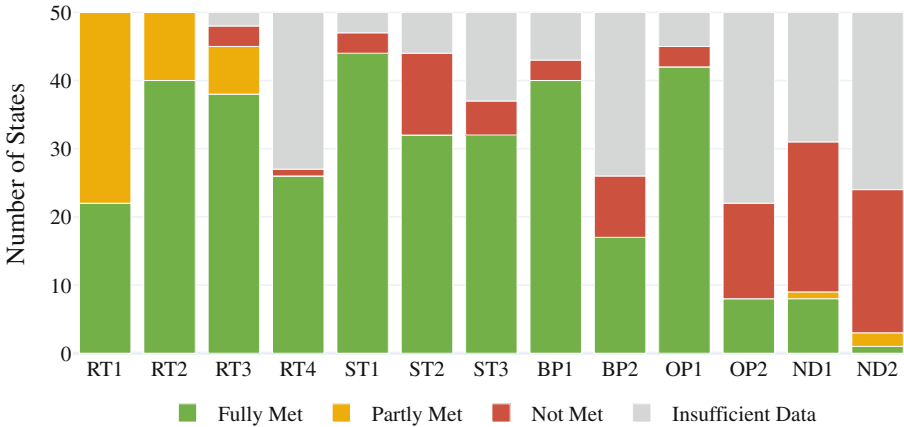


Fig. 1. We count the number of states that met, partly met, or did not meet each criterion. While states commonly require simple protections (BP1, OP1), most do not achieve more rigorous forms of error detection (e.g., OP2, ND1, and ND2).

4.1 Performance by Criterion

Figure 1 shows the number of states that met, partly met, or did not meet each criterion. All 50 states have laws that require L&A testing, but only 22 have a public, statewide document that details the steps necessary to properly conduct the tests (RT1). Of those that do not, several (such as California) merely instruct local jurisdictions to follow instructions from their voting equipment vendors, limiting the efficacy of logic and accuracy procedures to each vendor’s preferences.

States generally performed well with respect to transparency criteria. Every state has some public documentation about its L&A practices, with 40 states making this documentation readily available (RT2). At least 45 states perform some or all testing in public (RT3), although 7 of these impose restrictions on who is allowed to attend. At least 32 states test every machine in public (ST2). Just three states (Kentucky, Maryland, and Hawaii) do not conduct any public L&A testing, which may be a significant lost opportunity to build public trust.

Most states also scored high marks regarding the scope of their testing. We were able to confirm that at least 44 states require all equipment to be tested before each election (ST1). Exceptions include Tennessee, Texas, and Indiana, which only require testing a sample of machines. At least 32 states require every ballot style to be tested, but 5 or more do not (ST3), which increases the chances that problems will go undetected in these jurisdictions.

Consideration of several other criteria is more complicated, because we often lack evidence for or against their being met. We have insufficient data about 19 or more states for RT4 and most of the functional criteria (BP2, OP2, ND1, and ND2). Details concerning functional criteria tended to be less frequently described in public documentation, which potentially biases our analysis in cases where states were also unresponsive to inquiries. We treated such instances as unmet for scoring purposes, but it is also informative to consider the ratio of met to unmet, as depicted in Fig. 1. One example is whether states allow local jurisdictions to exceed their baseline requirements (RT4). Although we lack data for 23 states, the criterion is met by at least 26 states, and we have only confirmed that it is unmet in one state (New Mexico). This suggests that many of the unconfirmed states likely also allow local officials to depart upwards from their requirements.

After accounting for what data was available, states clearly perform much better on our procedural criteria than on our functional criteria. This suggests that many of the functional attributes we looked for are aspirational relative to current practice and indicates that L&A testing *could* provide much more value.

The only two functional criteria that most states meet are basic protections for voting targets (BP1) and overvotes (OP1), which are provided in at least 40 and 42 states, respectively. At least 17 states would detect transposed voting targets within contests (BP2), but as few as 8 fully validate overvote thresholds (OP2). These more rigorous protections require more complicated procedures and larger test decks, but that some states achieve them suggests that they would be practical to implement more broadly.

Policies facilitating even the basic form of nondeterministic testing were rare. Only 11 states scored even partial points for conducting nondeterministic testing, with 9 of them allowing public observers to mark test ballots (ND1) and 3 of them (Arizona, Connecticut, Vermont) confirming that election officials are required to mark random selections (ND2). Of the three, only Arizona confirmed that it required officials to use a random number generator, thus earning full points. These findings are surprising, since unpredictable or randomized testing can thwart certain kinds of attacks that predictable tests cannot. That nondeterministic testing is rare greatly limits the security benefits of typical state L&A practices.

4.2 Performance by State

When comparing states' overall L&A testing practices, we find wide variation across both procedural and functional criteria. As illustrated in Fig. 2, this vari-

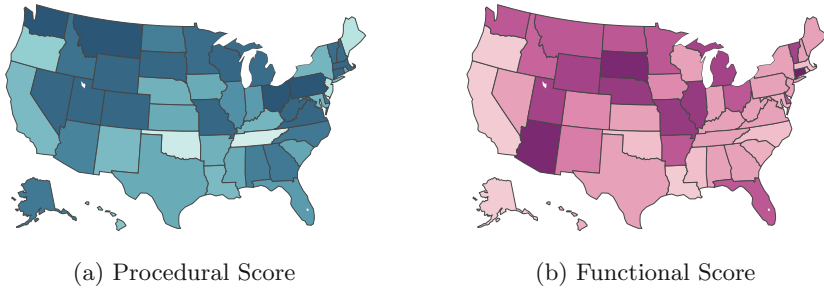


Fig. 2. Mapping state scores (darker indicates better performance) shows that L&A testing practices vary significantly within all geographic regions of the U.S.

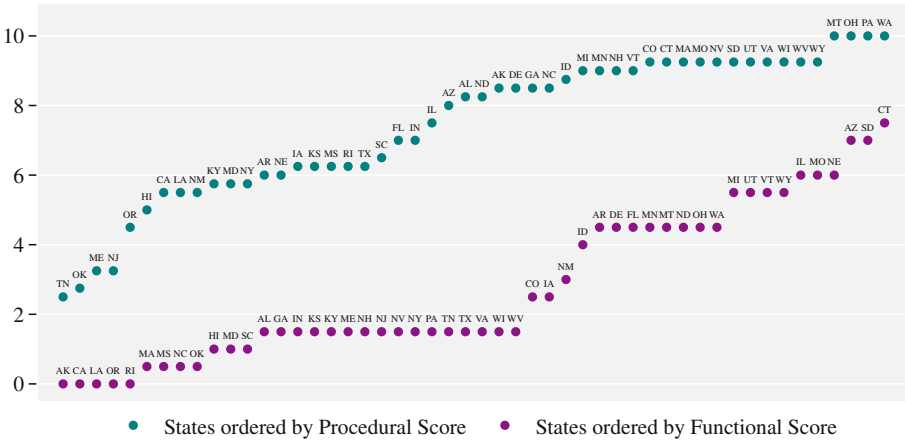


Fig. 3. Many states have perfect or near-perfect procedural scores, but functional scores are generally lower, reflecting opportunities for making L&A more effective.

ation is not clearly explained by regionalism. However, the plot in Fig. 3 reveals several notable features in the distributions of states’ scores.

