Using Recounts to Measure the Accuracy of Vote Tabulations: Evidence from New Hampshire Elections 1946-2002[1]

Stephen Ansolabehere
Department of Political Science, MIT

Andrew Reeves
Departments of Government, Harvard University

January, 2004

[1]Correspondence should be directed to Stephen Ansolabehere—email: sda@mit.edu. Mailing address: Department of Political Science, MIT, Building E53-457, Cambridge, MA 02139. We wish to thank the Carnegie Foundation and the Knight Foundation for their support of this research.

The 2000 presidential election exposed a surprisingly high level of inaccuracy in the tabulation of ballots. Differences between total ballots cast and votes counted were as high as 19 percent in some counties in Florida, and these discrepancies were widely attributed to the ballot formats, the handling of ballots, and machine operations.[1] For those involved in the administration of elections the recount was particularly troubling. Over the last 40 years the United States has introduced new technologies, especially punch card and optically scanned ballots, to improve vote tabulations. The problems revealed in Florida suggested that these newer technologies may not in fact represent an improvement over traditional hand-counted paper ballots.

A number of important studies of the performance and accuracy of voting technologies have sought to measure the error rate of vote tabulations. The main metric that emerges from these evaluations uses "residual votes" – the discrepancy between total ballots cast and votes cast for a particular office, such as president or governor. The incidence of residual votes should be unrelated to the type of technology used, and the difference in residual votes across technologies measures the extent to errors in the casting or tabulation of votes are attributable to specific the technology. Similar jurisdictions using different technologies ought to have the same residual vote rate, on average. By this metric, hand-counted paper ballots and optically scanned paper ballots have shown the better overall performance than punch cards, lever machines, and electronic voting machines.[2]

We examine a second measure of accuracy – the agreement between initial counts and recounts of ballots in contested elections. We term this the *tabulation validation rate*. Likewise, the discrepancy between the initial count and the recount is the tabulation error rate.

[1]The NORC coding of the ballots found that the most common problem with optical scanned ballots involved people voting for a candidate, say Bush, and then writing in that candidates name in the columns labeled "Write In." The machine tabulator failed to count these ballots, though voter intention was easily resolved.

[2]See Caltech/MIT Voting Technology Projecct Report, no. 1 (2001) and Ansolabehere and Stewart (2002) for further definition and discussion of residual votes. Studies using similar measures include Knack and Kropf (2001), whose results are consistent with the Caltech/MIT study, and Brady et al (2002), who find similar problems with punch cards but suggest somewhat better performance for electronic equipment.

When there is a legal challenge to an election or a mandated recount, state or local election offices conduct an audit of votes cast. For paper ballot systems, election officials reexamine the ballots to determine voter intention. Tabulations may change from the initial count to the recount for a variety of reasons: ballots may be mishandled; machines may have difficulty reading markings; people and machines may make tabulation errors. Because recounts are used to certify the vote, greater effort is taken to arrive at the most accurate accounting of ballots cast. The initial count of ballots, then, is treated as a preliminary count, and the recount as the official count. The recount, then, validates the initial tabulation.

Recounts allow us to see more precisely whether the introduction of voting machines has improved the tabulation of votes. Do machines have higher validation rates (and lower invalidation rates) than hand-counts?

We examine data on recounts for New Hampshire towns. There are several important reasons for studying New Hampshire. First, New Hampshire uses hand-counted paper ballots extensively as well as optically scanned ballots. This allows us to contrast new and old methods of counting directly. Second, New Hampshire has a large number of elections. Because recounts are rare, the high frequency of elections means that we have a large number of cases to study. Third, New Hampshire has a uniform reporting system for recounts dating back at least to the 1940s. The data on recounts are, then, comparable and the historical record allows us to establish a solid baseline against which to contrast new tabulation methods.

Two caveats accompany the use of recounts to measure the validity of tabulations. First, this measure only concerns tabulation. It ignores other factors, such as voter confusion about how to mark a machine-readable ballot, which might lead people to vote accidentally twice for an office or not at all. Such phenomena are captured by residual votes.

The metric differs for paper ballot systems and elections in which votes are cast on lever machines or electronic voting machines (also known as Direct Recording Electronincs, or DREs). Mechanical and electronic voting machines do not retain a separate record of the voters' intention, so it is impossible, at least with current technology, to compare voters'

intentions with the machines' recordings. Recounts with mechanical and electronic voting machines merely capture whether whether the election office made a recording error or whether the machine is functioning. With mechanical and electronic machines it is impossible to gauge the degree of malfunction or the disparity between voter intentions and machine recordings.

