A Voter Validation Experiment
SCREENING FOR LIKELY VOTERS IN PRE-ELECTION SURVEYS

Michael Dimock, Pew Research Center for the People and the Press
Scott Keeter, George Mason University
Mark Schulman, Schulman, Ronca and Bucuvalas, Inc.
Carolyn Miller, Princeton Survey Research Associates

May 16, 2001

Paper prepared for presentation at the 56th Annual American Association for Public Opinion Research Conference. May 17-20, 2001 Montreal, Quebec, Canada.

SCREENING FOR LIKELY VOTERS IN PRE-ELECTION SURVEYS

Dimock, Keeter, Schulman and Miller, 2001

INTRODUCTION

For all survey organizations, the accuracy of election projections hinges not only on having large and representative samples, but on accurately predicting who is and is not going to vote on election day. Discriminating between those who *say* they are going to vote and those who actually *are* going to vote has become a fine art, with numerous techniques. Though the process of culling likely voters from the larger pool of registered voters is often taken for granted, the process came to the forefront of many researchers' attention during the 2000 Presidential election, in which no candidate developed a clear lead and thus measurement precision became essential.

The Pew Research Center uses a procedure to arrive at likely voter estimates that was first developed in the 1950s and 1960s by Paul Perry, then chief election statistician at the Gallup Organization. The method is based on deriving a likely voter index from a number of questions that are known to relate to actual voter turnout. The purpose of this paper is to investigate the effectiveness of this approach using a dataset collected in the 1999 Philadelphia mayoral race in which the actual turnout of pre-election poll respondents was validated through precinct records. Using this data, we are able to test the effectiveness of the current likely voter index, whether expanding the index to include more items would improve predictions and whether trimming down the index would cause problems. We are also able to compare the effectiveness of the Guttman scaling technique applied by Gallup and Pew to methods in which respondents are assigned a probability of voting and weighted appropriately.

The results suggest that the standard Perry-Gallup likely voter index is as effective today as it has ever been, and is very difficult to improve upon. Expanding the 8-item index to include as many as 15-items has minimal impact on index efficacy. Moreover, more complicated probability models do nothing to improve the accuracy of the likely voter estimates that can be derived. Overall, our findings reinforce past research on predicting voter turnout from pre-election polls. Though it is impossible to accurately predict the behavior of all survey respondents, it is possible to accurately estimate the preferences of voters by identifying those most likely to vote.

In addition to studying pre-election likely voter screens, validation of non-responding households was also conducted. Based on this data, we are able to study whether turnout rates among non-respondents differs from that of survey participants.

BACKGROUND

The 1999 Philadelphia mayoral election turned out to be one of the closest in the city's history – just 9,447 votes separated the victor, Democrat John Street, from his opponent, Republican Sam Katz. This represents a 2.2% margin of victory among the 441,981 votes cast by residents of the city. Moreover, it was the most expensive municipal election in American history – with total spending well over $25 million, including $10 million by Street and $7 million by Katz (Committee of Seventy, 1999). According to the Philadelphia Board of Elections records, roughly 45% of registered voters turnout out on election day.

As far as can be determined, turnout among black constituents was relatively high, and overwhelmingly supportive of John Street, who is African-American himself. Roughly 42% of residents in overwhelmingly black wards voted, and Street received 91% of their vote. In overwhelmingly white wards, turnout was only slightly higher at 47%, with 83% of the vote going to Katz. This turnout disparity between black and white wards (42% to 47%) was the smallest in 16 years, according to a local public interest group.

The accuracy of these turnout figures is questionable, not because we don't know how many Philadelphia residents voted, but because of poor record keeping with respect to voter registration. In 1999, the Board of Elections identified 985,912 registered voters, or 93% of the 1,056,764 who were age eligible according to the Bureau of the Census. However, in the Pew Research Center's validation study, only 70% of respondents over the age of 18 claimed to be registered voters. And the evidence suggests that even this may be an overstatement. Just 86% of self-reported RVs who gave a name and address could actually be found in the voter registration list. Combined, this suggests that the true registration figure for Philadelphia may be around 60% of the voting age population, or roughly 600,000 instead of the nearly one million reported by the Board of Elections. Adjusting the total registration numbers based on this estimate, the 441,981 voters who participated in the November election represent roughly a 70% turnout rate among registered voters.

Though using a municipal election as a basis for a validation study has inherent external validity concerns, it provides a relatively low-cost means of accumulating and validating actual voting behavior.[1] Overall, we were able to match roughly 70% of the self-identified registered voters that we interviewed. The biggest factor in matching success was the willingness of the

[1] In the past, the Gallup organization has conducted numerous nationwide validation studies at greater expense, due to the costs of contacting multiple voter-record sources. The National Election Studies also conducted many validation studies in the 1980s, though they focused on post-election surveys.

interviewee to disclose their name and address at the end of the survey. We successfully matched 86% of those who gave us their name and address, just 43% of those who gave a name only.

The objective of the matching process is to uncover, using voting records, the actual behavior of our respondents on election day. Our matching process used five distinct identifying characteristics as a means of aligning interview subjects with voting records: phone number, address, last name, first name and birth year. Overall, 75% of the cases we were able to match met virtually all of these criteria – matching first and last name, birth date, and either phone or address or both. The remaining 25% were matched based on first and last name and birth year only (primarily among those who gave only their name), or those for whom we could match at least phone or address, first or last name, and at least a close match on birth year.

PART 1: The Elements of the Perry-Gallup Likely Voter Index

Typically, estimates of voter preferences in an election poll are based only on those who are registered to vote. An analysis of a 1984 Gallup validation study suggests that filtering out respondents who say they are not registered introduces very little error in horserace predictions. Just 6% of those who said they weren't registered actually were, according to voting records, and only 2% actually voted (Colasanto and Mattlin, 1987). As a result, all the analysis to follow will be based solely on respondents who report themselves as registered voters.[2]

But basing horserace predictions on all who claim to be registered is still problematic, since survey participants tend to both overstate their registration and their propensity to vote. In the 1984 Gallup study, fully 23% of those who claimed to be registered were not, and 30% did not vote on election day. Were this error distributed randomly across the population, we might overlook it. However, overestimation of registration and voting is highest among predominantly Democratic constituencies, leading to a systematic bias in favor of Democratic candidates unless some further filter is applied.

The likely voter screen used by Gallup and the Pew Research center is based on an index measuring each respondent's propensity to vote. In addition to registration, the likely voter index, originally developed by Paul Perry at Gallup, is made up of eight items intended to

[2] The Pew Research Center uses an even stricter standard when measuring voter registration. After being told that many people are so busy they can't find time to register or move around so often they don't get a chance to re-register, a respondent must affirm that they are registered *and* reaffirm that there is no chance that their registration has lapsed because they moved or for some other reason. Though the data is not available to test the effectiveness of this screen, it is doubtful that it mistakenly excludes any *more* voters than the less stringent Gallup screen.

TABLE 1: Elements of the Likely Voter Index

Points on Index	Question		Response Categories
1	Q2	Thought given to election	A Lot/Some
1	Q6	Follow government affairs	Most/Some of time
1	Q14	Plan to Vote	Yes
1	Q15	Likelihood of voting (10-pt scale)	7,8,9,10
1	D13	Voted in previous Presidential elect	Yes, recall candidate
1	Q7	How often do you vote	Always/Nearly/Part of time
1	Q4	Know where to vote	Yes
1	Q5	Ever voted in current election dist.	Yes
8			

Respondents are automatically coded zero (0) on the index if:
(1) they are not registered to vote
(2) they say they do not plan to vote
Respondents under 22 are not penalized for past voting behavior (Q5, Q7, D13)

identify four concepts related to voter turnout: voter interest, voter intentions, past voting behavior, and knowledge about where to vote, each of which will be discussed below. Though there are slight variations between the original Perry-Gallup index applied in the 1960s, 1970s and early 1980s and the one used today by the Pew Research Center, they are based on the same fundamental structure, outlined in Table 1 below.

This procedure results in a Guttman index with values ranging from zero to eight, with the highest values representing those with the greatest likelihood of voting. Both Gallup and the Pew Research Center then make a projection of voter turnout based on the past turnout rates and early indicators of turnout, such as particularly high or low levels of interest in the campaign. This turnout projection is used to define what percentage of respondents will be considered "likely voters" – the proportion of highest scoring respondents on which election estimates will be based. For example, in forecasting the 2000 presidential election, the Pew Center forecast that 50% of the age-eligible population would vote, and based its estimates on the 50% of respondents receiving the highest index scores. In the 1999 Philadelphia study, evidence suggested that roughly 70% of registered voters would turn out to vote.[3]

In addition to providing a more stable and reliable measure across distinct survey samples, the eight-item index provides a level of operational and content validity that no single item can achieve. But in order to fully investigate the effectiveness of the index and whether

[3] To reach a precise percentage of "likely voters" from the eight-point index, we often take all respondents from the one or two highest scoring categories, and a weighted proportion of the next category. For example, if we estimated a turnout of 50%, and 40% of the sample scored an "8" on the index and 15% scored a "7", we would count all of the "8"s and weight the category "7" by .6667 (10/15) to estimate the likely voter pool.

improvements can be made, the relevance and effectiveness of each index element will first be examined, grouped by the substantive concepts they measure.

Measures of Voter Interest

Citizens who are more interested in politics and who have been paying attention to the campaign are presumably more likely to vote than those who are disinterested, and a bivariate analysis of voting patterns suggests that this is true (see Table 2). To measure interest in politics, respondents are asked how much they follow what's going on in government and public affairs. According to the 1999 Philadelphia validation study, fully 84% of those who follow politics "most of the time" actually voted in the mayoral race, compared to 61% of those who follow politics "only now and then" and 55% of those who "hardly at all" follow government affairs.