Most obviously, procedural scores were much higher than functional scores. Again, this likely reflects both the relative scarcity of public documentation about functional aspects of L&A testing and that our chosen functional criteria were somewhat aspirational. No states achieved perfect functional scores, but 4 states (Montana, Ohio, Pennsylvania, and Washington) achieved perfect procedural scores. Four other states could potentially achieve this benchmark but did not provide missing information we requested. Eleven more states clustered just shy of perfect procedural scores, of which 10 could achieve full points simply by making detailed L&A procedures public (BP1)—potentially a zero-cost policy change.

Notable relationships occur between certain criteria. For instance, concerning the scope of testing, only 2 states that are known to require testing every ballot

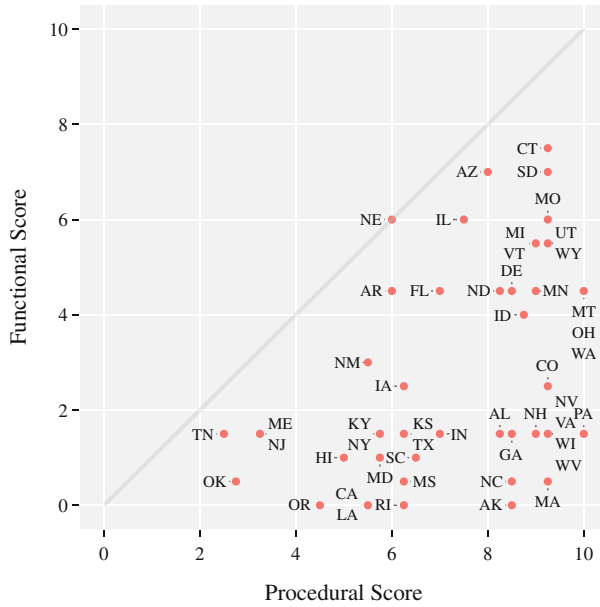


Fig. 4. No state’s functional score exceeds its procedural score, perhaps due to more limited data about functional aspects of testing. At most levels of procedural scores, states’ functional scores spanned a wide range, with no strong correlation.

(ST3) style do not also require testing every machine (ST1). It is much more common for states that require testing every machine to not require testing every ballot style (or remain silent), which 14 of 44 states did. This suggests that L&A policymakers are more likely to be aware of the potential for problems that affect only specific machines than of issues that can affect only specific ballot styles.

The distribution of functional scores highlights further relationships. The largest cluster, at 1.5, are 16 states that require basic protections for voting targets (BP1) and overvotes (OP1) but meet no other functional criteria. Interestingly, many of these states employ similar statutory language, with little or no variation.² Although we have so far been unable to identify the common source these states drew on, the situation suggests that providing stronger model legislation could be a fruitful way to encourage the adoption of improved L&A practices.

At least 21 other states accomplish basic voting-target and overvote protections plus one or more of the stronger functional criteria. Most commonly, they require additional testing to detect transposed voting targets within contests (BP2), which 17 states do. Eight of these states accomplish no further functional criteria, resulting in a cluster at score 4.5. Five others also fully validate overvote thresholds (OP2), as do only 3 other states, indicating a strong correlation

² An additional state, South Carolina, adopted nearly the same statutory formula but with a small change that *weakens* overvote protection, and so does not satisfy OP1.

between these more rigorous testing policies. In a surprising contrast, although nondeterministic testing is comparably uncommon (only 8 states fully achieve either ND1 or ND2), practicing it does *not* appear to be well correlated any of the other non-basic functional criteria. This may indicate that states have introduced nondeterministic testing haphazardly, rather than as the result of careful test-process design.

Considering both scoring categories together (Fig. 4), we see that although no state's functional score exceeds its procedural score, there is otherwise a wide range of functional scores at almost every level of procedural score. This may partly reflect limitations due to unresponsive states, but it may also suggest opportunities to better inform policymakers about ways to strengthen L&A functionality, particularly in states in the lower-right corner of the figure, which have a demonstrated ability to develop procedurally robust testing requirements.

The overall national picture shows that every one of our evaluation criteria is satisfied by at least one state, indicating that even the most rigorous functional criteria are realistic to implement in practice. Several states have the potential to serve as models of best practice across most dimensions of L&A testing, especially if procedural specifics are made readily accessible. In particular, Arizona and South Dakota each achieved full points in all but one criterion from each category, and Connecticut achieved the highest total score of any state under our metrics. We provide additional information and references regarding their procedures in our state-by-state summaries, found in Appendix A.

5 Discussion

Our findings support the need for strengthened L&A procedures nationwide. Current practice has room for substantial improvement in both transparency and substance, and state policy should seek to realize this potential.

Election security researchers and practitioners should work together to establish normative standards for L&A testing procedures and to draft model legislation to realize them. The precise mechanism for establishing this standard is beyond the scope of this paper, but a potential route would be for the National Institute of Standards and Technology (NIST) to issue L&A testing guidelines. Under the Help America Vote Act (HAVA), NIST is charged with the design of voting system standards, in coordination with the U.S. Election Assistance Commission (EAC) [30], and it has previously issued guidance for other aspects of election technology administration, such as cybersecurity and accessibility.

One challenge in the adoption of any technical standard is leaving safe and flexible opportunities for upward departure. It would be dangerous to lock in procedures that are later found to be insufficient, especially if every state would have to update its laws in response. For this reason, it is important that any L&A policy changes allow some degree of flexibility involved for local jurisdictions. Too much flexibility, however, can weaken security guarantees even with the best of intentions. One clerk we spoke with in the preparation of this paper offhandedly told us that she did not always follow the state requirement that

every candidate in a contest receive a different number of votes, since in real elections ties could occur and she felt it was important to test that behavior too. Despite the well-meaning nature of this deviation, it *decreased* the guarantees provided by her L&A testing, since it meant the two candidates who tied in the test deck could have had their votes unnoticeably swapped. Clerks do not have the resources to rigorously analyze all ramifications of deviations from procedure, so latitude to deviate should be provided only where it cannot reduce the integrity of the process, such as in optional, additional phases of testing.

We leave to future work determining what model L&A policies should look like. While the elements of transparency, openness, and security we considered in this paper are potential low-hanging fruit, there are other elements of successful L&A practice that we did not measure or describe. For instance, testing policies should consider not only ballot scanners but also ballot marking devices (BMDs), which are computer kiosks that some voters use to mark and print their ballots. Most jurisdictions use BMDs primarily for voters with assistive needs, but some states require all in-person voters to use them [70]. Errors in BMD election definitions can lead to inaccurate results [23], but carefully designed L&A testing might reduce the incidence of such problems. Another example of an intervention that would have detected real-world issues in the past is “end-to-end” L&A testing, where tabulator memory cards are loaded into the central election management system (EMS) and its result reports are checked against the test decks. One of the problems in Antrim County that caused it to report initially incorrect results in 2020 was an inconsistency between some tabulators and the EMS software, and “end-to-end” L&A testing could have headed off this issue [28].

We do, however, recommend that future L&A guidelines incorporate elements of nondeterministic testing. While our data shows that this practice is still quite rare in the U.S., using test decks that are unpredictable would make it more difficult to construct malicious election definitions that pass the testing procedure.

Election technology has evolved over time, but some L&A testing practices still carry baggage from the past. For instance, functional requirements in many U.S. states are suited for detecting common problems with mechanical lever voting machines but less adept at uncovering common failure modes in modern computerized optical scanners, such as transposed voting targets. Other nations, which may at this time be adopting optical scan equipment of their own, can learn from these standards and improve on them as they choose their own practices for the future. By applying careful scrutiny of existing process and incorporating the elements that most make sense in their own context, these polities can ensure their testing procedures are constructed in a way to meet their needs.

