

Date: 15 January 1996

To: NES Board

From: Larry M. Bartels

Re: Issue scales versus "effort" items on 1995 Pilot Study

The '95 Pilot Study included a series of experiments comparing traditional issue scales with "effort" items in the areas of environment, crime, welfare, and jobs. In the first case we have responses in both formats for all respondents (D3z versus D1a) but candidate placements only for the issue scale. For the other issues we have split-half comparisons (G1x, G3x, and G4x versus G5, G6, and G7) with candidate placements in both formats. Our aim was to test whether the shorter and cheaper "effort" items might usefully be substituted for traditional issue scales, and whether they could be used for candidate placements as well as self-placements.

Marginals. By comparison with the issue scales, the "effort" items have substantially higher mean values (.71 versus .44) and somewhat smaller standard deviations (.30 versus .42). The difference in means is especially large when the anti-effort side of the issue scale offers a potent counter argument to government activism (most notably "address conditions that cause crime" and "let each person get ahead on their own"). The "effort" items generate less missing data than the issues scales (1.5% versus 4.7% for self-placements, 9.2% versus 14.3% for Clinton placements, and 18.3% versus 24.3% for Dole placements).

Reliability. In each of the four relevant issue areas we have items in the 1994 study from which to compute test-retest reliability. These are the traditional "jobs and standard of living" 7-point scale and more/same/less spending items on environment, crime, and welfare. In the former case the 1994 measure correlates with both 1995 formats at .48. In the three cases in which the 1994 measure is in a format more similar to the 1995 "effort" items than to the 1995 issue scales, the correlations with effort average .41 and the correlations with the issue scales average .28. It appears from these test-retest correlations that the "effort" format is, if anything, more reliable than the issue scale format (and more reliable than many of our other standard measures).

Correlates. By comparison with the issue scales, the "effort" items are more strongly correlated with ideology (average .23 versus .18) and party identification (average .27 versus .20). In four separate regressions of Clinton thermometer ratings, Gingrich thermometer ratings, domestic spending versus budget deficit trade-offs, and domestic spending versus taxes trade-offs on party identification, ideology, education, age, race, and issue preferences, the "effort" items get coefficients that are, on average, about 10 percent smaller than the issue scale items, with standard errors that are, on average, about a third larger. The "effort" items do relatively better in the trade-off regressions and less well in the thermometer regressions. In both formats, the environment and government jobs items do better than the welfare and crime items. In a regression of Clinton thermometer ratings on party identification, ideology, and issue distances, the distance measures based on "effort" items get an average coefficient of

9.2 (with an average standard error of 3.8), while the distance measures based on issue scales get an average coefficient of 7.6 (with an average standard error of 3.6). In a parallel analysis for Dole, the distance measures based on "effort" items get an average coefficient of 3.4 (with an average standard error of 3.6), while the distance measures based on issue scales get an average coefficient of 10.7 (with an average standard error of 3.4). None of the individual items does noticeably better in either format.

Recommendations. The "effort" items seem in most respects to perform about as well as traditional issue scales, and sometimes better. Obviously, not every issue can be adequately addressed with an "effort" item. (For example, the crime effort item attracts 90 percent support and fails to tap the distinction between ameliorative and punitive strategies posed by our new issue scale; it shows no effects on Clinton and Gingrich thermometer ratings, whereas the issue scale gets t-ratios of 2.6 for Clinton and 1.6 for Gingrich.) And we don't know whether more/less "effort" items are better or worse than the more/less "spending" items included in previous surveys. (We dropped that experiment from the Pilot Study because we strongly suspected that the answer would be "neither.") However, since either "effort" or "spending" items take less than half as long to administer as issue scales with follow-ups, they should be used whenever possible to cover issue areas that we otherwise might not have room for on our surveys.