Table 2: Measures of Interest and Validated Voter Turnout

Follow gov't affairs	%		**Voted**
Most of the time	52	⇒	**84%**
Some of the time	32	⇒	**71%**
Only now and then	11	⇒	**61%**
Hardly at all	5	⇒	**55%**
DK/Refused	*	⇒	--
	100		
Thought given	%		**Voted**
A lot	58	⇒	**85%**
Some (Vol.)	8	⇒	**74%**
Only a little	30	⇒	**62%**
None (Vol.)	3	⇒	--
DK/Refused	1	⇒	--
	100		

In the 1984 Gallup pre-election poll a slightly different question achieved similar results. Seventy-nine percent of those who say they have a "great deal" of interest in politics turned out on election day 1984, compared to 71% of those who have a "fair amount" of interest, 60% of those with "only a little" interest, and 19% of those with "no interest at all."

Looking at actual attention to the campaign in the bottom of Table 2, we see that 85% of 1999 respondents who said they had given "quite a lot" of thought to the upcoming election actually voted, compared to 62% of those who said "only a little." The identical question achieved comparable figures in 1984, with 74% of those giving "quite a lot" of thought to the election actually voting, compared to just 57% of those who said "only a little."[4]

Measures of Voter Intentions

On its face, the most direct way of predicting voter turnout is to simply ask whether a person intends to vote or not. Unfortunately, such a straightforward question often gives us little traction, since nearly all who say they are registered to vote tell us that they plan to vote. Fully

[4] Overall, 76% of RVs (Waves 1 & 2 combined) voted in the 1999 Philadelphia survey, and 67% voted in the 1984 Gallup pre-election survey. The response categories enclosed in the boxes in Tables 2 through 5 are those coded "1" in the Perry-Gallup index.

97% of registered voters in the Philadelphia study told us they planned to vote, with only 2% saying they did not, proportions almost identical in the 1984 nationwide Gallup study. Though all who say they do not plan to vote are automatically coded at zero on the Perry-Gallup index, this question has a minimal effect on overall index accuracy.

Table 3: Measures of Intention and Validated Voter Turnout

Plan to vote	%		**Voted**
Yes	97	⇒	**77%**
No	2	⇒	--
DK/Refused	1	⇒	--
	100		

10-pt scale	%		**Voted**
10	77	⇒	**84%**
9	6	⇒	**71%**
8	6	⇒	**46%**
7	3	⇒	**33%**
1-6	7	⇒	**39%**
DK/Refused	1	⇒	--
	100		

A more promising measure of voter intention has respondents rate their chances of voting on a scale of 10 to 1. Though more than three-fourths of registered voters in both 1999 and 1984 rated their chance of voting as a 10, this index provides a bit more variance than the simple "do you plan to vote" question. Unfortunately, the Perry-Gallup index codes all responses above "6" as likely voters. This has two problems – first, over 90% (92% in 1999, 95% in 1984) rate their chance of voting as 7 or higher on the scale, leaving us with little variance. Second, in 1999 only 46% of those who rate their chances of voting at "8" and only 33% of those who rate their chances at "7" actually voted, introducing a high level of error into the likely voter index. In light of this, we will test whether moving the cutpoint up to "9", or even a solid "10" would improve index effectiveness.[5]

Measures of Past Voting Behavior

Those who have voted in the past are the most likely to turn out in any given election, and measures of past voting behavior are central to any measure of the likelihood of voting. The Perry-Gallup index uses two general measures of past voting: whether an individual voted in the previous presidential election, and the individual's own assessment of how regularly they vote. Each proves to be a powerful predictor of turnout in both the 1999 mayoral race and the 1984 general election. Since respondents aged 18-21 may not have had the opportunity to vote in previous national elections, past voting behavior is not included as part of the likely voter index for these respondents.

Table 4 shows that those who say they voted in the 1996 Presidential election were roughly twice as likely as those who did not to participate in the 1999 Philadelphia mayoral

[5] The 1984 Gallup pre-election poll exhibited the same problem. Among those rating their chance of voting as a "7", just 30% actually voted, and just 65% of those rating their chance of voting as a "7" did likewise. By comparison, 73% of those coding their chances at "9" or "10" voted.

election. Interestingly, the 8% of the sample who couldn't recall if they had voted, or refused to say, also exhibited high turnout in the mayoral race.

Our baseline likely voter index codes respondents as likely voters only if they say they voted in 1996 *and* can recall the name of the person they voted for. The assumption underlying this coding choice is that we know that many people over-report past voting (in this poll fully 80% of registered voters told us they voted in 1996), and those who say they voted but can not recall who they voted for are the most likely to be the non-voters in the crowd. The validation study suggests otherwise. Turnout among the 10% who say they voted in 1996 but can't recall who they voted for is not statistically different from turnout among those who say they voted and can recall for whom. Below we will test whether altering the index to include these respondents in the likely voter index might improve index accuracy.

Table 4: Measures of Past Voting and Validated Voter Turnout

Voted in '96 Presid.	%		**Voted**
Yes, Voted	70	⇒	**81%**
Voted, forgot who	10	⇒	**76%**
Did not vote	12	⇒	**40%**
DK/Refused	8	⇒	**82%**
	100		

How often do you…	%		**Voted**
Always	60	⇒	**85%**
Nearly always	25	⇒	**74%**
Part of the time	9	⇒	**43%**
Seldom/(Never-vol.)	5	⇒	**21%**
Other/DK/Refused	1	⇒	**--**
	100		

Fully 85% of those who say they always vote turned out on November 2, 1999, along with 74% of those who say they nearly always vote. By comparison, just 43% of those who say they vote part of the time went to the polls, and just 21% of those who said they seldom or never vote. Unfortunately, the Perry-Gallup likely voter index codes those who say they vote just part of the time as likely voters, which this bivariate analysis suggests may be inaccurate.[6] Altering the cutpoint on this question to include only those who always or nearly always vote will be tested.

Measures of Knowledge about Where to Vote

The Gallup index also includes two measures related to the practicalities of voting – the sorts of things that might keep a person who intends to vote from making it to the polls on election day. First, respondents are asked if they know where people in their neighborhood go to vote. Second, respondents are asked if they have ever voted in their precinct or election district where they now live. Though this latter question is similar to measures of past voting behavior,

[6] The 1984 Gallup pre-election poll reinforces this concern, as just 29% of respondents who said they vote "part of the time" actually turned out on election day, compared to 71% of those who said they "nearly always" vote, and 78% of those who said they "always" vote.

it also encompasses the more practical question of whether a person knows how to make it to the polling booth when election day comes.

Though the vast majority of registered-voter respondents answer both of these questions in the affirmative, they do serve to discriminate between likely and unlikely voters fairly effectively.

Table 5: Voting Knowledge and Validated Voter Turnout

Know where to vote	%		**Voted**
Yes	90	⇒	**79%**
No	9	⇒	**53%**
DK/Refused	1	⇒	**--**
	100		

Ever voted in prec't	%		**Voted**
Yes	86	⇒	**81%**
No	14	⇒	**43%**
DK/Refused	1	⇒	**--**
	100		

Combined	%		**Voted**
Yes to both	81	⇒	**82%**
Yes to one	13	⇒	**55%**
No/DK to both	6	⇒	**40%**
	100		

These last questions provide an opportunity to display the power of constructing an index to measure voter turnout. While each question alone separates voters and non-voters fairly effectively, the proportion answering each in the affirmative makes them unwieldy – we would prefer our measure of likely voters to include fewer than 90% of the registered-voter sample. By combining the two questions, we find that about 80% answered both questions in the affirmative, and fully 82% of those who did actually voted according to the validation study. Of the other nearly 20% who answered one or both in the negative, just 50% voted. The combination of the two questions, in other words, provides us with a more useful proportional division of the population, and a more accurate screen of likelihood of voting. To further improve both of these qualities, a full index of all eight items will be constructed.

Part 2: Index Accuracy

Before analyzing the effectiveness of the likely voter index, we will first explore the accuracy of each individual indicator as a measure of voting behavior. In other words, if respondents' answers to any single question were used to predict whether each individual would vote or not, what proportion of the sample would be classified correctly as voters and non-voters, and would we be able to accurately predict the preferences of true voters. Our initial analysis will focus on Wave 2 of the experiment, collected the weekend immediately prior to election day.

For each question included in the likely voter index, Table 6 shows the proportion of respondents who would be coded as likely voters according to that question alone, along with the net proportion of respondents who would be correctly classified as voters and non-voters, according to the post-election validation. The final column shows the accuracy of the horserace prediction each question would produce if it were used as a likely voter filter.

In the top row of the table, we see that 77% of wave 2 registered voters actually voted in the 1999 mayoral election, (if we included all RVs as likely voters, we would be correct 77% of the time). We also see that we would overestimate the Democratic candidate's lead by nearly 3% if we were to base our estimate on all RVs.

We can attempt to eliminate this bias by further identifying each registered voter's propensity to vote using the screening questions that we have already seen to be correlated with voter turnout. Each individual indicator would identify some proportion of RVs as likely voters based on their answers, with horserace predictions based on this subset of the population.