6 Conclusion

We performed the first detailed comparative analysis of L&A testing procedures, based on a review of L&A requirements and processes across all fifty U.S. states.

Although L&A testing can be a valuable tool for spotting common configuration errors and even certain kinds of low-tech attacks, our analysis shows that there is wide variation in how well states' testing requirements fulfill these prospects. We hope that our work can help rectify this by highlighting best practices as well as opportunities for improvement. Rigorous, transparent L&A testing could also be a valuable tool for increasing public trust in elections, by giving voters a stronger basis for confidence that their votes will be counted correctly.

Acknowledgements. We are grateful to the many election workers who corresponded with us to provide data for this study. We also thank our shepherd Nicole Goodman as well as Doug Jones, Dhanya Narayanan, Mike Specter, Drew Springall, and the anonymous reviewers. This work was supported by the Andrew Carnegie Fellowship, the U.S. National Science Foundation under grant CNS-1518888, and a gift from Microsoft.

A State-by-State Practices

Here we summarize notable features of each state's L&A testing practices with respect to our evaluation criteria. We list each state's score and rank under the procedural and functional criteria (each /10 points) and their total (/20 points).

Alabama (AL) *Proc.*: 8.25 (25th) *Func.*: 1.5 (23rd) *Total*: 9.75 (28th)

Alabama's L&A [5] public testing requirements vary by class of device. Direct-recording electronic voting devices (DREs), which are no longer used [70], had to be tested at an event open to the general public. Testing of the state's current optical scan machines occurs in two phases: all devices are tested in a process observable by candidates or their representatives, and a subset of the devices is tested again in view of the general public. Interestingly, the state requires each candidate receive a minimum of *two* votes, although there is no requirement that candidates within a contest receive a *different* number of votes. Alabama permits local jurisdictions to implement more stringent testing practices in addition to required testing, so practices could be independently improved by local jurisdictions.

Alaska (AK) *Proc.*: 8.5 (21st) *Func.*: 0.0 (46th) *Total*: 8.5 (32nd)

Documents describing Alaska's L&A procedures are not publicly accessible, but the state's Division of Elections provided them in heavily redacted form in response to our requests [6–9]. Although Alaska requires each voting machine to be publicly tested, the redactions precluded our finding evidence responsive to most of our other criteria, resulting in a low score. Alaska has not responded to a request for clarification.

Arizona (AZ) *Proc.*: 8.0 (27th) *Func.*: 7.0 (2nd) *Total*: 15.0 (4th)

Arizona uses a random number generator to mark test ballots, creating a non-deterministic test deck that would be difficult to predict and earning the state full credit for ND2. The state also firmly bounds the overvote thresholds by voting at least one ballot for the maximum number of permissible choices in each

contest and then exactly one additional vote than what is allowable [11]. Lastly, the state permits observers from local political parties to contribute their own votes to a test ballot. The state could earn a perfect score overall by testing each ballot style and ensuring candidates within a contest receive distinct vote totals.

Arkansas (AR) *Proc.:* 6.0 (37th) *Func.:* 4.5 (11th) *Total:* 10.5 (25th)

Arkansas allows the public to observe the testing of all devices prior to an election. The state ensures every candidate receives at least one vote and that at least one overvoted ballot must be cast [12]. Its procedures can be strengthened by specifying proper overvote validation practice, ensuring different vote totals for each candidate in a contest, and introducing nondeterministic test elements. Arkansas has not responded to requests for additional information.

California (CA) *Proc.:* 5.5 (42nd) *Func.:* 0.0 (46th) *Total:* 5.5 (44th)

California's Election Code §15000 [18] requires the state to conduct logic and accuracy testing. The state, however, relies on vendor-provided procedures that vary from machine to machine instead of implementing statewide requirements that are independent of the vendor. We assessed that this reliance on vendor-provided testing material does not satisfy our functional criteria, since even if every vendor's manual happens to meet a particular requirement now, updated manuals could weaken these provisions without any conscious regulatory action. Accordingly, the state scores lower than it likely would have otherwise.

Colorado (CO) *Proc.:* 9.25 (5th) *Func.:* 2.5 (21st) *Total:* 11.75 (18th)

Colorado has robust procedural L&A policies that are publicly applied to every machine [20]. Still, Colorado has room to improve in their functional protections. The state currently does not require vote totals to differ between various candidates for an office, but instead encourages it by providing a "ballot position calculator." Furthermore, Colorado does not require a strict test of overvote protection. Most of the information for our assessment was provided via communication with the state as there is no statewide procedure document.

Connecticut (CT) *Proc.:* 9.25 (5th) *Func.:* 7.5 (1st) *Total:* 16.75 (1st)

Connecticut's public statues require L&A testing [21]. Communication with the state election division revealed that the state comes close to fulfilling almost all of our criteria and made it the highest-scoring state overall. The main opportunity we find for improvement would be to strengthen nondeterministic testing. Connecticut already earns partial credit for ND2 since its procedures demonstrate an understanding of the importance of randomized testing, even though the source of randomness is not specified. To earn a perfect functional score, the state should require the use of a random number generator to mark some number of test ballots and also allow public observers to mark and cast test ballots arbitrarily.

Delaware (DE) *Proc.:* 8.5 (21st) *Func.:* 4.5 (11th) *Total:* 13.0 (14th)

Delaware's L&A policies are unclear based on its statutes [1], but a FOIA request for testing policies submitted at our behest by a Delaware resident yielded more

informative documents [22]. We therefore consider criterion RT1 to be met, although the process of obtaining the documentation was needlessly difficult. In addition to a lack of transparency around their L&A documentation, it is unclear whether the state allows local jurisdictions to practice more stringent testing requirements. However, all other procedural testing criteria were met. In terms of functional protections, both basic protections were included as part of testing requirements. Overvote protections are not strictly checked; test decks include ballots where all choices are marked for each office but do not include ones marked with exactly one more vote than is permissible. Additionally, election officials are not required to mark test ballots randomly.

Florida (FL) *Proc.:* 7.0 (29th) *Func.:* 4.5 (11th) *Total:* 11.5 (19th)

Florida tests all tabulators during a “100% Logic and Accuracy Testing” event and a sample during a public L&A test [25]. The state meets all procedural testing criteria except for requiring all machines to be tested publicly and with each ballot style. Florida recognizes that some counties use a traditional 1-2-3 test deck pattern (satisfying BP2) but notes that it is not the most accurate way to verify that ballots are being counted correctly. To supplement this testing mechanism, the state encourages (although does not require) the creation of an “enhanced test deck with non-traditional vote patterns.”

Georgia (GA) *Proc.:* 8.5 (21st) *Func.:* 1.5 (23rd) *Total:* 10.0 (27th)

Georgia maintains a step-by-step document that describes how to implement its L&A rules [26]. The procedures, however, do not meet many criteria including requiring testing of all ballot styles and ensuring candidates for an office have different vote totals. Additionally, the procedures could not be located from the state’s election website.