Historical Recounts

Historical recounts provide an important baseline. Over a long period of time across many offices, we can examine the percent difference between initial counts and recounts.

We examine historical recount data in the state of New Hampshire from 1946 to 1962. The first use of punch card machinery in the United States occurs in 1964. In the general elections in this span, 108 races had recounts. The offices involved range from registrar of voters to representative of Congress as well as town questions. The majority of the recounts (69) occurred in races for state representative. All recounts are aggregated to the town level except those for Congressional races and state senate races which were aggregated to district levels.

These data provide complete and consistent information about the accuracy of hand counted votes for 9 elections over sixteen years. Since all ballots during this time were hand counted once in the initial count and again in the recount we are able to evaluate accuracy rates of the counts independent of residual votes.

Our analysis consists of 415 cases where a case is a reported ballot tabulation of a town or (district) for each candidate running in a race. We compute the invalidation rate for each candidate for each office in each jurisdiction. That is, we compute the percent difference between the initial count and the recount for the votes recorded in each jurisdiction (i.e., town or ward) for each candidate seeking an office.

Two different average invalidation rates are of interest. First, the simple average invali-

dation reflects what happens in a typical jurisdiction. In a state with many small towns and a handful of larger towns, the typical jurisdiction will be a small one. Second, the weighted average, weighting by total ballots cast, equals the percent of ballots cast that differed between the counts. This measure guards against the occasional aberrant tabulation in a small community. The weighted average measures the frequency with which one's vote is counted in the initial count with each type of equipment. We will focus on the latter, but will report both.

Historically, hand-counted paper tabulations have an invalidation rate just under 1 percent. The weighted average of the discrepancy between the initial count and the recount is approximately 0.83 percentage point with a 95% confidence interval from 0.61 to 1.05.

Hand-Counts vs. Machine-Counts

Fast forward to 2002. While most counties and towns continue to use hand-counted paper, many of the towns in three counties had shifted to optically scanned paper ballots. Two sorts of scanners were used: Election Systems and Software's Optech and Global's (now Diebold's) Accuvote.[3]

Six races were recounted following the 2002 general election. Data on the recount are reported at the town and sometimes ward level. Again, we distinguish between the average percent difference and the percent of all ballots cast (regardless of size of jurisdiction).

Our analysis here considers the changes between the initial counts and recounts for each method of vote tabulation. Again, we compute a weighted mean of the average absolute percent difference between the initial count and the recount. We present the hand-counted data with and without the town of Bradford, where the initial tabulations differed from the recount by as much as 20 percent.

[3] Voting machinery of New Hampshire towns obtained from http://www.nh.gov/sos/voting%20machines.htm

Table 2
Tabulation Error Rates of Hand-Counted and Scan Ballots
New Hampshire 2002 General Election

	N	Weighted Average	95 % Confidence Interval	Maximum*
Hand-Counted	92	2.49	(1.44, 3.55)	21.88
Hand-Counted (without Bradford)	79	0.87	(0.69, 1.06)	4.08
Machine Counted	217	0.56	(0.42, 0.70)	8.27
OPTECH	21	0.55	(0.40, 0.70)	1.36
ACCUVOTE	196	0.56	(0.41, 0.71)	8.27

* The minimum value is 0 in all cases.

The results for hand-counted paper, excluding Bradford, are nearly identical to the historical average. We take this as the baseline. The percentage difference for optically scanned paper was .56 percent, approximately one half of one percent and significantly lower than the hand-counted paper.

The differences in the table may owe to the type of equipment or the size of the community. Smaller communities are more likely to have a large discrepancy in their tabulations and they are more likely to use paper ballots.

We can untangle these effects by controlling for the office at stake and the size of the vote cast. We performed two regression analyses to estimate the effect of *ScanBallot*, which equals 1 if the town used optically scanned ballots and 0 if the town used hand-counted paper ballots. We control for total votes cast in a town and the initial count for a candidate. We also control for the office at stake, as some elections spanned several towns. This allows us to hold constant the candidates on the ballot. Table 2 presents two analyses, one of which removes the effect of office (see column labeled Fixed Effects) and one does not (see column labeled OLS). We present both as a robustness check.