Table 6: Individual Items as Likely Voter Predictions
Based on 1999 Wave 2 respondents (N=856)

	Answer	Percent Giving Answer	Percent Correctly Classified	Percent Democr. Overest.
Total	Registered voter	100%	77%	+2.7%
Perry-Gallup LV Index Items				
Q14: Plan to vote	Yes	97	78	+2.5
Q7: How often vote	At least part time	95	79	+1.7
Q15: Scale of vote intention	7,8,9,10	92	77	+2.1
Q4: Know where to vote	Yes	89	77	+1.4
Q5: Ever voted in precinct	Yes	87	78	-0.2
Q6: Follow gov't affairs	Most/Some	85	74	+2.2
D13: Voted in '96 Pres.	Yes, recall name	79	71	+2.7
Q2: Thought given election	A lot/Some	73	73	+0.4
Alternate Cutpoints				
Q7: How often vote	Always/Nearly	88	80	+2.5
Q15: Scale of vote intention	9,10	84	78	+2.4
D13: Voted in '96 Pres.	Yes, all	84	75	+2.7
Perry-Gallup 8-Item Index	Highest 70%	70	73	+0.2
Alternate 8-Item Index	Highest 70%	70	75	+1.2

There are three ways to measure the accuracy of a likely voter indicator. First, we can focus on the percent of the sample who are correctly identified as voters and non-voters, shown here as the percent correctly classified. Second, we can compare the distribution of candidate preferences among those coded as "likely voters" to those of respondents in the sample who actually voted on election day. Third, we can compare the demographic characteristics of the likely voter pool to the demographics of actual voters. As Colasanto and Mattlin succinctly stated in their 1987 Joint Statistical Meeting paper, "in order to work successfully, [a] scale need not necessarily be a foolproof method of selecting individual voters as long as the socio-economic and demographic composition of the voting electorate is accurately reflected in the socio-economic and democraphic composition of predicted likely voters." (1987, 8). Our analysis suggests that Colasanto and Mattlin are correct in their assessment, and that the percent correctly classified is in fact a rather poor means of assessing the accuracy of a likely voter measure.

With respect to correctly classifying voters and non-voters, the column labeled "percent correctly classified" in Table 6 shows that each individual indicator does a fairly good job of predicting voter behavior prior to the election, though few exceed the "null" model of counting all RVs as likely voters. In fact, the table highlights one of the fundamental problems of using the percent correctly classified as a measure of the accuracy of a likely voter screen – it is heavily influenced by the overall percent who are classified as likely. As we will discuss later, increasing the proportion identified as likely voters almost invariably increases the percent correctly classified.

With respect to accurately predicting the preferences of voters, the column labeled "percent Democratic overestimate" in Table 6 shows that some items clearly outperform others, though within a margin of error that makes generalizability questionable. Some questions provide horserace predictions that are just as flawed as what an estimate based on all registered voters would produce -- overestimating the Democratic candidate's margin by nearly 3 percentage points. Others, namely whether the respondent has ever voted in their election precinct or district and how much thought the respondent has given to the election, produce horserace estimates that nearly perfectly capture the preferences of those who actually voted on November 2.

Table 6 also provides an initial test of alternate cutpoints on three index items. In all three cases, the bivariate analysis in Tables 1 through 4 suggested that our original cutpoints may have mistakenly included many non-voters as likely voters and visa versa. The summary analysis here further suggests that adjusting the cutpoints on these three items might improve index accuracy. In all three cases, adjusting the cutpoint improves the percent correctly classified over the original cutpoint, though doing so tends to overstate Democratic candidate support.

Combining individual items into an index to measure the likelihood of voting has many advantages. Most directly, an index based on multiple items provides greater validity in that it is based on a range of items that are all known to be related to turnout, rather than one single item. Perhaps more importantly, an index has greater reliability – the error present in each individual question will be minimized when all are combined into a single measure. Finally, an index allows the researcher to determine what proportion of respondents on which to base the horserace prediction, rather than being constrained by the distribution of a single indicator. Most questions, as we have seen, have natural cutpoints, which, since respondents generally overstate their propensity to vote, typically include too large a proportion of the registered voter base to be useful as a likely voter screen individually. Combining all eight questions into an index, ranging from zero to eight, allows the researcher to use ex-ante information to determine what proportion should be considered likely to vote, and to cut the index at precisely this point.

The bottom of Table 5 shows the results of both the original likely voter index and the alternate index based on the adjusted cutpoints on Q7, Q15 and D13. We estimated that 70% of registered voters would turn out to vote on election day – a slight underestimate for this particular sample in which 77% of self-reported registered voters in this sample actually voted. This likely voter definition resulted in all respondents with scale scores of eight considered as likely voters and a portion of those with scores of seven (84% for the original index, 96% for the alternate). Respondents with scores of seven were weighted down to reflect their appropriate share of the predicted likely voter pool.

Interestingly, neither the original nor the alternate index was able to achieve a higher percent correctly classified than the null of counting all RVs as likely voters, or even to achieve the percent correctly classified by some of the individual elements of the index on their own. As we will show below, however, this seeming failure on the part of the 8-item indices can be attributed primarily to the selection of a cutpoint of 70%, which is significantly lower than the proportion coded as likely voters by most individual index elements.

The advantage of the index over the individual items can be seen in the horserace predictions, which tend to be more accurate in estimating the actual candidate preferences of those who voted within the sample. Even though the original index only classified 73% of RVs correctly, it nearly perfectly predicted the actual preferences of those who voted. This result mirrors the findings of the 1984 Gallup validation study conducted by Colasanto and Mattlin, in which the Perry-Gallup likely voter index correctly classified just 69% of RVs, yet estimated the candidate preferences of voters almost exactly.

The test of the alternate index provided mixed results. Though the percent correctly classified by the alternate index was better than the original, the horserace prediction overestimated the Democratic candidate's margin by just over one percentage point. Neither difference is statistically significant.

The advantage of the 8-point likely voter index over individual items can be seen most clearly in Figure 1, in which we can visually compare the accuracy of all measures in terms of both percent correctly classified and horserace prediction, controlling for the proportion of the population identified as likely voters. The X-axis in Figure 1 shows the percent of RVs identified as likely voters by each question. Each "X" in the figure shows the percent correctly classified by each question if used as an independent likely voter measure. Since the researcher has the discretion to achieve any proportion of likely voters we want from the 8-point likely voter index, the percent correctly classified by the index is represented here by a bold line, ranging from a cutpoint in which 48% of RVs are coded as likely to the full 100% of RVs coded as likely. Note that this line predicts the voting of 77% of RVs correctly when 100% of RVs are coded as likely voters, reflecting the fact that 77% of registered voters actually voted.

What this comparison shows us is that achieving a higher percent correctly classified is more a function of the proportion of the population that is coded as likely voters than it is a measure of index accuracy. If we were to inflate the cutpoint of our 8-point index to include 92% of RVs as likely voters, we could achieve a percent correctly classified of over 80%. Similarly, as we saw in Table 6 above, those individual items which classify higher percentages of RVs as likely voters tend to classify a higher proportion of the population correctly.

Correctly classifying respondents does not lead to better horserace predictions, however. The disconnect between percent correctly classified and index accuracy is most clearly seen by comparing the arc of the bold like with the horserace predictions below. Figure 1 clearly shows, consistently with our expectations, that the higher the proportion of the population identified as likely voters, all the way up to 100%, the better the Democratic candidate appears to do. But among the 659 (77%) registered voters who actually voted, support was evenly split at 42% for each candidate. Any deviation from this even split represents an erroneous horserace prediction.

Though we can increase the percent correctly classified by raising the proportion identified as likely voters, doing so decreases the accuracy of our horserace predictions by overstating Democratic support. The best horserace prediction comes when 70% of registered voters are coded as likely (serendipitously, our ex-ante cutpoint), even though this has an inferior percent correctly classified relative to higher cutpoints.

One of the most striking results of this analysis is the inability of an index measure of likelihood of voting to significantly outperform individual items in predicting voting behavior. Figure 1 clearly shows that for any given cutpoint, the 8-item likely voter index only slightly outperforms individual index elements in terms of the percent correctly classified as voters and non-voters, and does no better than individual items in terms of horserace predictions.

This analysis suggests that the main advantage of the index is not so much improved accuracy, but reliability across surveys and elections, and the fact that the index allows the researcher to select the proportion of respondents who will be classified as likely voters, rather than being constrained by response rates to any one or two questions. While the calculation of a turnout estimate can never be considered a precise science, analysis of the 1999 Validation study suggests that one needn't predict turnout rates precisely to be successful. Any turnout estimate between 69% of RVs and 90% produced the same horserace prediction when rounded to whole numbers, -- a virtually perfect estimate of the preferences of actual voters in the survey.

Demographic Analysis

A third way to assess the accuracy of the likely voter index is to see how well it corrects for known variations in turnout across different demographic groups. One of the key faults of any horserace prediction based on all RVs is that it overstates the preferences of younger, less educated, and often minority voters who tend to participate at lower rates than older, more educated whites. Since these demographics are highly correlated with partisanship, estimates based on all RVs are inherently biased toward the Democratic candidate.

Table 7 compares the actual turnout among key demographic groups with the percent of each group that is predicted to be a voter by the likely voter index. Even though the index may not classify all respondents correctly, as long as it creates a likely voter base that has similar demographic characteristics to those in the sample who actually do vote, it will provide more accurate electoral estimates.

As noted above, 77% of the wave 2 RV sample turned out in the 1999 Philadelphia mayoral election, whereas our ex-ante prediction was 70% turnout. Therefore, we underestimated turnout by 7% overall. For the most part, our underestimation was fairly evenly distributed across all demographic groups. For example, we can see that turnout was lower among less educated voters than among the college educated, a pattern that the likely voter index does a fairly good job of capturing.

There are two noticeable differences between the distribution of actual voters and the distribution of those predicted as likely voters, however. First, though the likely voter index accurately predicts that turnout will be lower among younger voters, it actually seems to overcompensate for this age effect in the Philadelphia race, predicting that only 40% of RVs under 30 would vote, when in fact 57% voted. As a result, the preferences of this demographic are clearly underrepresented in the likely voter estimate.

