Federal Certification Test for Vote-Counting Accuracy Cannot Determine the Error Rate of the Equipment

By Ellen Theisen, VotersUnite.Org *March 21, 2009, revised March 30, 2009*[1]

We live in a world of complex computerized systems -- medical equipment such as MRIs and CAT scans; flight control software; an automotive fuel injection system; databases that store, retrieve, and collate data. All these systems undergo stringent testing.

The primary purpose of any testing of computerized systems is to show either that the system works as designed or that it doesn't. If the system does **not** work as designed, then the test results are expected to provide information on the nature, and perhaps the cause, of the defective behavior.

Using these criteria, this report evaluates one specific test -- a test designed to confirm or refute that a specific computerized vote-counting system meets the accuracy level required by federal law. The details that follow contain many numbers and many technical details, but the discussion centers on a single, simple question: did the test, which was approved by the federal agency tasked with certifying voting equipment, provide evidence that the accuracy requirement was met, or did the test provide evidence that the accuracy requirement was not met?

As the author demonstrates in this specific instance, the answer is: the test provided neither. The test, which was designed by a federally-accredited test lab for the sole purpose of testing the accuracy of the vote-counting equipment, **cannot** prove whether or not the system accurately counts votes, nor can the test yield any measure of the accuracy rate.

The author wants the reader to understand that such testing provides false assurance that a federally-certified vote-counting machine meets the accuracy requirements of federal law.

The Requirements of Federal Law

Federal law mandates that voting systems meet a very stringent standard for accuracy in counting votes. The Help America Vote Act of 2002 (HAVA), Section 301(a)(5) states:[2]

> *The error rate of the voting system in counting ballots … shall comply with the error rate standards established under section 3.2.1 of the voting systems standards issued by the Federal Election Commission which are in effect on the date of the enactment of this Act.*

The error rate standard referenced in the law is "a target error rate of no more than one in 10,000,000 ballot positions, with a maximum acceptable error rate in the test process of one in 500,000 ballot positions."[3] A "**ballot position**," also called "**vote position**" is any choice presented to the voter, such as a single candidate in a contest or a single response to an issue. For example, a ballot with 10 contests and 3 candidates in each contest would have 30 vote positions.

While the maximum legal error rate of 0.00001% may be unrealistic, it is law. However, it is rare that any officials analyze the error rate of a voting system used in a real election. So, for practical purposes, the error rate allowed for the test process is the acceptable rate – 0.0002% – still a very high standard of accuracy. Testing properly to such a high standard would be extremely time consuming.

HAVA tasked the U.S. Election Assistance Commission (EAC) with establishing a process for certifying that a voting system meets a set of standards (which states may voluntarily adopt or not), including the accuracy standard. But the accuracy standard is not voluntary; federal law requires it. Therefore it is logical to assume that the EAC would pay very close attention to the test process to ensure that it proved the maximum error rate was not exceeded.

It is also reasonable to assume that a national testing lab accredited by the EAC would treat federal law very seriously and pay special attention to ensuring that its test plan for the accuracy of any voting system could prove whether or not the system yielded an error rate below the legal maximum.

However, neither of these assumptions is correct.

A Hypothetical Test Plan

Consider a test plan in which 2,000 ballots are counted, and every candidate in every race receives the exact same number of votes. **Even if the actual test results match the expected results, the error rate cannot be known.**

- The test engineer has no way of knowing whether the votes for Candidates A and B in Race Z were counted correctly or switched with each other.
- Further, the test engineer has no way of knowing whether the votes for Candidate A in Race Z were counted correctly or whether they were switched with the votes for Candidate C in Race Y.

The error rate might be 0.0%. But if the system made multiple tabulation errors that canceled each other out in the totals, the actual test results would match the expected results, and yet the test would have failed to yield information about the true error rate.

The Test Plan for the Premier Assure 1.2 OSX Optical Scanner

Now consider a test plan in which 792 ballots are counted. Each ballot has 30 contests with 66 candidates in each contest for a total of 1980 candidates in all contests, so there are 1980 vote positions on the ballot; and all but 200 of those candidates all receive the same number of votes as other candidates. It would be impossible for a test engineer to know whether the votes for those 1780 candidates had been accurately counted or whether they had been switched with other candidates who received the same totals.

Since votes in real elections have been switched to opposing candidates,[4] it is obvious that the equipment must be tested for just such a situation. Yet, SysTest, a national test lab accredited by the EAC, designed a plan similar to the one described in this scenario, and the EAC approved the plan. SysTest executed the plan on the Premier Assure 1.2 AccuVote OSX optical ballot scanner and claimed the system passed accuracy testing.