Table 2 **Estimated Effect of Machine v. Hand Count on Tabulation Error Rate, 2002**		
	OLS Regression **b** (S.E.)	Fixed Effects For Office **b** (S.E.)
Constant	.825 (.113)	.800 (.113)
Scan Ballot	-.450 (.170)	-.523 (.171)
Town Vote (in 1000's)	.014 (.004)	.012 (.004)
Candidate's Initial Vote (in 1000's)	-.072 (.095)	.063 (.106)
R^2	.05	.08
N	306	306

The key coefficient of interest is the effect of *ScanBallot*. The coefficient in both specifications is approximately -.5, which means that the percentage difference between the initial count and the recount is approximately one half of one-percentage point lower in towns using scanners than in towns using hand-counted paper, holding constant the initial count, the total vote, and the office sought. These coefficients are slightly larger than the observed difference between the mean for hand-counted and machine counted (which is .87-.56 = .31).

To see the practical effect of these estimates consider an election with 10,000 votes where the candidate in question received exactly half of the votes.[4] The predicted discrepancy between counts is 1.24 percent if the town tabulates by hand (i.e., $.80 + .012 \times 10 + .063 \times 5 = 1.24$) and .71 percent if the town uses an optical scanner (i.e., $.80 + .012 \times 10 + .063 \times 5 = .712$).

[4]We caution against making out of sample predictions. The largest community in 2002 was Manchester, which recorded approximately 25,000 votes. The effect of town vote likely tends to an asymptote and the linear specification used cannot capture the predicted discrepancy for larger communities.

Conclusions

Recounts provide information on the reliability of voting tabulation methods. In recounts of paper ballots, election officials and judges make the best attempt to resolve voter intentions. The discrepancy between the initial (and preliminary) count and the recount captures the degree to which a tabulation method incorrectly tabulates the votes. Estimated validation rates from recounts supplements other metrics, especially residual votes. Residual votes capture a wide variety of ways that voting systems fail to capture the voters' intentions, including voter confusion and errors in marking ballots, mishaps in handling ballots, machine failures, and tabulator errors.

As an example of how this information can be used to refine our understanding of the sources and causes of errors in voting consider what percent of the residual vote might be tabulation error. We can calculate this using data on the residual vote in presidential rates in New Hampshire from 1988 to 2000. In the counties and towns using only paper ballots, the residual vote rate was 2.3 percent; in the towns using optical scanning, the residual vote rate was 1.7 percent. Using the data on the validation rate from the historical record and from the 2002 elections, the tabulation invalidation rate was .83 percent for paper and .56 percent for optical scanning. Roughly one third of the residual vote, then, is pure tabulation error. The remainder is either unrecoverable ballots (i.e., people who accidentally voted twice) or blank ballots.

Stepping back from the details of the data, New Hampshire's recounts speak directly to two important questions in election administration.

Have we made progress? Do machine counts improve on hand counts? At least in the comparison of optical scanning and paper, the answer is yes. Historically, there is about a 1 percent difference between initial counts and recounts when ballots are tabulated by hand. The discrepancy between initial counts and recounts falls to about .5 percent with the optically scanned ballots. Looking at recounts from 2002, controlling for total vote, initial

vote, and office sought, optically scanned paper produced a lower discrepancy between the initial count and the recount compared to hand-counted paper.

Considering these tabulation errors, how confident should we be in vote counts, and when should we have a recount? The tabulation invalidation rate was low, especially for optical scanning. However, it was not trivial. In a US House election with 250,000 votes, the invalidation rate of .005 for scanners amounts to 1250 votes. The tabulation errors may swing toward any of the contestants in a recount. Assuming a uniform distribution of tabulation errors, any race decided by less than .5 percent of the vote will have a non-trivial probability of being reversed in a recount.

Most states have no set standard for an automatic recount; the courts or state election officer decides whether a recount is appropriate. Some states do have provisions for automatic recounts – typically, if the election is closer than one-half of one percent of ballots cast (National Commission on Federal Election Reform, 2002, pages 343-346). This standard is approximately equal to the average tabulation error rate with optical scanning equipment, but smaller than the rate for hand counted paper. The average, of course, reflects the typical case, so, for both optical scanning and hand counted paper, the discrepancy between initial counts and recounts in many jurisdictions in our data exceeded the .5 percent standard. State election officials and courts should treat an automatic recount standard of .5 percent as a minimum threshold, rather than an absolute standard.