The index also seriously underestimates Republican turnout rates, which were just as high as turnout among Democrats. Of course, this is one of the areas where the generalizability of this validation study is weakest. Just 16% of RVs consider themselves to be Republicans, compared to two-thirds who identify themselves as Democrats. As a result, even though the index underestimates the preferences of Republican voters, it has little aggregate effect on the horserace estimates due to the small number of cases involved.[7]

Table 7:
Demographic Analysis of Perry-Gallup Index

	Percent Actually Voting	Percent Predicted Likely	*Percent Underest.*
Total	77	70	*-7*
Male	74	69	*-5*
Female	80	71	*-9*
White	80	71	*-9*
Black	74	69	*-5*
18-29	57	40	*-17*
30-39	69	66	*-3*
40-49	72	73	*+1*
50-64	89	78	*-11*
65+	85	78	*-7*
H.S. incomplete	68	60	*-8*
H.S. graduate	73	66	*-7*
Some college	79	73	*-6*
College graduate	88	81	*-9*
Republican	79	67	*-12*
Democrat	78	72	*-6*
Independent	69	66	*-3*

Part 3: Wave 1 and Wave 2 Analysis

Perhaps the most important value of the index is its consistency over time. The above analysis is based on Wave 2 respondents in the 1999 validation study, conducted 3 to 6 days before election day. But pollsters typically want to make likely voter estimates far earlier in the election cycle if possible.

Table 8 shows that the original likely voter index serves nearly as well for this purpose two to three weeks before election day as it does on the eve of the election. The index correctly

[7] Freedman and Goldstein (1996) test a model in which demographic data such as age, race and education are used to weight the subset of registered voters who say they intend to vote to achieve a demographic profile similar to that of the validated voting population. By using demographics, instead of an index of indicators as applied here, they are able to achieve a high level of predictive accuracy.

classifies 72% of respondents (compared to 73% of Wave 2 RVs), and though it doesn't come up with quite as accurate an estimate of the preferences of actual voters during Wave 1 as it did during Wave 2, it provides a significant improvement over an estimate based on all RVs, and outperforms predictions based on any single item as a likely voter screen. In Wave 1, the likely voter index would underestimate the Republican candidate's lead at that time by only 1.7%, compared to the 5.5% error in an estimate based on all RVs.

The above tests clearly show the key advantages of the likely voter index over basing estimates on all RVs or on any single likely voter indicator. In addition, we tested the alternative index based on adjustments to the cutpoints within three key indicators, to little improvement in overall index accuracy.

This experiment suggests that the effectiveness of the likely voter index two weeks prior to election day is comparable to election weekend. Of course, most survey researchers want to estimate voter preferences months prior to election day. Our experiment provides no clear test of index efficacy in this type of time frame.

Table 8: Individual Items as Likely Voter Predictions
Based on Wave 1 Form 1 respondents (N=408)

	Answer	Percent Giving Answer	Percent Correctly Classified	Percent Democr. Overest.
Total	Registered voter	100%	72%	+5.5
Perry-Gallup LV Index Items				
Q14: Plan to vote	Yes	96	74	+6.1
Q7: How often vote	At least part time	92	76	+5.6
Q15: Scale of vote intention	7,8,9,10	91	75	+5.1
Q4: Know where to vote	Yes	91	71	+4.0
Q5: Ever voted in precinct	Yes	84	75	+3.6
Q6: Follow gov't affairs	Most/Some	81	71	+6.0
D13: Voted in '96 Pres.	Yes, recall name	66	66	+3.0
Q2: Thought given election	A lot/Some	56	63	-3.1
Perry-Gallup 8-Item Index	Highest 70%	70	72	+1.7

Part 4: Testing New Index Items

But the question remains as to whether we might improve on index accuracy by expanding the scope of the likely voter index to include other indicators associated with voter turnout. The 1999 validation study included eight additional items that might be used to identify likely voters. These are:

Q3: Campaign News Interest: How closely the respondent has been following news about candidates and the election campaign.

Q17: Voting Difficulty: How difficult the respondent says it is for him/her to get out and vote in the mayoral election.

Q22: Learned Enough: Whether the respondent feels he/she has learned enough to make an informed choice.

Q23: Contacted by Party: Was respondent contacted by party or candidate's campaign.

Q26: Recently Moved: Whether the respondent has moved in the last two years.

D15: 1998 Congressional Vote: Whether respondent voted in the 1998 Congressional election.

Q12: Strength of Support: How strong the respondents expressed candidate preference is.

D28: Interviewer Assessment: Interviewer's impression of the respondent's interest in the upcoming mayoral election.

Table 9: Measures of Interest and Voting Behavior

Campaign News	%		**Voted**
Very closely	38	⇒	**84%**
Fairly closely	42	⇒	**77%**
Not too closely	14	⇒	**58%**
Not at all closely	6	⇒	**55%**
DK/Refused	*	⇒	--
	100		

Difficult to vote	%		**Voted**
Very easy	61	⇒	**83%**
Easy	30	⇒	**70%**
Difficult	7	⇒	**54%**
Very difficult	2		
Neither/DK/Ref	*	⇒	--
	100		

Learned enough	%		**Voted**
Yes, enough	67	⇒	**79%**
No, not enough	30	⇒	**68%**
DK/Refused	3	⇒	--
	100		

Contacted by Party	%		**Voted**
Yes, contacted	24	⇒	**85%**
No	75	⇒	**73%**
DK/Refused	1	⇒	--
	100		

Moved in 2-years	%		**Voted**
No, didn't move	85	⇒	**79%**
Yes, moved	15	⇒	**58%**
DK/Refused	*	⇒	--
	100		

Voted in 1998 Cong	%		**Voted**
Yes, certain voted	64	⇒	**86%**
No	29	⇒	**54%**
DK/Refused	7	⇒	**80%**
	100		

Strength of support	%		**Voted**
Strongly	40	⇒	**81%**
Moderately	31	⇒	**71%**
Lean	12	⇒	**72%**
No Preference	16	⇒	**77%**
DK strength	1	⇒	--
	100		

Interviewer assessment			**Voted**
Very interested	60	⇒	**84%**
Somewhat interest	33	⇒	**67%**
Not too interested	6	⇒	**52%**
Not at all interest	1	⇒	--
	100		

With the exception of strength of support, each of these items is correlated with voter turnout in the expected direction. Interestingly, respondents who say they "strongly" support their favored candidate are not significantly more likely to turn out on election day than are those who express no preference – even within a week of election day. As a result, this item is dropped from further analysis.

Table 10 shows that these new items show similar properties to items already included in the Perry-Gallup index. Most do a fairly good job of separating voters from non-voters, though items such as Q.23, whether the respondent was contacted by a party or candidate, fail as individual indicators because too small a proportion of registered voters are classified as voters.

Table 10: Possible New Items as Likely Voter Predictions
Based on Wave 2 respondents (N=856)

	Answer	Percent Giving Answer	Percent Correctly Classified	Percent Democr. Overest.
Total	Registered voter	100%	77%	+2.7%
Possible New Index Items				
Q3: Campaign news interest	Very/Fairly close	83	73	+0.9
Q17: Voting difficulty	Easy	88	75	+3.6
Q22: Learned enough	Yes	71	66	-0.9
Q23: Contacted by party	Yes	31	46	-1.6
Q26: Didn't move in 2 yrs	Yes	87	74	+1.3
D15: 1998 Cong. Voter	Yes, certain	66	70	+1.3
D28: Interviewer assessment	Very interested	65	68	-0.1
	Reliability			
Original 8-Item LV Index	Alpha = .61	70	73	+0.2
New 7-Item Index	Alpha = .56	70	73	-0.1
New 15-Item Index	Alpha = .74	70	76	-1.8

Interestingly, a likely voter index based solely on the 7 new items performs identically to the 8-item Perry-Gallup index. With a cutoff that classifies 70% of registered voters as likely voters, both indices correctly classify 73% of RV respondents, and estimate the candidate preferences of actual voters nearly perfectly. This analysis, combined with the above analysis of altering item-cutpoints within the Perry-Gallup index, suggests that virtually any combination of these measures can produce a reasonably accurate likely voter index. Perhaps more importantly, likely voter estimates derived from different surveys using different indexes can be considered relatively comparable.

A scale adding the 7 new items to the 8 original to create a 15 item index has greater internal reliability (as would be expected from the increase in number of correlated items) and produces a higher percent correctly classified. However, the horserace prediction overcompensates for the Republican turnout advantage in this case, and produces an inferior horserace prediction.

This result suggests that a researcher could create a likely voter index from almost any group of items and produce similar levels of accuracy. Increasing the number of items leads to predictable improvements in scale reliability, but no real improvement in overall index accuracy. The solution to the variability in likely voter predictions that we saw in the 2000 tracking polls is not a more comprehensive screen for likely voters.

Now the question is whether we can create an index that is just as accurate as the 8-item original index, but based on fewer indicators (thus saving minutes on survey questionnaires).

Part 5: Reducing the Size of the Index

Trimming down the likely voter index is clearly desirable, but the question is how to do so. As detailed above, the index is designed to cover four concepts related to an individual's propensity to vote: interest, intentions, past behavior and practicalities. In attempting to trim back the instrument, we will work from the assumption that we want to include at least one measure of each concept in the final index, to maintain our content validity. Thus, instead of going from an eight-item index with two indicators for each concept, we will test the effectiveness of moving to a four-item index, with one indicator for each concept.

Eliminating questions from an index is always difficult, however, because it affects the overall validity of the measure. Our examination of the accuracy of a shortened likely voter scale will be based on a four-item index based on Q2, Q4, Q7 and Q15 as a measure of voter interest. The substantive reasons for eliminating Q5, Q6, Q14 and D13 are outlined below.