Hawaii (HI) *Proc.:* 5.0 (45th) *Func.:* 1.0 (39th) *Total:* 6.0 (43rd)

Hawaii’s L&A procedures [29] differ substantially from most other states. The state does not permit the general public to observe testing. Instead, “Official Observers” must be designated by a political party, news media organization, or chief election officer to “serve as the ‘eyes and ears’ of the public.” In practice, “Official Observers” are quasi election officials who are tasked with conducting testing however they choose. Given Hawaii’s unique situation, we assigned the state partial credit for ND1. However, since testing is not open to members of the general public unless they become “Official Observers,” Hawaii did not earn credit for RT3. Hawaii earned full credit for requiring all tabulators to be tested, but it failed to meet other key criteria including ensuring that each candidate receives at least one vote and that all ballot styles are included during testing.

Idaho (ID) *Proc.:* 8.75 (20th) *Func.:* 4.0 (19th) *Total:* 12.75 (15th)

Although Idaho does not have a statewide document dedicated to explaining the state’s testing procedures, communication with state officials revealed that Idaho meets all functional criteria except for requiring proper overvote validation (BP2) and for election officials to mark ballots using a source of randomness (ND2).

Illinois (IL) *Proc.*: 7.5 (28th) *Func.*: 6.0 (4th) *Total*: 13.5 (12th)

The Illinois L&A testing best practice guide [33] fulfills many functional and procedural criteria. Importantly, the state tactfully juxtaposes an example unsatisfactory test with an example satisfactory test in way that efficiently conveys its L&A requirements—a method that states hoping to improve procedural clarity should consider. Illinois would benefit from introducing nondeterministic elements to its L&A testing and requiring that all ballot styles be tested on each tabulator.

Indiana (IN) *Proc.*: 7.0 (29th) *Func.*: 1.5 (23rd) *Total*: 8.5 (32nd)

Indiana distinguishes between public testing and logic and accuracy testing in its procedure manual [34], requiring the former and leaving the latter undefined. What the state terms “public testing,” however, is analogous to other states’ L&A. While the state requires testing of all optical scan (“ballot card”) tabulators, only 5% of DRE voting systems must be tested. Since much of Indiana still uses DREs [70], it should strongly consider requiring all devices to undergo testing prior to each election as well as introducing explicit requirements regarding basic protections, overvote threshold validation, and nondeterminism. Indiana does require that all ballot styles be tested on each device subject to testing.

Iowa (IA) *Proc.*: 6.25 (32nd) *Func.*: 2.5 (21st) *Total*: 8.75 (31st)

Iowa has limited public documentation of its election procedures and has not responded to any of our email inquiries regarding its testing practices. Therefore, the state was scored in accordance with what could be found in its Election Code [32] and an election security informational video on its website [35].

Kansas (KS) *Proc.*: 6.25 (32nd) *Func.*: 1.5 (23rd) *Total*: 7.75 (35th)

Kansas includes L&A testing requirements as part of its state code, §25-4411 [36]. This statute requires public testing with a test deck that includes at least one overvoted ballot for each machine. Kansas also requires that L&A testing be open to the public and that all machines are tested. This gives the state a somewhat better procedural testing score than it would have had otherwise, but there are still many improvements that can be made, such as specifying a requirement for the testing of each ballot style and detection of transposed targets. The state’s election division has not responded to our inquiries regarding testing requirements and so has been scored purely on the basis of public information.

Kentucky (KY) *Proc.*: 5.75 (39th) *Func.*: 1.5 (23rd) *Total*: 7.25 (38th)

Kentucky does not have any statewide procedural L&A document, but the state does have L&A requirements under its administrative code [3]. Kentucky earned points for requiring all of its voting equipment to undergo testing; however, the state does not require public L&A testing. Instead, it permits a representative from each political party and representatives of news media to be present at a “Public Examination” as the county elections board ensures: ballots are properly arranged, counters are set to zero, equipment is locked, and that the equipment’s assigned serial number is recorded.

Louisiana (LA) *Proc.*: 5.5 (42nd) *Func.*: 0.0 (46th) *Total*: 5.5 (44th)

Louisiana Election Code [38] requires testing of every machine before each election, but it does not appear to establish any minimum standards for what this testing entails. Unusually, only Louisiana citizens are permitted to observe testing, so the state earns only partial credit for RT3. The state still earned full credit for ST2 because it ensures that the permitted public observers may witness the preparation and testing of each machine.

Maine (ME) *Proc.*: 3.25 (47th) *Func.*: 1.5 (23rd) *Total*: 4.75 (46th)

Found in 1.21–A M.R.S.A [2], Maine’s L&A policy has several opportunities for improvement. It is unclear whether each tabulator is required to be tested, whether all ballot styles are included in testing, whether two candidates from the same contest can receive the same number of votes, and so forth. Observers are also not permitted to test ballots and machines themselves. Maine should consider producing publicly accessible procedure or guideline documents that expand upon its policy. Maine’s Bureau of Corporations, Elections and Commission has not responded to our email inquiries regarding its testing procedures.

Maryland (MD) *Proc.*: 5.75 (39th) *Func.*: 1.0 (39th) *Total*: 6.75 (40th)

Described in Maryland COMAR (33.10.01.14–16) [19], the state’s L&A testing consists of a “Pre-Election Test” of all tabulators. It is followed by a “Public Demonstration” that is limited to attendance by one representative of each political party and independent candidate. The public demonstration only consists of documentation completed during the pre-test and an overview of the testing process. Maryland should consider permitting the general public to observe “Pre-Election Testing” and creating a readily accessible statewide procedural document so that the public can make informed observations.

Massachusetts (MA) *Proc.*: 9.25 (5th) *Func.*: 0.5 (42nd) *Total*: 9.75 (28th)

Massachusetts does not have an approved statewide document regarding L&A procedures, but it does have regulations on the subject [4]. Communication with state officials uncovered that the state’s internal guidelines meet most procedural testing criteria. However, the state performs poorly in terms of test functionality, where the only criterion it meets is basic overvote protection (OP1).

Michigan (MI) *Proc.*: 9.0 (16th) *Func.*: 5.5 (7th) *Total*: 14.5 (7th)

Michigan’s L&A procedures are documented in a public manual [39] and consist of a preliminary accuracy test in which all tabulators and BMDs are tested, as well as a public test in which only a sample of tabulators are required to be tested. Each candidate in a contest must receive a different number of votes, and overvoted ballots must be cast. Additionally, observers can mark and cast test ballots. We recommend that Michigan further ensure that all machines undergo public testing, that overvote thresholds are fully validated, and that some ballots are marked using a source of randomness.

Minnesota (MN) *Proc.*: 9.0 (16th) *Func.*: 4.5 (11th) *Total*: 13.5 (12th)

The state’s public L&A guidelines [40] identify two testing events—a preliminary

test in which all ballot counters are required to be tested and a public accuracy test in which only a sample is tested. Minnesota includes step-by-step instructions for creating test deck spreadsheets and provides samples that jurisdiction can utilize. The state does well in ensuring that its guidelines protect against problems identified in the basic protections category. The state could benefit from introducing nondeterministic elements to its testing as well as requiring election jurisdictions to fully validate overvote thresholds.

Mississippi (MS) *Proc.:* 6.25 (32nd) *Func.:* 0.5 (42nd) *Total:* 6.75 (40th)

Mississippi's L&A requirements can be found in its state code [41], which only requires that each machine be publicly tested for basic overvote rejection (OP1). The state has not responded to an inquiry regarding other procedural and functional elements of its testing.