SysTest explained its methodology for determining accuracy as simply: "Review the outcome of the test(s) against the expected result(s)."[5] Neither the test plan nor the test report describes any method of verifying that the machine correctly counted each voted and non-voted voting target area on each ballot. It appears that the lab merely compared the actual tabulated results to the expected results and concluded that the votes had been counted accurately.

Before all the testing was complete, the EAC removed SysTest's accreditation and another national test lab, iBeta, took over. iBeta reviewed SysTest's documentation and recommended accepting SysTest's accuracy testing. The EAC agreed:

> *After a careful technical review of iBeta's review and audit the EAC approves the above recommendations made by iBeta.*[6]

For those interested in details of the test plan, they are presented below, along with the calculations that led to the author's easily reproducible conclusion that **at least 89.5% of the candidates received the same number of votes as other candidates.** If SysTest did nothing more than compare actual results to expected results (as their test plan indicates), the testing evaluated the ability of the OSX to accurately count **at most 10.5%** of the ballot positions.

This design, approved by the EAC, **cannot** ensure that the error rate was less than one vote in 500,000 ballot positions. In fact, there is no way of determining the number of errors that occurred in a test process using this design. The EAC's certification of the system would, without any justification, assure jurisdictions purchasing the AccuVote OSX ballot scanner that the equipment complies with federal accuracy requirements.

Details of the OSX Test Plan

A description of the AccuVote OSX test plan, formulated and used by SysTest, and subsequently approved by iBeta and the EAC, is provided in the latest version of the SysTest test plan.[7]

# of Contests	***30***
# of Candidates	***66***
Ballots per batch	***66***
Vote Positions per ballot	***1,980***
Vote Positions per batch	***130,680***
# of batches	***12***
# of times a batch is run	***1***
Total Vote Positions scanned	***1,568,160***

SysTest's test plan gives no information about how the 792 ballots are marked, so it is necessary to make some assumptions. Let us assume:

1) All contests are "vote for one" contests. The defect in the plan is equally severe if some or all contests are multi-vote contests.

2) No contest is overvoted, that is, no ballot has more than one candidate marked in a given contest. The defect in the plan is equally severe if the plan includes some overvoted ballots.

So, if all contests are "vote for one" and no contest is overvoted:

♦ Since there are 30 contests, each ballot will have at most 30 ballot positions marked (one per contest). All other ballot positions will be blank.

♦ Since each contest has 66 candidates, 66 ballots are needed to allow each position of each contest to be tested once, with each position having one vote.

- If the votes on those 66 ballots are arranged optimally, 726 (792-66) ballots remain to allow some positions to receive more than one vote in the totals. (12 batches multiplied by 66 ballots equals 792 ballots.)
- However, it would take an additional 2,145 ballots to ensure that each ballot position for a particular contest has a unique number of votes within that contest. 66 votes for one position, 65 for the next, 64 for the next, and so on. [67*66/2=2,211 and 2,211-66=2,145]
- With only 792 ballots, a maximum of 39 ballot positions can be unique numbers. [39*40/2=780] That means **at least 27 of the 66 ballot positions in each contest received the same number of votes as other positions.** There is no way of knowing if the votes for those 27 positions were counted for the correct positions or for other ballot positions.

But the defect in the test plan is even worse than that. In real elections, votes from one contest have been given to another contest,[8] and it is essential to ensure that does not happen. Even the 2,211 ballots required to ensure that each ballot position within a given contest has a unique total are not sufficient to ensure that votes for one contest are not being switched with votes for another contest.

- With 1,980 ballot positions on each ballot, at least 1,961,190 ballots are required to ensure that each ballot position has a unique result. First position has 1980 votes, second has 1979 votes, third has 1978 votes, and so on. [1981*1980/2]
- But with 30 contests and only 792 ballots, a maximum of 208 positions can have unique totals. **At least 89.5% (1772/1980) of the ballot positions received the same number of votes as other positions,** and there is no way of knowing if the votes for those positions were counted for the correct positions or for other ballot positions.