Voter Intentions

An obvious candidate for removal might be Q14, which asks respondents whether they plan to vote or not. Consistently, over 95% of respondents answer this question in the affirmative, giving the question little impact on the overall index, and of the remaining 5%, most would be coded as unlikely voters even without this question's presence. Even though this question is given more weight in the Perry-Gallup index since any respondent who says "no" is automatically coded a zero on the scale, fully 852 of the 856 validated respondents in the wave 2 sample would be coded identically if Q.14 were removed from the index. Moreover, a likely voter index calculated without Q.14, exhibits the same accuracy in terms of both the percent correctly classified and the horserace estimates.

By comparison, the other measure of voter intentions used in the Perry-Gallup index – respondents' rating of their chances of voting on a scale from 1 to 10 – has a good deal more variance and versatility. Removing Q.15 from the index causes greater disturbance in terms of the percent of cases that change classifications.

A comparison of the importance of each of these questions to the Perry-Gallup index can be seen in Table 11, where the original 8-item index is compared to each possible 7-item index that would result from the removal of a single index element. We see here that a 7-item index which excludes the "Q14: Plan to vote" question is virtually identical to the original 8-item index: less than 1% of cases are coded differently, and the percent correctly classified and horserace predictions remain the same. By comparison, removing the Q15 "10-point scale" question has only a slightly larger impact (3% of cases change categorization, with 1% fewer correctly classified).

Table 11: Eliminating LV Index Items, One-by-One
Based on Wave 2 respondents (N=856)

		Reliability (alpha)	Percent of cases Changed	Percent Coded Likely	Percent Correctly Classified	Percent Democr. Overest.
Total (All RVs)		--	--	100%	77%	+2.7%
Original 8-Item LV Index		.64	--	70	73	+0.2
	7-Item index…					
Measures of Intention	without Q.14	.59	*%	70	73	+0.4
	without Q.15	.60	3%	70	72	+0.4
Past Behavior	without Q.07	.62	1%	70	72	+0.6
	without D.13	.59	12%	70	73	-0.6
Measures of Interest	without Q.02	.61	17%	70	71	+0.2
	without Q.06	.59	8%	70	73	-0.7
Practicalities	without Q.04	.62	4%	70	72	+0.0
	without Q.05	.63	5%	70	72	+0.5

One factor acting in favor of keeping the Q14 question is its brevity, and the fact that it acts as a natural setup for Q15, where respondents are asked to rate the chances that they will vote in the election on a scale from 1 to 10. Such practical factors, in addition to its obvious face validity, may merit the retention of the Q14 even though our evidence suggests its impact on the accuracy of a likely voter screen is marginal at best.

Past Voting Behavior

Another problematic question in the original likely voter index is D13, whether the respondent voted in the previous presidential election. Aside from the applicability of past national voting behavior to state or local voting behavior, there are some serious concerns associated with this as a filter for future voting behavior.

First, the accuracy of this type of specific recall question is apt to change across the four-year presidential election cycle. In other words, the measure may be more accurate one or two years after a presidential election than it is three to four years after. In the interest of a fully generalizable and stable likely voter index, such fluctuations are less than desirable.

Second, and more importantly, there are biases built into the presidential voting report based on the nature of each presidential election. Past research has shown that not only do respondents tend to exaggerate participation in elections, but surveys tend to overstate levels of support for the winning candidate. Much of this error in favor of the winning candidate comes from under-reporting of voting from supporters of the losing candidate, and over-reporting from supporters of the winning candidate who did not actually vote. In short, any measure of past voting behavior that asks respondents about particular behavior in a previous election tends to be biased in favor of the candidate who won that election.

This was certainly the case in surveys following the 1996 Clinton victory over Dole. The Pew Research Center's 2000 pre-election survey found that Clinton lead Dole by a 54%-30% margin among RVs who say they voted in 1996, a far greater margin of victory than the actual 1996 outcome.

Because of this over-reporting among those allied with the winning candidate, past presidential vote is the only likely voter index item in which respondents coded as likely to vote are actually more Democratic and more supportive of Democratic candidates than those who are considered unlikely to vote. This was the case in both the 1999 Philadelphia Validation study and the 2000 Pew Election-Weekend study, but NOT in the 1984 Gallup study, in which the bias was in the opposite direction, as would be expected following Ronald Reagan's victory on 1980.

Table 11 shows that fully 12% of respondents change their classification in terms of whether they are perceived as likely or unlikely voters when the index is calculated without D.13. And, because the item raised the index scores of Democrats more than Republicans, the new horserace prediction favors the Republican candidate more than the full 8-item index does. This 7-item scale classifies just as many respondents correctly as voters and non-voters.

Clearly an index constructed without D.13 is substantively different from the full 8-item index, given the 12% who change classifications with it's removal. However, given the problematic nature of the question, we believe it is the appropriate choice for removal. We will retain, in it's stead, the respondents self-reported frequency of voting as our measure of past voting behavior in the trimmed-down, four-item index.[8]

Another possible solution to this problem would be to use past congressional vote instead of past presidential vote as a measure of past turnout. Though the data are not shown here, we investigated this alternative, but it does little to improve overall index effectiveness.

Voter interest

The Perry-Gallup index includes two measures of voter interest – one a general measure of voter interest in politics (Q.6), the other a more direct measure of how much thought the respondent has given to the election being studied (Q.2). Both are strongly correlated with turnout, and deciding which to remove is difficult. Though Q.2 has, perhaps, more face validity since it asks directly about the current campaign, it has a related problem in that the proportion who answer it in the affirmative rises dramatically as the campaign season progresses and more voters have given thought to the election. Though this is not a fatal flaw in the usefulness of the item, it does lead to a certain level of instability as the index is applied months prior to election day. The distribution of responses to Q.6, by comparison, tends to remain largely stable across the election period.

Table 11 suggests that removing Q.2 has a dramatic effect on the way respondents are coded in the likely voter index. The categorization of fully 17% of respondents changes with the removal of this one item from the full 8-item Perry-Gallup index. Though the horserace estimate remains unchanged, this reorganization of respondents suggests that Q.2 plays a central role in the makeup of the original index. Given the substantive relevance of the amount of thought a respondent has given to the campaign, and the statistical centrality of Q.2 to the original 8-item index, we will test a 4-item scale that eliminates the slightly less essential Q.6 about general political interest.

Knowledge about Where to Vote

Neither measure of the practicalities of voting stands out as superior to the other in general terms. Both have similar response rates (90% say they know where people go to vote in their neighborhood, and 86% say they have voted in their current district). Table 11 suggests that each has a similar effect on the overall index, if removed. Roughly 5% of respondents

---

[8] During the 2000 election campaign, the incompatibility of the 1996 Presidential Vote item in the likely voter index was identified early in the summer by the Pew Research Center staff, and, even though it has traditionally been used, it was excluded from all likely voter analyses by the Center throughout the campaign season.

would be categorized differently with the removal of either item, and the percent correctly classified and the horserace predictions remain largely unchanged if either item is removed.

As a practical matter, Q.4, asking respondents if they know where people go to vote, is slightly more useful as an index item. As mentioned above, measures of past voting behavior are not applied to respondents who are under 22 years of age. Though we classify Q.5 here as a "practicality" since it addresses whether a person knows where to vote, it is also a measure of past behavior, and negative answers are not counted against those aged 18-21. If we remove Q.4 and keep Q.5, the categorization of respondents under 2 will be based solely on their interest and intention to vote. If we retain Q.4, however, it applies to all age groups.

Analyzing the 4-item trimmed index

Interestingly, the 4-item index produces likely voter estimates that are at least as good as the estimates based on the full 8-item Perry-Gallup index. Table 12 shows that though the interitem reliability drops, as would be expected with half the number of items, the index performs equally well at both classifying voters and non-voters correctly and estimating the preferences of actual voters. This accuracy is maintained even though more than a quarter of respondents change classifications as a result of the elimination of the other four index elements.

Table 12: Testing a Four-Item LV Index
Based on Wave 2 respondents (N=856)

	Reliability (alpha)	Percent of cases Changed	Percent Giving Answer	Percent Correctly Classified	Percent Democr. Overest.
Total (All RVs)			100%	77%	+2.7%
Original 8-Item LV Index	.64	--	70	73	+0.2
New 4-Item LV Index	.44	27%	70	74	-0.2

Though we are not suggesting that all researchers shift to such a truncated likely voter index in future election polls, this analysis supports a central argument of this paper – that any errors in the accuracy of likely voter indices have little to do with the size or comprehensiveness of the index overall. We have seen that single items, 4-item indices, 8-item indices and 15 item indices can all perform about equally well at separating voters from non-voters and deriving reasonably accurate estimates of the preferences of those who are bound to turn out.

Part 6: Probability Models

A different approach to computing a likely voter estimate is to apply regression analysis to determine a probability of voting for each respondent. A logistic regression procedure is used to model the independent effect of each likely voter indicator on actual voting turnout as derived from the validation study. In addition to deriving a coefficient measuring the relationship of each item to behavior, the procedure produces a predicted probability of voting for each respondent.

These probabilities can be used in two ways to create likely voter predictions of the election horserace. First, each respondent can be weighted by their predicted probability of voting according to the regression model. Unlike the standard methodology used above, where the top 70% of respondents are coded as "likely" and estimates are based solely on likely voters, the preferences of all respondents are taken into account, but those the model deems most likely to participate carry more weight.

A second approach is to apply our standard index methodology to the regression predictions. In other words, we can take the 70% of respondents the regression model derives the highest probability of voting for.

Table 13 compares the results of these regression-based analyses with our original eight-item likely voter index. First, taking the 70% of RVs with the highest predicted probability of voting according to the logistic regression model produces a horserace estimate virtually identical to the one achieved by taking the top 70% of the Perry-Gallup index, thought it does achieve a slightly higher percent correctly predicted.