Missouri (MO) *Proc.:* 9.25 (5th) *Func.:* 6.0 (4th) *Total:* 15.25 (3rd)

Missouri's L&A policy is defined in 15 CSR 30–10.040 [42]. Testing ensures that each candidate for an office receives a distinct and nonzero number of votes. Missouri is one of the few states that require full validation of overvote thresholds, with a requirement as follows: “In situations where a voter can legally vote for more than one person for an office, at least one card shall be voted for the maximum number of allowable candidates; one card shall [then] be marked to have one more vote for each candidate or question than is allowable.” This language could serve as a model for other states. We recommend introducing nondeterministic testing but commend the state for its unusually clear functional policies.

Montana (MT) *Proc.:* 10.0 (1st) *Func.:* 4.5 (11th) *Total:* 14.5 (7th)

Under Montana's detailed public L&A procedures [43], testing is divided into three dimensions—functional, diagnostic, and physical—and fulfills almost all of our criteria. Montana could strengthen its procedures by fully bounding overvote thresholds and introducing forms of nondeterministic testing.

Nebraska (NE) *Proc.:* 6.0 (37th) *Func.:* 6.0 (4th) *Total:* 12.0 (17th)

Nebraska's logic and accuracy test is commonly referred to as a “Mock Election.” All tabulators undergo three independent tests using different test decks—one by the election official, one by the chief election commissioner, and one by the person who installed the election definition on the voting device. The state's Election Act [46] does not indicate that testing observation is open to members of the public; however, its test procedures meet all other criteria in the scope of testing, basic protections, and overvote protection categories. Nebraska could further improve its test functionality by requiring nondeterministic testing.

Nevada (NV) *Proc.:* 9.25 (5th) *Func.:* 1.5 (23rd) *Total:* 10.75 (21st)

In Nevada, L&A testing is primarily described by state law [47]. All tabulators must be tested before an election in a process that can be observed by the general public, and local jurisdictions are allowed to exceed state testing requirements. Nevada would benefit from creating a public, statewide guide that describes the steps necessary to properly conduct L&A testing.

New Hampshire (NH) *Proc.*: 9.0 (16th) *Func.*: 1.5 (23rd) *Total*: 10.5 (25th)
 New Hampshire’s *Election Procedure Manual* [48] meets all scope of testing requirements but otherwise does not address several key criteria, including the more rigorous of basic and overvote protections. Although the state broadly requires that election officials mark ballots with “as many combinations as possible,” we deem this to fall short of satisfying our nondeterministic testing criteria.³ Calling for some test ballots to be marked truly at random would strengthen this provision.

New Jersey (NJ) *Proc.*: 3.25 (47th) *Func.*: 1.5 (23rd) *Total*: 4.75 (46th)
 New Jersey’s L&A policy is described by state statutes 19:53A-8 [51]. Testing is conducted publicly and includes one vote for each candidate as well as the casting of overvoted ballots. State law was unclear regarding other key aspects of testing, including whether all machines are to be tested prior to each election. The state did not respond to additional requests for information.

New Mexico (NM) *Proc.*: 5.5 (42nd) *Func.*: 3.0 (20th) *Total*: 8.5 (32nd)
 New Mexico has a handbook of relevant election code and legislation, but no procedure document [49]. While this legislation requires public testing before each election, no other information was present relevant to our criteria. Further communication with state officials revealed that only “party and organization representatives, election observers and candidates” are allowed to observe testing. We did not receive a response on whether all ballot styles are tested on each tabulator. Functionally, the state’s testing implements both overvote protections but only the first basic protection (BP1).

New York (NY) *Proc.*: 5.75 (39th) *Func.*: 1.5 (23rd) *Total*: 7.25 (38th)
 New York has a relatively minimal set of requirements defining their L&A procedure [50], which they term “Prequalification Testing.” Under these rules, all tabulators and all ballot styles are tested, but other important properties are not met: testing is not open to the public, overvote thresholds are not well-bounded, and multiple candidates in the same contest may receive an equal number of votes. We recommend addressing these shortfalls and introducing nondeterministic testing elements. The New York State Board of Elections did not respond to an email inquiry requesting additional information.

North Carolina (NC) *Proc.*: 8.5 (21st) *Func.*: 0.5 (42nd) *Total*: 9.0 (30th)
 North Carolina’s L&A testing is briefly described on the state’s election website [52]; we were able to obtain the state’s procedures via correspondence with the State Board of Elections. The step-by-step list meets most of our procedural criteria but almost none of our functional criteria.

North Dakota (ND) *Proc.*: 8.25 (25th) *Func.*: 4.5 (11th) *Total*: 12.75 (15th)
 Email communication with state election officials provided a great deal of insight

³ Taken literally, this is intractable. There are 2^n ways to mark n voting targets; an election with 50 candidates would require approximately as many ballots as there are grains of sand on earth.

regarding North Dakota's L&A testing [44]. The state met the more advanced criteria requiring that all ballot styles be tested on every machine (ST3) and that a different numbers of votes be assigned to each option for each contest. It does not, however, require validating overvote thresholds or any nondeterministic testing.

Ohio (OH) *Proc.:* 10.0 (1st) *Func.:* 4.5 (11th) *Total:* 14.5 (7th)

Ohio's L&A practices [53] indicate that state election officials are required to test tabulating computer programs prior to a given election and tabulating equipment prior to its use to count ballots. In addition to fulfilling each scope of testing and basic protections criterion, the state met all of our procedural criteria. The state would earn a perfect combined scored if it were to specify a requirements that election officials fully validate overvote thresholds and introduce nondeterministic elements into the test process.

Oklahoma (OK) *Proc.:* 2.75 (49th) *Func.:* 0.5 (42nd) *Total:* 3.25 (50th)

State law in Oklahoma contains some provisions related to testing but leaves significant gaps relative to our criteria. When we contacted the State Election Board to request more information, they provided a page from the state's *Uniform Election Reference* [54]. This still failed to answer many of our questions, but contained some language which suggests local election officials may have access to additional private documents further defining L&A requirements. The state did not answer follow-up requests for additional information, so we were only able to use public information and the short excerpt of the testing documentation when scoring the state's practice.

Oregon (OR) *Proc.:* 4.5 (46th) *Func.:* 0.0 (46th) *Total:* 4.5 (48th)

Oregon conducts all elections by mail and so defines L&A policy in its *Vote by Mail Procedure Manual* [55]. The state's L&A testing is divided into a preparatory test in which all tabulators are required to be tested and a public certification test in which only a sample is tested. Observation of public testing is limited to one representative of each party and each nonpartisan candidate or their designated representative. There is no indication that any of the criteria in basic protections, overvote protection, and nondeterminism are met.

Pennsylvania (PA) *Proc.:* 10.0 (1st) *Func.:* 1.5 (23rd) *Total:* 11.5 (19th)

Pennsylvania has robust testing requirements [56] with an appropriate scope and excellent transparency. The state also has several good functional recommendations, including encouraging jurisdictions to assign a different number of votes to each candidate in a contest. We recommend that Pennsylvania turn its recommended practices into requirements. The state should also introduce elements of nondeterminism into its testing practices.

Rhode Island (RI) *Proc.:* 6.25 (32nd) *Func.:* 0.0 (46th) *Total:* 6.25 (42nd)

Rhode Island's State Board of Elections works in conjunction with the voting equipment vendor to publicly test all tabulators before each election. Its L&A policy [57], however, is very vague regarding our functional criteria, which led

to the state earning no credit for that category. The Board did not respond to our correspondence seeking additional information.

