poll showed that education was the most important issue, and voters that had education on their minds supported Democratic candidates by a two-to-one margin. Since the election, national surveys continue to show a great deal of support for more spending on education and closer monitoring of HMOs; and the fact that the public is more likely to trust Democrats to

Reading the Electorate: Internet-Based Polls Were Shown to Work

By Humphrey Taylor

The most dramatic and revolutionary polling event this past election season was the successful use of Internet-based polling, a development which marks an important milestone in polling history.

• In 1936, Gallup's success, and the *Literacy Digest*'s failure, established the credibility of "scientific polling," and showed that sample design is more important than sample size.

• In 1948, Truman's come-from-behind victory discredited quota sampling (in the US, if not elsewhere), and pointed to the need for continuing to poll as late as possible.

• In 1980, telephone polls replaced in-person polls as the

Table 1:
The Increasing On-Line Population

	All adults using the Internet
1998	
October/November	45%
May/June	37
1997	
November/December	36%
May/June	30
April/May	27
1996	
September	19%
April	15
1995	
November	10%
September	7

Note: "On-line" is defined as accessing the Internet, the World Wide Web, or an on-line service at home, work, or some other location. The data for 1995 and 1996 are based on nationwide cross-sections of 1,000. The data for 1997 and 1998 are based on rolling averages for samples of 2,000 adults.
Source: Surveys by Louis Harris and Associates.

handle important issues like these suggests Republicans lost seats because they had the wrong agenda, not because they didn't have one at all.

From now on, I am going to pay more attention to the numbers and less attention to the conventional wisdom. I hope you do, too.

preferred methodology for most opinion polls and election predictions.

Now, a new landmark must be added to this list:

• In 1998, Internet-based polling was shown to be a reliable methodology for opinion polls and election prediction.

The remorseless rise of the Internet, like the Energizer Bunny, just keeps on going. According to the latest Harris data, based on 2,023 interviews conducted in October and November, fully 45% of all adults have access to the Internet (see Table 1). This includes people who access the Internet from their homes, their workplaces, their colleges, or other locations.

The pace of Internet growth has been truly astonishing. Internet penetration—using Harris' definition—has grown from 7% of all adults in September 1995, just a little over three years ago, to almost 20% by September 1996, to 30% in June 1997, to 37% by mid-1998. By early 1999, more than half of all adults will be on-line.

The latest Harris numbers also show that, more and more, the demographic profile of the on-line population resembles the whole country. As the total number of people on-line has increased, the differences between the Internet population and the total population have diminished. For example, African Americans have grown from 1% of the on-line population to 11%, and women from 21% to 49% (see Table 2).

However, the on-line population in general, and those choosing to respond to our *Harris Poll On-Line* surveys in particular, are still very different from the total population. Quite substantial weighting is needed to correct for biases in telephone surveys, particularly with relation to age and income. Even more weighting is needed to correct for the substantial biases in the on-line samples.

Telephone and On-line Survey Methodologies Compared

The most common criticism of on-line polling is that, unlike telephone research, it is not based on probability sam-

Table 2: Demographics of Internet Users—
Increasingly Representative

	1995	1997	1998	All adults*
Black	1%	8%	11%	12%
Hispanic	9	8	10	8
Women	21	46	49	52
College	57	38	34	22
Under $25,000	15	18	18	27
65+	3	3	5	16

*Current Population Survey.
Source: Surveys by Louis Harris and Associates.

pling. However, this ignores the nasty truth that even good telephone polls when conducted over three or four days—even with the best designs, RDD, and multiple call-backs to reach those not available and convert refusals—often fail to interview 60 to 70% of all possible eligible respondents because of refusal rates (typically 35 to 45%) and non-response (typically estimated at 30% or more). Indeed, the great majority of both in-person and telephone surveys have such low total response rates that they should really be described as using convenience samples, not probability samples.

This does not mean that they are inaccurate or unreliable; the track record of both in-person and telephone polls in predicting presidential elections is remarkably good. These polls have been accurate in spite of the fact that they use convenience samples. It turns out that the people who can be reached and are willing to be interviewed by telephone are—after weighting by five or six demographic variables— like the total adult population, at least in terms of voting behavior.

How Well Did the Internet-Based Polls Do?

The on-line polls of the 1998 mid-term elections have established that Internet-based polls (which also rely on convenience samples) can produce remarkably accurate election predictions, but only after the data are weighted by demographic and other variables.

The *Harris Poll On-Line* invited residents of fourteen states who had signed up as members of our on-line panel to be surveyed about their votes in 22 elections for senator, governor, or both. The polls correctly projected the winners in 21 of the 22 races. The one misstep was the failure to predict Roy Barnes' victory in Georgia; interestingly, all the telephone polls in Georgia also failed to get this race right.

While 21 correct predictions out of 22 is not bad, this was a learning experience. The average error on the forty-four (22 x 2) main candidates was four percentage points, which is not good enough, even if it compares quite favorably with the early days of in-person polling in the 1930s and 1940s and is nearly identical to the average error for telephone polls on these races in 1998. (In fact, there were five telephone polls which "picked the wrong winners" in the twenty-two races.)

Remember Quota Sampling?

The success of the Internet-based election forecasts will be less surprising to European researchers than to Americans. They never stopped using quota samples—almost the ultimate in convenience sampling—and have never been as committed to probability sampling as have Americans. They never suffered the trauma of the 1948 presidential election. American researchers have often been shocked by the widespread use of quota sampling in Europe, and by the acceptance of quota sampling by respected, and widely published, European political and other social scientists.

There are good reasons to believe that on-line polls will continue to become more accurate. As the percentage of the on-line population grows, it will become more and more like the total population. And, with each passing month, we are learning more about how to weight these data.

As on-line polls get better, telephone polls may get worse. Response rates have been declining because of public resistance to telemarketing and the use of answering machines to screen calls.

It won't be long, I believe, before most opinion polls are conducted on-line.

AATP