Table 13: Regression Models
Based on Wave 2 respondents (N=856)

	Percent Predicted Likely	Percent Correctly Classified	Percent Democr. Overest.
Total (All RVs)	100%	77%	+2.7%
Original 8-Item LV Index	70	73	+0.2
Regression Model: Top 70%	70	76	+0.2
Regression Model: Weighted	--	--	+1.4

Using the predicted probability of voting as a weight meets with slightly less success. In this case, the candidate preferences of all voters are taken into account, though some are weighted more than others. The majority of RVs receive fairly high predicted probabilities, however (fully 50% are assigned probabilities of voting between 80% and 90%, with only 12% given a less than 50% chance of voting) which leads to a horserace prediction that is slightly too favorable to the Democratic candidate.

The validation study suggests that, for the extra effort involved, neither of these approaches produces election predictions that are any more accurate than our original index technique.

Part 7: Non-Response Error

Apart from the question of using a survey to predict who will vote is the related question of whether survey respondents in general are more likely than nonrespondents to vote. If our pool of respondents is a biased sample of the electorate, it may make it more difficult to use a comparison of estimated turnout among respondents against known parameters of turnout from election statistics. That is, if one means of calibrating a likely voter scale is the creation of an estimate of likely turnout among registered voters (e.g., 70%) or the voting age population (e.g., 50%) and establishing a cutoff within the survey at that percentage, we could miss many likely voters if the pool of survey respondents was already more likely to vote than the population from which it was drawn.

Beyond this practical issue is the broader concern that survey samples may overrepresent politically interested and active people. The evidence for this is mixed, but it remains a concern, especially in light of declining response rates (Brehm, 1993; Keeter et al. 2000).

There are many reasons to believe that survey samples, especially for political surveys, will overrepresent likely voters. One is that interest in the survey topic is a predictor of cooperation, and there is much evidence that people interested in politics are more likely to vote. Another is that most telephone survey samples overrepresent better educated and more affluent individuals, and education and income are predictors of voter turnout. And telephone surveys, especially if conducted over a relatively short period of time (as many election surveys must be), tend to underrepresent younger and more mobile individuals, who are less likely than average to vote.

The present study provides at least the possibility of comparing voter turnout between households that cooperated with the survey and those that did not. The Philadelphia voter registration list contains telephone numbers for many registered voters and these can be matched against the telephone numbers dialed in the two surveys (and of course, part of the sample was drawn from this list). Conceptually, we should be able to compare turnout in cooperating and noncooperating households. Unfortunately, there are a number of challenges to doing so.

First is the obvious problem that while we have telephone numbers for presumed households, more than one registered voter may live in the household. We have turnout records for all registered voters in the household, but for households from which no interview was obtained we do not know who claimed to have voted nor do we usually know who in the household refused an interview or was ultimately "responsible" for a noncontact. Thus we do not have the ability to make a direct inference about the connection, if any, between a voter's behavior and his or her cooperation with a survey. Among the possible solutions to this problem are (1) to describe the percentage of households in which *someone* cast a vote, or (2) to compute, for each household, the percentage of individuals therein who cast a vote, and to then report the average percentage for the group of households. Neither is ideal but both could provide a basis for a relatively unbiased comparison of households that cooperated and those that did not.

A second and perhaps more serious problem is the lack of certainty that the telephone number in the voter registration list is, in fact, the telephone number of the individuals whose voting behavior is documented on the record. For survey respondents from whom we obtained a name or other identifying information, we were able to achieve matches in a high percentage of cases and thus have a great deal of confidence in the validity of the analysis described earlier. But even though we were able to create a match, the telephone number in the voter registration list did not always correspond to the telephone number used to reach the respondent. For the listed portion of the sample, we found a match between the listed phone number and 83% of the validated cases matched. This means that 17% of the telephone numbers found in the voter registration list actually match a household (people) different from the one to which it was associated in the list.[9]

If this percentage were applicable to all of the records in the registration list (and not just to those cases for which we obtained an interview and a match), then our judgment about turnout would be compromised in nearly one-in-five cases. But the critical question for a comparison of turnout in households that yielded an interview and those that did not is whether the rate of telephone matching error is different for the two. Or, viewed another way, is there any reason to suspect that telephone numbers once associated with a registered voter but now assigned to a different household will be more or less likely to yield an interview than those where the phone number in the registration list actually reaches the individuals whose voting records are associated with that number on the list? The most likely difference may be mobility. In the survey among listed households for which there was a phone match, 92% reported having been in their home for at least two years. Among listed households for which there was not a phone

[9] A visual inspection of the voter registration list, sorted by telephone number, quickly reveals the nature of the problems. Numerous instances occur in which telephone numbers appear twice. In some instances, they are obviously linked with the same household (but different individuals in the household) because of slight variations in the addresses given by the registrants. In other instances, the duplicate numbers are linked with entirely different addresses and voters. The obvious implication is that voters have moved or died and their telephone numbers have been assigned to other households and thus other voters. This phenomenon is certainly more widespread than is evident solely from the duplicate numbers. Individuals previously registered with their phone numbers listed may have moved, changed phone numbers, and not corrected the phone number on the record.

match, only 72% had resided there for two years. But we have no reason to believe that recent arrivals would be significantly less likely to grant an interview, even though they may be less likely to vote.

With these various and perhaps somewhat convoluted caveats in mind, let us proceed to an examination of the data. All of the telephone numbers dialed in the surveys were matched against the voter registration list. Of the TK total numbers dialed, we were able to match 8,233 to the list. To help us remember that the phone numbers are not definitively associated with the household in the voting record, we will refer to phone/households. For each phone/household on the list, we coded voter turnout in the election, assigning a code of "voted" if any member of the phone/household turned out. For all 8,233 phone/households, 56.4% were coded as voting.[10]

Table 14 shows the turnout percentage by categories of survey cooperation. Someone voted in 69% of the phone/households that granted an interview. By comparison, someone voted in 57% of the phone/households that did not grant an interview. Our confidence that this is a valid comparison is undermined somewhat by the turnout figures in the other two categories in the table. Note that 55% turned out in phone/households deemed ineligible to participate in the survey. This group included phone/households discarded because a gender quota for the survey had been reached (and thus it is not surprising to find voters in those phone/households), but also included were phone/households out of the sample area and thus presumably ineligible to vote in the election, and phone/households in which no one over 18 resided. Even more troubling is the fact that someone turned out to vote in 34% of the phone/households determined to be nonresidential. That is, people associated with many telephone numbers in the voter registration list turned out to vote even though those telephone numbers no longer are in service, or are now associated with a business or government office. The logical conclusion is that many of these people changed their phone numbers but stayed within the city.

Table 14: Turnout Rates in Non-Responding Phone/Households

Final Disposition	Voted %	Did Not Vote %	N
1 Completion	69	31=100	1738
2 Complete -- ineligible	55	45	163
3 Nonrespondent - presumed HH	57	43	5162
4 Not a household	34	66	1170
Total	56	44	8233

[10] This figure of 56.4% for our matched sample of dialed numbers is very close to the 55.9% turnout found among all phone/households in the voter list with telephone numbers.

Our best comparison may be between respondent phone/households and those clearly determined to be residential households but which refused to participate. Table 15 shows voter turnout broken down by all of the disposition codes in the sample. Turnout in phone/households that refused to participate was lower than in those that participated, but the difference is not large. Compared with 69% turnout in phone/households that granted an interview, turnout was 61% in phone/households for which the final disposition was an answering machine, 60% in those for which a callback had been scheduled, 65% in which there was a single refusal, 65% in which there was a "soft refusal," and 62% in which a hard or second refusal occurred.[11]

Table 15: Turnout Rates by Disposition

	Final Disposition	Voted %	Did Not Vote %	N
3	other unaccounted for	100	0=100	2
5	line problems, incomplete	50	50	20
6	no answer	40	60	627
7	busy	46	54	112
8	**answering machine**	**61**	**39**	**730**
10	callback unspecified	42	58	12
11	**callback scheduled**	**60**	**40**	**2256**
12	initial refusal	65	35	20
13	**soft refusal**	**65**	**35**	**633**
14	breakoff	73	28	80
15	**second or hard refusal**	**62**	**38**	**338**
16	R away for duration	55	45	71
17	business/govt	44	56	173
18	computer line	38	62	133
19	disconnected	30	70	844
20	health, hearing	57	43	139
21	language	30	70	142
22	no one 18 in HH	54	46	41
23	outside sample	54	46	63
24	quota filled	58	42	59
25	**completion**	**69**	**31**	**1738**
	Total	56	44	8233

Thus it appears that voter turnout was slightly higher in households that cooperated with the survey than those that didn't (4-10%). What conclusions can we draw from this analysis? The first is that the differences are small enough that it's unlikely that an aversion to politics or lack of interest in the election is a serious factor in survey nonresponse, since there are good reasons to suspect that other sources of nonresponse will also contribute to the difference. Lack of interest in politics would be the most serious source of bias, but since we expect differential nonresponse by education and age (for reasons unrelated to attitudes toward politics and elections), the purely attitudinal component of the difference is apt to be very small.

A second conclusion is that voter registration records are a mess. Because the city of Philadelphia (like other localities) can no longer purge the registration rolls for lack of voter

[11] An alternative but perhaps slightly riskier basis of comparison uses the turnout of the first voter on the registration list for a household, rather than a measure of turnout by *anyone* in the household. Using the first voter, the differences between respondent and nonrespondent phone/households are somewhat larger. Among respondents, 64% of "first voters" turned out, compared with 54% among callback phone/households, 56% among answering machine phone/households, 58% among soft refusals, and 56% among hard or second refusals.

activity, the registration list is grossly inflated and contains numerous errors. Reliance on registration lists for estimating voter turnout rates is highly problematic.

REFERENCES:

Brehm, John. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.