South Carolina (SC) *Proc.:* 6.5 (31st) *Func.:* 1.0 (39th) *Total:* 7.5 (37th)

Even though South Carolina’s State Election Commission provided us with an excerpt of the state’s L&A procedures [60], the document was so heavily redacted that we were not able to obtain any useful information relative to our criteria. Instead, we relied on state code [59], which uses similar language to several other states but is unique in that it can be satisfied by the inclusion of overvoted *or* undervoted ballots and thus fails to meet OP1. The State Election Commission did not responded to our request for additional information.

South Dakota (SD) *Proc.:* 9.25 (5th) *Func.:* 7.0 (2nd) *Total:* 16.25 (2nd)

South Dakota does not have an “approved” statewide document for conducting L&A, instead relying on its statutes [61,62]. Communication with state election officials revealed strong functional requirements. Notably, the state ensures proper overvote validation by requiring, for each contest, that election officials mark the maximum number of allowable votes on a test ballot and then exactly one more vote than what is permissible on another. South Dakota would benefit from requiring election officials to mark test ballots using a source of randomness and by creating publicly-accessible documentation that details statewide requirements.

Tennessee (TN) *Proc.:* 2.5 (50th) *Func.:* 1.5 (23rd) *Total:* 4.0 (49th)

Tennessee’s L&A policy can be found in state law [64,65]. It allows candidate, news media, and (depending on the type of election) political party representatives to observe testing, but not the general public. Notably, the state earned no credit for ST1, ST2, and ST3 because it only requires testing of tabulators in a number of precincts equal to at least 1% of the total number of precincts in an election. The state did not responded to requests for additional information.

Texas (TX) *Proc.:* 6.25 (32nd) *Func.:* 1.5 (23rd) *Total:* 7.75 (35th)

Texas is positioned to greatly strengthen its L&A testing [66] by explicitly requiring the implementation of more rigorous practices. For example, although the state requires testing prior to a given election, it does not require that every device be tested, leaving the devices that are not tested susceptible to preventable errors. The state also requires that every candidate receive at least one vote, but does not ensure that a different number of votes is assigned to each candidate in a contest, thus leaving the possibility of transposed targets untested. Texas does, however, position its policy as a baseline requirement, so local jurisdictions have latitude to perform more comprehensive testing.

Utah (UT) *Proc.:* 9.25 (5th) *Func.:* 5.5 (7th) *Total:* 14.75 (5th)

Although Utah’s testing requirements are currently only set forth by statute [69], the state’s election division is developing a best practices guide during the summer of 2022. Email correspondence revealed that the state already performs well, meeting most of our criteria in both procedural and functional categories. Utah

could earn a perfect score by fully validating overvote thresholds and ensuring that some test ballots are marked using a source of randomness.

Vermont (VT) *Proc.:* 9.0 (16th) *Func.:* 5.5 (7th) *Total:* 14.5 (7th)

Vermont's most recent L&A procedures [71] were produced in June 2022 and require that public testing incorporate all ballot styles and at least one overvote. Notably, the state requires election officials to randomly fill ten ballots while keeping different vote counts for all candidates. Since the state demonstrated an awareness of the importance of marking test ballots in a way that makes the outcome less predictable, we awarded partial credit for ND2, even though the policy does not explicitly require use of a random number generator. Vermont's election division did not respond to a request for additional information.

Virginia (VA) *Proc.:* 9.25 (5th) *Func.:* 1.5 (23rd) *Total:* 10.75 (21st)

Virginia's L&A procedures are outlined in Chap. 4 of its *General Registrar and Electoral Board Handbook* [72]. The state requires each locality to test all of its voting equipment with each ballot style prior to an election. However, its testing does not ensure that transposed targets are detected nor does it ensure that overvote thresholds are fully validated. The test procedures also do not incorporate nondeterministic elements. In addition to addressing these functional issues, Virginia should consider permitting members of the public who are not representatives of a candidate or political party to observe testing.

Washington (WA) *Proc.:* 10.0 (1st) *Func.:* 4.5 (11th) *Total:* 14.5 (7th)

Washington State meets all our procedural requirements. It tests all machines twice, first at a pretest and then again at a public proceeding [74]. For all elections in Washington, voters are allowed to mark only one option. Proper overvote validation for jurisdictions in this state would therefore look like ensuring every contest has at least one ballot that votes for precisely two options. Incorporating this practice would be a simple way to strengthen the state's functional score.

West Virginia (WV) *Proc.:* 9.25 (5th) *Func.:* 1.5 (23rd) *Total:* 10.75 (21st)

West Virginia [75] meets almost all of our procedural criteria. It requires testing of all tabulators prior to an election, permits additional testing in local jurisdictions, and ensures that every ballot style is included in the test deck. The state could strength functional aspects of its requirements by incorporating greater protections offered by BP2 and OP2 as well as by providing nondeterminism.

Wisconsin (WI) *Proc.:* 9.25 (5th) *Func.:* 1.5 (23rd) *Total:* 10.75 (21st)

Wisconsin maintains a state L&A policy that provides for the public testing of all machines with all ballot styles before each election [77]. Functionally, the policy requires at least the simple forms of basic protections and overvote protections. There is also a non-binding recommendation to test "as many vote combinations as possible." We recommend strengthening this so that election officials are required to test a different number of votes for each candidate in a contest and to mark some ballots using a source of randomness. The Wisconsin Elections Commission did not respond to a request for additional information.

Wyoming (WY) *Proc.*: 9.25 (5th) *Func.*: 5.5 (7th) *Total*: 14.75 (5th)

Wyoming has a strong L&A testing policy [78] which, among other provisions, requires that each candidate in a receive a different number of votes. From additional email correspondence with Wyoming's election division, we were able to verify that all tabulators are publicly tested. Wyoming would benefit from consolidating its L&A requirements into one resource and incorporating a source of randomness for marking some test ballots.

References

1. 2 DE Code § 5523 (2014 through 146th Gen Ass) (2014). <https://law.justia.com/codes/delaware/2014/title-15/chapter-55/section-5523/>
2. 21-A Me Rev. stat. § 854 (2011). <https://legislature.maine.gov/statutes/21-A/title21-Asec854.html>
3. 31 KY Admin. Regs. § 2:020 (2021). <https://apps.legislature.ky.gov/law/kar/titles/031/002/020/>. Accessed 15 Dec 2021
4. 950 Mass. Reg. § 54.02 (2022). <https://casetext.com/regulation/code-of-massachusetts-regulations/department-950-cmr-office-of-the-secretary-of-the-commonwealth/title-950-cmr-5400-voting-and-counting-procedures-for-electronic-voting-systems/section-5402-testing>. Accessed 19 Apr 2022
5. Alabama Electronic Voting Committee: Chapter 307-X-1, Procedures For Electronic Vote Counting Systems (2002). <https://www.alabamaadministrativecode.state.al.us/docs/evc/307-X-1.pdf>. Accessed 31 Mar 2002
6. Alaska Division of Elections: Democracy Suite Election Check List: Regional Offices (2020). Obtained via private communication. Accessed 14 Sept 2020
7. Alaska Division of Elections: LAT Testing for Early Vote: Absentee Voting Tablet HP (2021). Obtained via private communication. Accessed 19 Mar 2021
8. Alaska Division of Elections: LAT Testing for Precinct Scanner and Voting Tablet HP (2021). Obtained via private communication. Accessed 1 Mar 2021
9. Alaska Division of Elections: LAT Testing for Voting Tablet VVPAT and Precinct Scanner (2021). Obtained via private communication. Accessed 19 Mar 2021
10. Ansolabehere, S., Hersh, E.: The quality of voter registration records: a state-by-state analysis (2010). <https://www.vote.caltech.edu/reports/6>
11. Arizona Secretary of State: Arizona Elections Procedures Manual (2019). https://azsos.gov/sites/default/files/2019_ELECTIONS_PROCEDURES_MANUAL_APPROVED.pdf. Accessed Dec 2019
12. Arkansas State Board of Election Commissioners: County Board of Election Commissioners Procedures Manual (2022). https://static.ark.org/eeuploads/elections/2022_CBEC_Manual_FINAL.pdf
13. Ballotpedia: Michigan official sample ballots (2020). https://ballotpedia.org/Michigan_official_sample_ballots,_2020
14. Bartlett, B.: Report on the analysis of voting machine issues in Northampton county, county of Northampton: office of the county executive (2019). <https://www.northamptoncounty.org/CTYEXEC/Documents/121219%20Press%20Conference%20Recap.pdf>. Accessed 12 Dec 2019
15. Bernhard, M., et al.: Public evidence from secret ballots. In: Krimmer, R., Volkamer, M., Braun Binder, N., Kersting, N., Pereira, O., Schürmann, C. (eds.) E-Vote-ID 2017. LNCS, vol. 10615, pp. 84–109. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68687-5_6