The testing requirements of the voting system standards state that:

> *If the system makes one error before counting 26,997 consecutive ballot positions correctly, it will be rejected. The vendor is then required to improve the system.*
>
> *If the system reads at least 1,549,703 consecutive ballot positions correctly, it will be accepted.*
>
> *If the system correctly reads more than 26,997 ballot positions but less than 1,549,703 when the first error occurs, the testing will have to be continued until another 1,576,701 consecutive ballot positions are counted without error (a total of 3,126,404 with one error).*[9]

With a maximum of 208 ballot positions having unique results for a total of 792 ballots, a successful test may be able to provide a high degree of assurance that the system has read 164,736 (792*208) ballot positions correctly – but not 164,736 **consecutive** ballot positions, and not even close to the required 1,549,703 consecutive ballot positions.

Conclusion

Since the total votes for at least 89.5% of the candidates are the same number as the total for other candidates, and SysTest appears to be evaluating only the vote totals, the test cannot prove that every vote accrued for the intended candidate. Multiple, possibly unrelated, tabulation errors could have canceled each other in the totals and thus remained undetected.

Mistakes such as shifting candidate votes within a race or among races have been reported in real elections involving real candidates. A test plan that does not detect ballot-scanning errors like those that are known to have occurred in the field fails in its primary mission: to confirm or refute that the vote-counting system under test meets the accuracy requirement of federal law. The EAC-approved test did not yield the essential information for which the test was intended.

The question to be answered by the EAC-approved accuracy testing is simply this: "Does this vote-counting machine count votes accurately?" Before the test is run, the answer to the question is unknown. After the test is run, the answer to the question is still unknown. The **entire purpose** of the EAC-approved accuracy test was to answer this question, and since the question was never answered, the test was useless.

The EAC-approved test was a waste of time and money. Worse, the test gave false assurance that the vote-counting machine meets the HAVA-mandated accuracy rate, even though the actual accuracy rate remains completely unknown.

Thanks to John Washburn, Certified Software Test Engineer, whose review comments improved this report. Thanks to Barbara Simons for pointing out the calculation error that led to the March 30 revision.

[1] The original version of this paper contained an approximation and a miscalculation. The number of unique totals in a single contest with 66 candidates is 39, but had been approximated to 40. And because there are 30 separate contests with 66 candidates each, the maximum number of unique totals for all 1980 ballot positions is 208, not 40. The correct numbers are reflected in the current version but do not change the conclusion of the analysis, except that the maximum percentage of unique tallies is 10.5% rather than 2%.

[2] "Help America Vote Act of 2002." http://www.fec.gov/hava/law_ext.txt

3 "Voluntary Voting System Standards, 2002. Volume I." Section 3.2.1. Page 3-52 (pdf page 60). http://www.eac.gov/voting%20systems/voluntary-voting-guidelines/docs/voting-systems-standards-volume-i-performance.pdf/attachment_download/file .

4 One of many examples of machines counting votes for an opponent occurred in Pottawattamie County, Iowa in 2006. "Faulty voting machines delay results; counting under way." The Daily Nonpareil Online. June 7, 2006. by Tim Rohwer, Staff Writer. Archived at: http://www.votersunite.org/article.asp?id=6571

5 "Certification Test Report." Report Number 06-V-DB-058-CTR-01. Premier Election Solutions Assure 1.2 Voting System. Prepared by SysTest Labs. February 22, 2008. http://www.eac.gov/files/voting_systems/Premier%20Assure%201.2%20Certification%20Test%20Report%20Rev01.pdf . Page 85.

6 "EAC letter to iBeta on reuse of prior testing conducted by SysTest Laboratories." http://www.eac.gov/program-areas/voting-systems/docs/2-10-09-approval-reuse-of-testing-data-accuracy-final/attachment_download/file . Page 1.

7 "Premier Assure v1.2 Rev 11 Draft Test Plan" by SysTest. http://www.eac.gov/program-areas/voting-systems/docs/premier-assure-1-2-rev-11-test-plan.pdf/attachment_download/file . Page 35. An email sent by James Nilius of SysTest to Matt Masterson of the EAC and forwarded to the author defines this same ballot design as the one used for the AccuVote TSX touch screen voting machine. iBeta recommended redoing SysTest's testing of the TSX because the system had been set up wrong – not because the test plan was flawed.

8 "Arkansas Election Officials Baffled by Machines that Flipped Race." Wired Blog. By Kim Zetter May 29, 2008. http://blog.wired.com/27bstroke6/2008/05/arkansas-voting.html

[9] "Voluntary Voting System Standards, 2002. Volume II." Page 4-60 (pdf page 64). http://www.eac.gov/voting%20systems/voluntary-voting-guidelines/docs/voting-systems-standards-volume-ii-test.pdf/attachment_download/file.