Colasanto, Diane and Jay A. Mattlin. 1987. "Evaluation of Gallup's Methodology for Predicting Likelihood of Voting." Paper prepared for presentation at the 1987 Joint Statistical meetings in San Francisco, Aug. 17, 1987.

Committee of Seventy, The. 1999. "Seventy's Reflections On the November 1999 Election" Newsletter. Contact: 2 Penn Center Plaza Suite 770, Philadelphia, PA 19102.

Freedman, Paul and Ken Goldstein. 1996. "Building a probable Electorate from Preelection Polls: A Two-State Approach." *Public Opinion Quarterly* 60:574-587.

Scott Keeter, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64 (Summer): 125-148.

Perry, Paul K. "Certain Problems in Election Survey Methodology." *Public Opinion Quarterly* 43:312-325.

Traugott, Michael W. and Clyde Tucker. 1984. "Strategies for Predicting Whether a Citizen Will Vote and Estimation of Electoral Outcomes." *Public Opinion Quarterly* 48:330-343.

APPENDIX ONE: ACCURACY AND BIAS IN THE MATCHING PROCESS

One inherent problem in validation studies is accurately and evenly matching respondents to actual voting records. By focusing on a single political sector, the City of Philadelphia, many of the complications of gathering data from multiple voter registration offices was bypassed. However, certain complications remained.

In order to make the match to voter registration and turnout records, respondents who identified themselves as registered voters were asked for their name and address at the conclusion of the survey using the following text: "We are going to send you a small gift as a token of our appreciation for participating in this survey. If you like, we will also send you a copy of the final report. So that we can send you this gift, may I please have your name and current mailing address? We promise that this information WILL NOT be used to contact you at a later date." If a respondent refused, further assurances about confidentiality were made, including the provision of a 1-800 number that can be called to speak to a supervisor. If the respondent continues to refuse, they were asked to simply give their name, with no address.

Overall, 76% of respondents who identified themselves as registered voters were willing to give their name and address, with another 12% giving their name only. This information, in addition to the respondent's phone number and year of birth (if given) were used in the matching process. If a respondent refused to give their name, no match could be attempted. No attempt at a match was made for respondents who did not identify themselves as registered voters.

Overall, 70% of completed RV respondents were successfully matched using the Philadelphia voting records. The demographic distribution of the matching process is shown on the next page, in the column labeled "Total Percent Matched". We see here that our inability to match certain respondents did not distort the demographic makeup of the sample, for the most part. The gender distribution of the matched sample remained identical to the overall sample, as did the distribution of income levels, whites and blacks. We can see that our matching process had the most trouble identifying Hispanics (62% matched), respondents under 30 (67% matched), college graduates (66% matched), and political independents (67% matched).

The reasons for our difficulty in matching certain groups varies. College graduates, for example, are as easy to match to voting records as any other education group, however they were less willing to give us their name and address, and thus the proportion matched is lower than other educational categories. The table on the next page shows that just 68% of college graduates were willing to give us their address, and fully 17% wouldn't even tell us their names. Though we could match 88% of those giving us their full information, our overall percent matched falls because of the lack of identifying information on the rest.

We include the category "refused" under the income distribution to highlight how much trouble refusal to provide identifying information can cause. Not surprisingly, respondents who do not feel comfortable giving out their household incomes also tend to shy away from providing their names and addresses – just 56% did so, with

28% giving no identifying information. Even though we were able to match respondents who did provide identifying information as well as all other income levels, our overall percent matched was a mere 58% for this group, far below average. As a result, all analysis based on our matched sample underrepresents this demographic group.

On the other hand, Hispanics and young RVs were no less likely to be willing to provide their names and addresses when asked. In these cases, our relatively low matching rate derives directly from problems in the matching process itself. For example, we were only able to match 72% of Hispanic respondents who gave us their names and addresses, compared to 90% of Whites. And our difficulty finding younger voters, particularly young men, can also be seen in the relatively low percent of those providing their names and addresses who were successfully matched. Either these respondents were in fact not registered, or their names did not appear on the voter registration rolls due to delays in the registration process or respondents recently moving. Once again, these demographic groups are slightly underrepresented in the overall matched sample.

Overall, whether respondents were willing to give their name and address at the end of the survey was far and away the single most important factor in the success of the matching process. While we were successfully able to match 86% of those giving their name and address, we could match just 43% of those who gave their name only, and, of course, none of those unwilling to give us their name.

MATCHING AND VOTER PARTICIPATION BY DEMOGRAPHICS
(Based on self-reported registered voters, all waves and samples)

	Percent Offering Identifying Information			Percent of N&A Matched	Total Percent Matched	N-Matched/ N-Respondents	*Percentage of matched RVs who…*	
	Name & Address	Name Only	Ref-used				Voted	Didn t Vote
	%	%	%	%	%		%	%
Total	**76**	**12**	**12=100**	**86**	**70**	**(1694/2415)**	**76**	**24=100**
Sex								
Male	78	10	12=100	85	70	(796/1134)	75	25=100
Female	74	13	13=100	87	70	(898/1281)	77	23=100
Race								
White	72	15	13=100	90	72	(877/1226)	80	20=100
Non-white	82	8	10=100	82	70	(774/1104)	71	29=100
Black	83	8	9=100	82	71	(684/967)	73	27=100
Hispanic*	84	6	10=100	72	62	(86/138)	55	45=100
Age								
Under 30	83	11	6=100	76	67	(246/366)	50	50=100
30-49	78	11	11=100	84	70	(657/946)	72	28=100
50-64	73	13	14=100	91	72	(340/473)	87	13=100
65+	75	11	14=100	93	77	(419/547)	87	13=100
Sex and Age								
Men 18-29	87	9	4=100	72	65	(113/175)	50	50=100
Men 30-49	81	10	9=100	85	72	(354/489)	71	29=100
Men 50+	73	11	16=100	92	73	(322/444)	87	13=100
Women 18-29	81	12	7=100	81	70	(133/191)	50	50=100
Women 30-49	75	13	12=100	84	66	(303/457)	73	27=100
Women 50+	75	13	12=100	92	76	(437/576)	87	13=100
Education								
College Grad.	68	15	17=100	88	66	(412/621)	84	16=100
Some College	77	11	12=100	88	73	(370/510)	75	25=100
H.S. Grad	79	12	9=100	83	71	(690/968)	73	27=100
<H.S. Grad.	83	5	12=100	88	74	(212/285)	71	29=100
Family Income								
$75,000+	69	18	13=100	93	73	(143/196)	83	17=100
$50-$74,999	80	11	9=100	88	73	(221/302)	85	15=100
$30-$49,999	81	10	9=100	84	72	(405/566)	73	27=100
$20-$29,999	85	9	6=100	83	75	(249/332)	74	26=100
<$20,000	84	9	7=100	85	75	(383/510)	69	31=100
Refused	56	16	28=100	88	58	(293/509)	80	20=100
Party ID								
Republican	77	12	11=100	84	70	(271/390)	78	22=100
Democrat	79	10	11=100	88	74	(1094/1484)	77	23=100
Independent	72	15	13=100	82	67	(260/391)	67	33=100

* The designation Hispanic is unrelated to the white-black categorization.

APPENDIX TWO: LIST VS. RDD SAMPLES

The 1999 Philadelphia Validation Study also included a test of the effectiveness of different sampling techniques as a basis for matching names to voting records. Roughly one-third of both the Wave One and Wave Two samples were drawn not from a standard RDD base, but directly from the Philadelphia voter registration lists. The experiment was designed to see if matching was substantially easier with numbers drawn directly from registration records.

The results say more about the inaccuracy of the Philadelphia voter registration records than about the matching process. Overall, we were able to match 75% of respondents in our list sample (the sample drawn from registration records), and 68% in our RDD sample. Despite a slightly higher matching rate from the list sample, it is remarkable that we were unable to match fully one-in-four respondents to the very list that they were drawn from. Clearly, many telephone numbers connected with names on the voter registration list have been reassigned to new residents, with no change on the registration rolls. As mentioned earlier, in 1999 the Philadelphia voter registration list identified 985,912 registered voters, in a city with only 1,056,764 age-eligible residents.

The bulk of the likely voter analysis in this report is based on all wave 2 respondents, regardless of whether they were drawn from the list or RDD sample. Though unusual, there are numerous reasons for this analytical choice, not the least of which was the need to have a sufficient number of cases on which to base our election estimates. More importantly, however, we believe that though each wave's respondents are drawn from two distinct sampling frames, they are, in effect, all part of an identical pool of registered voters. In other words, in order for either a listed or RDD respondent to be included in our analysis, we had to be able to match their name to the registration lists in our possession. In this sense, though the RDD sample may have represented a more representative cross-section of the Philadelphia population at the time, all respondents analyzed were fundamentally pulled from the same frame of registered voters.

More importantly, a comparison of the demographic distribution of the list and RDD samples finds very few substantive differences, further supporting our analytical decision to merge the two datasets.

One unfortunate outcome of this merging procedure is our inability to effectively weigh the data to mirror the overall Philadelphia population. Since our sample is effectively drawn from the Philadelphia registration records, and not a random sampling of the population, no clearly identifiable weighting parameters were available.

APPENDIX THREE: DISPOSITION OF TELEPHONE NUMBERS AND RESPONSE RATES

The following table presents the final dispositions of all telephone numbers dialed in the survey. The tabulation is based on a hierarchical analysis of all call records for the study, in which each number is assigned a disposition based on the highest-coded outcome of any call during the calling period. For example, a number that yielded four "no answer" results (code 6 in the table below) and one "initial refusal" (code 12) would be coded as "initial refusal." Insofar as the coding system employed by the interviewers permitted, this tabulation follows the guidelines for presentation of final dispositions as described in the document, " Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys."