16. Bowler, S., Donovan, T.: A partisan model of electoral reform: voter identification laws and confidence in state elections. *State Polit. Policy Q.* **16**(3), 340–361 (2016)
17. Brown, E., Gardner, A.: Georgia county under scrutiny after claim of post-election breach, *The Washington Post* (2022). <https://www.washingtonpost.com/investigations/2022/05/13/coffee-county-misty-hampton-election/>. Accessed 13 May 2022
18. California Elections Code, Division 15: Semifinal Official Canvass, Official Canvass, Recount, and Tie Vote Procedures; Chapter 2: Vote by Mail Ballot Processing. (2022) https://leginfo.ca.gov/faces/codes_displayText.xhtml?lawCode=ELEC&division=15.&title=&part=&chapter=2.&article=. 1 Jan 2022
19. Code of Maryland Regulations: 33.10.01.14-16. <https://www.dsd.state.md.us/COMAR/ComarHome.html>
20. Colorado Secretary of State: C.R.S. Title 1 Elections, Article 7 Conduct of elections (2021). https://www.sos.state.co.us/pubs/info_center/laws/Title1/Title1Article7.html
21. CT Gen. Stat. § 9-247 (2012). <https://law.justia.com/codes/connecticut/2012/title-9/chapter-147/section-9-247/>
22. Delaware Department of Elections: Ballot Accuracy Procedures. <https://www.muckrock.com/foi/delaware-236/ballot-accuracy-procedures-delaware-126706/?#files>. Obtained via open records request
23. Estep, T.: DeKalb elections officials offer more insight into May primary issues. *Atlanta J. Const.* (2022). <https://www.ajc.com/neighborhoods/dekalb/dekalb-elections-officials-offer-more-insight-into-may-primary-issues/KJPO7DK5KFD77KMQXXSWO2MXSA/>. Accessed 14 July 2022
24. Feldman, A.J., Halderman, J.A., Felten, E.W.: Security analysis of the Diebold AccuVote-TS voting machine. In: *USENIX Workshop on Electronic Voting Technology (EVT)* (2007), August 2007
25. Florida Division of Elections. Guidelines for logic and accuracy (L&A) testing: DE reference guide 0019 (2020). https://soe.dos.state.fl.us/pdf/DE%20Guide%200019%20-%20Guidelines%20for%20Logic%20%20Accuracy%20Testing_rev20200709.pdf. Accessed July 2020
26. Georgia Secretary of State: Secure the Vote: Logic and Accuracy Procedures (2020). <https://archive.org/download/logic-and-accuracy-procedures-v-1-02-2020-stv-1/Logic%20and%20Accuracy%20Procedures%20v1.02-2020%28STV%29%20%281%29.pdf>. Accessed Jan 2020
27. Green, R.: Election observation post-2020. *Fordham Law Rev.* **90**(2) (2021)
28. Halderman, J.A.: The Antrim county 2020 election incident: an independent forensic investigation. In: *31st USENIX Security Symposium* (2022)
29. Hawaii Office of Elections: Counting center manual (2020). <https://elections.hawaii.gov/wp-content/uploads/2020-Counting-Center-Manual.pdf>
30. Help America Vote Act of 2002, Pub. L. 107-252, 116 Stat. 1666
31. Hursti, H., Lindeman, M., Stark, P.B.: New Hampshire SB 43 Forensic Audit Report. (2021). <https://www.doj.nh.gov/sb43/documents/20210713-sb43-forensic-audit-report.pdf>. Accessed July 2021
32. IA Code Title 2: Elections and Official Duties (2022). <https://www.legis.iowa.gov/law/iowaCode/chapters?title=II&year=2022>
33. Illinois State Board of Elections, Division of Voting and Registration Systems: State of Illinois Voting System: Testing & Security Best Practices, Obtained via private communication
34. Indiana Election Division: Indiana Election Administrator’s Manual. <https://www.in.gov/sos/elections/files/2022-Election-Administrators-Manual.FINAL.pdf>

35. Iowa Secretary of State: Election Security in Iowa (2018). <https://sos.iowa.gov/electionsecurityiniowa.html>
36. KS Stat §25-4411 (2016). <https://daw.justia.com/codes/kansas/2016/chapter-25/article-44/section-25-4411>
37. Lindeman, M., Stark, P.B.: A gentle introduction to risk-limiting audits. *IEEE Secur. Priv. Spec. Issue Electron. Voting* (2012). <https://www.stat.berkeley.edu/~stark/Preprints/gentle12.pdf>. Accessed 16 Mar 2012
38. Louisiana Sec. of State: State of Louisiana Election Code (2021). <https://www.sos.la.gov/ElectionsAndVoting/PublishedDocuments/ElectionCode.pdf>. Accessed Oct 2021
39. Michigan Department of State: Bureau of Elections: Test Procedure Manual For Tabulators and Voter Assist Terminals (VAT) (2019). https://www.michigan.gov/-/media/Project/Websites/sos/01vanderroest/TEST_DECK_MANUAL05.pdf
40. Minnesota Secretary of State: 2014 Minnesota Voting Equipment Testing Guide (2014). <https://www.sos.state.mn.us/media/1108/2014-equipment-testing-guide.pdf>
41. Mississippi Secretary of State: Mississippi County Election Handbook (2015). <https://www.sos.ms.gov/content/documents/Elections/Mississippi%20%20County%20Election%20Handbook.pdf>. Accessed Sept 2015
42. Mo. Code Regs. 15 §30–10.140 (2022). <https://casetext.com/regulation/missouri-administrative-code/title-15-elected-officials/division-30-secretary-of-state/chapter-10-voting-machines-electronic/section-15-csr-30-10140-electronic-ballot-tabulation-counting-preparation-and-logic-and-accuracy-testing-dres-and-precinct-counters>. Accessed 2 May 2022
43. Montana Secretary of State: Uniform Ballot and Voting Systems Procedures Guide: Security, Testing, Inventory, and Troubleshooting. <https://sosmt.gov/wp-content/uploads/Uniform-Voting-Systems-Guide.pdf>
44. N.D. Cent. Code § 16.1-06-15 (2011). <https://casetext.com/statute/north-dakota-century-code/title-161-elections/chapter-161-06-ballots-voting-machines-electronic-voting-systems/section-161-06-15-mandatory-testing-of-electronic-voting-systems-before-each-election-and-after-tabulation-of-ballots>. Accessed 1 Aug 2011
45. National Conference of State Legislatures: Election Administration at State and Local Levels (2020). <https://www.ncsl.org/research/elections-and-campaigns/election-administration-at-state-and-local-levels.aspx>. Accessed 3 Feb 2020
46. Nebraska Revised Statute 32-1049 (2007). <https://nebraskalegislature.gov/laws/statutes.php?statute=32-1049>
47. Nevada Revised Statutes: Title 24, Chapter 293B. <https://casetext.com/statute/nevada-revised-statutes/title-24-elections/chapter-293b-mechanical-voting-systems-or-devices/testing-of-equipment-and-programs>
48. New Hampshire Department of State: New Hampshire Election Procedure Manual: 2020–2021. <https://sos.nh.gov/media/kzupydju/epm-2020.pdf>
49. New Mexico Secretary of State: Election Handbook of the State of New Mexico (2021). <https://www.sos.state.nm.us/wp-content/uploads/2021/10/NM-Election-Handbook-SOS.pdf>
50. New York State Board of Elections: State of New York: 2021 Election Law (2021). <https://www.elections.erie.gov/PDFs/2021ElectionLaw.pdf>
51. NJ Rev. Stat. §19:53A-8 (2014) (1975). <https://law.justia.com/codes/new-jersey/2014/title-19/section-19-53a-8/>