	Wave One: Listed		Wave Two: Listed		Wave One: RDD		Wave Two: RDD	
	N	%	N	%	N	%	N	%
3 other unaccounted for	2	0.1	--	--	--	--	--	--
5 line problems, incomplete	7	0.3	15	0.6	1	0.0	1	0.0
6 no answer	234	8.4	226	9.2	991	13.8	1074	15.0
7 busy	42	1.5	36	1.5	206	2.9	202	2.8
8 answering machine	234	8.4	278	11.3	412	5.7	617	8.6
10 callback unspecified	3	0.1	8	0.3	--	--	1	0.0
11 callback scheduled	695	24.9	539	21.9	1587	22.1	1801	25.2
12 initial refusal	4	0.1	15	0.6	2	0.0	7	0.1
13 soft refusal	196	7.0	202	8.2	524	7.3	223	3.1
14 breakoff	24	0.9	24	1.0	34	0.5	50	0.7
15 second or hard refusal	106	3.8	107	4.4	213	3.0	178	2.5
16 R away for duration	27	1.0	9	0.4	76	1.1	35	0.5
17 business/govt	49	1.8	55	2.2	839	11.7	763	10.7
18 computer line	51	1.8	36	1.5	289	4.0	272	3.8
19 disconnected	417	15.0	312	12.7	528	7.4	536	7.5
20 health, hearing	45	1.6	26	1.1	75	1.0	63	0.9
21 language	50	1.8	54	2.2	121	1.7	108	1.5
22 no one 18 in HH	4	0.1	9	0.4	30	0.4	103	1.4
23 outside sample	16	0.6	20	0.8	63	0.9	46	0.6
24 quota filled	60	2.2	--	--	5	0.1	1	0.0
25 completion	521	18.7	488	19.8	1173	16.4	1060	14.8
TOTAL	2263	100.1	2041	100.1	5512	100.0	5569	99.7

The final dispositions listed above were used to compute response rates for the four samples. These are shown in the following table. The completion percentage is the sum of the "completes" and the "complete -- ineligible" cases (the latter include households that yielded an interview but reported that they were outside of the city limits, or that included no one 18 or older, or were over the gender quota). The computation follows the AAPOR guidelines for the calculation of response rate 1 (RR1), the minimum response rate. Numbers with a final disposition of "business/government," "computer line," and "disconnected" are excluded as nonhouseholds. All other numbers are included in the denominator.

	Wave One: Listed		Wave Two: Listed		Wave One: RDD		Wave Two: RDD	
1 Completion	521	23%	488	24%	1173	21%	1060	19%
2 Complete – ineligible	80	4%	29	1%	98	2%	150	3%
3 Nonrespondent – presumed household	1662	73%	1524	75%	4241	77%	4359	78%
TOTAL	2263	100%	2041	100%	5512	100%	5569	100%

APPENDIX FOUR: SURVEY METHODOLOGY

The data used in this analysis are from two telephone surveys (Wave 1, Oct 12-21, 1999; Wave 2, Oct 27-30, 1999) conducted in the City of Philadelphia by the Pew Research Center for the People & the Press, under the direction of Schulman, Ronca and Bucavalis, Inc. Each wave of the survey consists of approximately 1,600 interviews, drawn from two distinct samples (see below for details). Roughly two-thirds of respondents in each wave were drawn from a standard random-digit sample of telephone numbers selected from telephone exchanges in the City of Philadelphia. The random digit aspect of the sample is used to avoid "listing" bias and provide representation of both listed and unlisted numbers (including not-yet-listed). The design of the sample ensures this representation by random generation of the last two digits of telephone numbers selected on the basis of their telephone exchange and bank number.

The other third of each wave is drawn directly from voter registration lists maintained by local government agencies. Registration lists were used to identify households encompassing at least one registered voter, with standard household randomization applied once telephone contact was made. This alternative sampling methodology was utilized to test whether "matching" survey respondents to voter registration lists more or less efficient using different sampling techniques. Though there are many possible sources of bias in the voter-list sample (not all registered voters provide a phone number when registering, registration records may not be completely up-to-date), the respondents drawn from each separate sampling procedure were similar in most demographic and political characteristics.

For both RDD and listed samples, numbers were released for interviewing in replicates. Using replicates to control the release of sample to the field ensures that the complete call procedures are followed for the entire sample. At least 10 attempts were made to complete an interview at every sampled telephone number. The calls were staggered over times of day and days of the week to maximize the chances of making a contact with a potential respondent. All interview breakoffs and refusals were recontacted at least once in order to attempt to convert them to completed interviews. In each contacted household, interviewers asked to speak with the "youngest male 18 or older who is at home." If there is no eligible man at home, interviewers asked to speak with "the oldest woman 18 or older who is at home." This systematic respondent selection technique has been shown empirically to produce samples that closely mirror the population in terms of age and gender.

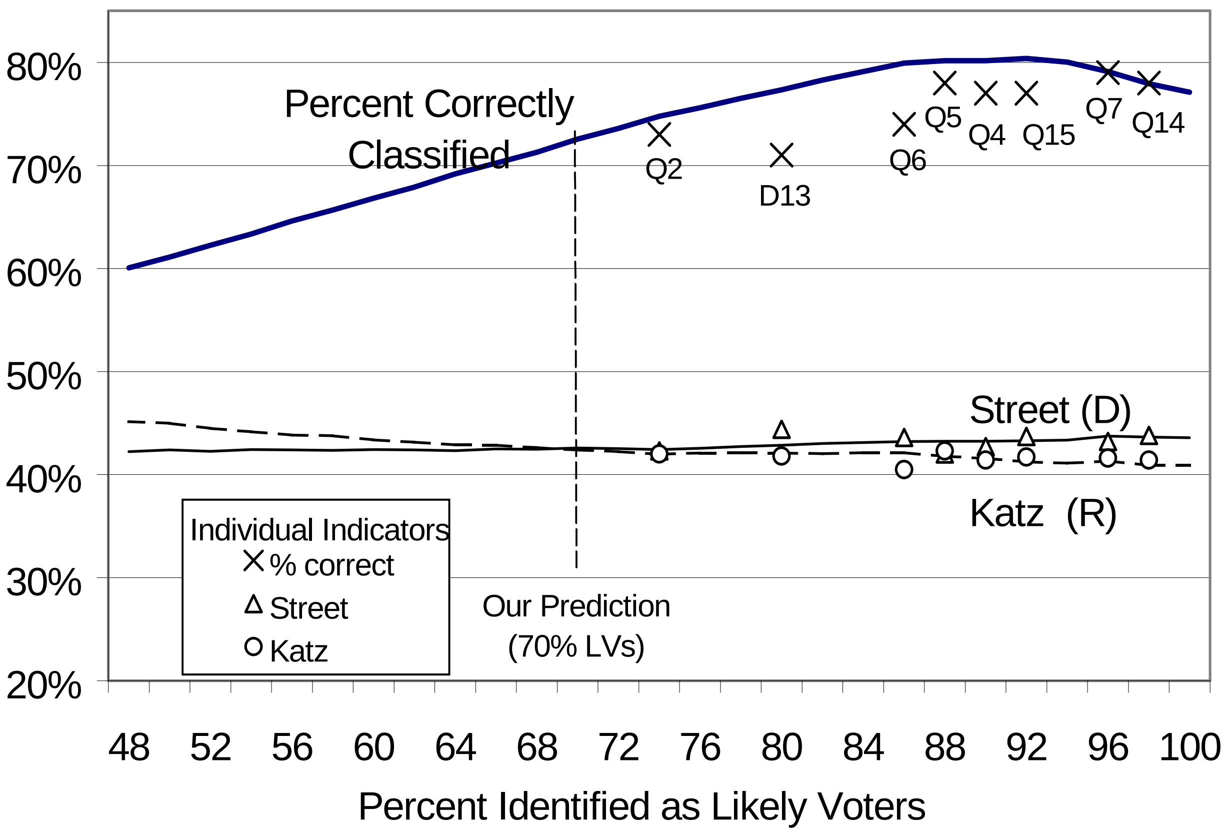
Survey respondents were matched to voter registration lists after election day to validate their voting behavior. The matching process took into account five parameters: phone number, first name, last name, address, and the respondent's age. Overall, we successfully matched 70% of respondents who identified themselves as registered voters to the voter registration lists (68% from the combined RDD samples, 75% from the combined listed samples). The inability to match 30% of respondents who claim to be registered reflects three factors, each of which might affect the representativeness of the sample. First, many respondents overreport voter registration. Second, many respondents refused to give their name and address, making matching difficult or impossible. Third, registration lists maintained by local government agencies may not be completely up-to-date. See Appendix One for a more complete analysis.

Non-response in telephone interview surveys produces some know biases in survey-derived estimates because participation tends to vary for different subgroups of the population. Both respondents and non-responding households were "matched" to voter registration lists, in order to gauge the relationship between survey participation and turnout (see analysis in text).

Though each wave of telephone interviewing is drawn from two separate sampling frames, the analysis of likely voter methodology is based only on matched cases which, in effect, are all drawn from the same sampling frame of the registration lists of local government agencies. As a result, all analysis is conducted on the combined listed and RDD samples. Data are not weighted to census parameters due to the fact that the registration lists do not represent a random distribution of the city's population.

	Wave One			**Wave Two**		
Number of Cases	Listed	RDD	**Total**	Listed	RDD	**Total**
Total Respondents	509	1170	**1679**	490	1060	**1559**
Registered (Self-Reported)	413	801	**1214**	400	801	**1201**
Matched	302	536	**838**	306	550	**856**

FIGURE 1: Comparing Accuracy of Original LV Index
to Accuracy of Individual Indicators



Based on Wave 2 Respondents
(N=856 Registered Voters)

Among the 77% (N=659) we
confirmed as voters:
42% favored Street
42% favored Katz