52. North Carolina State Board of Elections: Preparing for Accurate Elections. <https://www.ncsbe.gov/about-elections/election-security/preparing-accurate-elections>
53. Ohio Secretary of State: Ohio Election Official Manual. https://www.sos.state.oh.us/globalassets/elections/directives/2022/eom/eom_fullversion_2022-02.pdf. Accessed 2 Feb 2022
54. Oklahoma State Election Board: Oklahoma Uniform Election Reference. <https://archive.org/download/page-from-uniform-election-reference-200401/Pages%20from%20Uniform%20Election%20Reference%20200401.pdf>. Obtained via open records request
55. Oregon Secretary of State: Vote by Mail Procedures Manual. https://sos.oregon.gov/elections/Documents/vbm_manual.pdf. Accessed Jan 2022
56. Pennsylvania Department of State: Directive on Logic & Accuracy Testing (2020). https://www.dos.pa.gov/VotingElections/OtherServicesEvents/Documents/PADOS_Directive_Logic_Accuracy%20with%20attestation.pdf
57. RI Gen. L. § 17-19-14 (2014) (1996). <https://law.justia.com/codes/rhode-island/2014/title-17/chapter-17-19/section-17-19-14/>
58. Root, D., Kennedy, L., Sozan, M., Parshall, J.: Election security in all 50 states, center for American progress report (2018). <https://www.americanprogress.org/article/election-security-50-states/>
59. SC Code §7-13-1390 (2019). <https://www.scstatehouse.gov/code/t07c013.php>
60. South Carolina Election Commission: South Carolina Election Preparation Results and Accumulation Guide (2019). https://archive.org/2/items/2021-10-18-sec-eprag-13-17-redacted/2021-10-18%20SEC%20EPRAG%2013-17_Redacted.pdf. Obtained via open records request
61. South Dakota Secretary of State: South Dakota Administrative Rule (2021). <https://sdsos.gov/elections-voting/assets/2021ARSD.pdf>. Accessed 6 Oct 2021
62. South Dakota Secretary of State: South Dakota Election Code (2021). <https://sdsos.gov/elections-voting/assets/2021ElectionCode.pdf>. Accessed 17 Aug 2021
63. Statutes of California: California Assembly Clerk Archive (1907). https://clerk.assembly.ca.gov/sites/clerk.assembly.ca.gov/files/archive/Statutes/1906_07/1907.pdf#page=691
64. Tenn. Code §29-9-105 (2020). <https://law.justia.com/codes/tennessee/2020/title-29/chapter-9/section-29-9-105/>
65. Tenn. Rules of Sec. of State, State Coordinator of Elections: Ch. 1360-2-1 (1999). <https://publications.tnsosfiles.com/rules/1360/1360-02/1360-02-01.pdf>
66. Tex. Elec. Code §1.001 (2021). <https://statutes.capitol.texas.gov/Docs/SDocs/ELECTIONCODE.pdf>
67. Thompson, D.M., Wu, J.A., Yoder, J., Hall, A.B.: Universal vote-by-mail has no impact on partisan turnout or vote share. *Proc. Natl. Acad. Sci.* **117**(25), 14052–14056 (2020). <https://doi.org/10.1073/pnas.2007249117>
68. U.S. Election Assistance Commission: Election Audits Across the United States (2021). https://www.eac.gov/sites/default/files/bestpractices/Election_Audits_Across_the_United_States.pdf. Accessed 6 Oct 2021
69. Utah State Code 20A-4-104: Counting ballots electronically (2022). https://le.utah.gov/xcode/Title20A/Chapter4/C20A-4-S104_2022050420220504.pdf. Accessed 4 May 2022
70. Verified Voting. The Verifier: Polling Place Equipment. <https://www.verifiedvoting.org/verifier/>

71. Vermont Secretary of State: Elections Division: Vermont Vote Tabulator Guide (2022). <https://outside.vermont.gov/dept/sos/Elections%20Division/town%20clerks%20and%20local%20elections/election%20supplies/vermont-vote-tabulator-guide.pdf>
72. Virginia Department of Elections: General Registrar and Electoral Board Handbook - Chapter 4 (2022). [https://www.elections.virginia.gov/media/grebhandbook/2022-updates/4_Voting_Equipment_\(2022\).pdf](https://www.elections.virginia.gov/media/grebhandbook/2022-updates/4_Voting_Equipment_(2022).pdf)
73. Voting Rights Act of 1965, Pub. L. 89-110, 79 Stat. 437
74. Washington Secretary of State: Presidential Primary Advisory 2020 #06 - Logic and Accuracy Testing (2020). https://www.sos.wa.gov/_assets/elections/06%20&a%20testing_2020b.pdf. Accessed 21 Jan 2020
75. West Virginia Code: Ch. 3, Ar. 4, §26. <https://code.wvlegislature.gov/3-4A-26/>
76. Wheaton, S.: Comparing Ballot Lengths Around the Country, *The New York Times* (2013). https://archive.nytimes.com/thecaucus.blogs.nytimes.com/2013/02/06/comparing-ballot-lengths-around-the-country/?_r=0. Accessed 6 Feb 2013
77. Wisconsin Elections Commission: Election Administration Manual for Wisconsin Municipal Clerks (2020). <https://elections.wi.gov/sites/elections/files/2022-03/Election%20Administration%20Manual%20%282020-09%29.pdf>
78. Wyoming Admin. Rules: Sec. of Stat —Election Procedures— Ch. 29: Voting Systems (2020). https://rules.wyo.gov/DownloadFile.aspx?source_id=17964&source_type_id=81&doc_type_id=110&include_meta_data=Y&file_type=pdf&filename=17964.pdf&token=12222099183083056006234136082008020226028193044
79. Yu., J.: Does state online voter registration increase voter turnout? *Soc. Sci. Q.* **100**(3) (2019). <https://onlinelibrary.wiley.com/doi/abs/10.1111/ssqu.12598>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

