EVALUATING AND CHOOSING VALUES/PREDISPOSITIONS
FOR INCLUSION IN NATIONAL ELECTION STUDIES SURVEYS

Richard G. Niemi
University of Rochester

Discussion paper for the Conference on Values and Predispositions, National Election Studies,
I. WHAT WE HAVE DONE

Since 1984 we have devoted time, energy, and Pilot Study or regular survey space to assessing more than a dozen Values/Predispositions (perhaps as many as 20, depending on how one counts various subscales and revised scales). Despite this effort, there has been little continuity in the measurement of values. Only Equalitarianism, Racial Prejudice, and Moral Traditionalism have been measured in the same way over four or more surveys. If one thinks, as I do, that the real value (so to speak) of measuring Values and Predispositions comes from assessing them over a long time, we have barely begun to make the kind of investment that is likely to have a real payoff.

II. HOW WE SHOULD EVALUATE WHAT WE HAVE DONE

In the nine evaluation reports on values/predispositions written since 1984, the following nine questions have generally been asked in evaluating the performance of proposed items and scales:
1) Are the questions answered by almost all respondents? 2) Are the marginals not too skewed? 3) Are the items, if asked in a Likert format, half "positive" and half "negative"? 4) Does each item correlate moderately highly with others in the proposed measure? 5) Do the items collectively form a reliable scale (usually measured as a high alpha value)? 6) Is the resulting scale correlated with criterion variables as expected? 7) Are the scales related to at least a few important dependent variables in a regression equation with "standard" controls (usually party identification, liberal/conservative self-placement, gender, race, education, perhaps income)? 8) Do the items form coherent and distinguishable factors in a factor analysis? 9) (Receiving less attention due to absence of panel data) Are responses to the items stable over time?

On the basis of these criteria, the scales measuring Equalitarianism, Racial Prejudice, Moral Traditonality, and Governmental Interventionism are evidently very good. It is obvious from the relevant codebooks that Criterion 1 is satisfied—i.e., that almost all respondents answer the component items. Tables 1-4 of NES Technical Report #50, prepared for this conference, show that the distributions of responses are not highly skewed (Criterion 2). There is only one exception—the very first item in Table 1 (Tech Report), which shows 90 percent agreeing that everyone should have an equal opportunity to succeed—and even here there is diversity in the strength of agreement. Tables 1-3 in the Report show that the scales using an Agree/Disagree response format contain equal numbers of positively and negatively worded items (Criterion 3).

Success with respect to Criteria 4-5 is somewhat harder to judge because inevitably there are variations with respect to the fit of component items and scale reliability. Nevertheless, the items are generally quite well correlated with one another, and scale reliability is good (for a summary, see Table 1 of this paper). It is true that selected inter-item correlations are very low. See especially Table 5-A of the Technical Report, in which several of the entries are less than .10, and several others are between .10 and 15. Still, even these items are correlated at higher levels in other years, and, as various memos pointed out, some reduction of inter-item correlations and reliability scores comes from the fact that the items alternate between positive and negative wording (see below). (In addition, these are Pearson correlations, and one might argue that
gamma coefficients, which usually register higher values, are theoretically more appropriate.)

Criteria 6-7 are still more difficult to test, and we do not provide new evidence here. (Part of the difficulty is that one can argue about which variables should be correlated with any given scale and which should be significantly related to the scale in a regression context.) In earlier memos, however, versions of at least three of the four scales have been related to other variables with very positive results. The clearest test is for Equalitarianism. Feldman (9/29/87) regressed 18 dependent variables on a scale based on precisely the items shown in the Technical Report (along with other appropriate independent variables); in 2/3 of the regressions (especially those involving policies rather than ratings of individuals), the Equalitarianism scale was significant. For Governmental Interventionism, Markus (2/1/90) correlated a scale containing three of the current items and one additional item with some 41 other variables, about 85 percent of these variables were significantly related to his scale. Regressions (with controls) also showed the scale to be significantly related to a number of policy areas. For Moral Traditionalism, Conover and Feldman (1985) regressed 37 dependent variables on an eight-item scale (the four items shown in the Technical Report and four related measures) and appropriate controls, in about 70 percent of the cases (higher if one excludes candidate evaluations) the morality scale was significantly related to the dependent variable.

Criterion 8 is difficult to assess because of the multiple ways in which one can do a factor analysis. Still, on the basis of a confirmatory factor analysis using the items making up the four values analyzed here, the results are quite positive (Tables 9A-9C of the Technical Report). There are variations with respect to the fit of component items, but all have significant factor loadings on the anticipated scale. Few items fit on more than one scale, and when they do, the factor loading is always much greater on the anticipated scale. There is some overlap between the Equalitarianism and Racial Prejudice scales (average scale correlation = .39), but I think they are sufficiently distinct conceptually and empirically so as not to be a problem.

Finally, over-time stability (Criterion 9) is, well, about what we have come to expect. For the individual items, average coefficients in 1990-92 were .33, .51, .38, and .40 for Equalitarianism, Racial Prejudice, Moral Traditionalism, and Governmental Interventionism, respectively (Tables 10A-10C of the Technical Report).1 None of the correlations is especially low, the minimum being .27 for one of the Equalitarianism items. Moreover, the stability coefficients for the scales, with corrections having been made for attenuation, are all above .75.

This is not to suggest that there is complete agreement about the properties and virtues of these scales. In various memos, suggestions have been made about how one might improve them, and doubts have been expressed about just what they measure. Feldman (9/29/87, 2-3), for example, suggested replacing two items in the Equalitarianism scale with two new items. Markus

---

1These coefficients are in the same range as the two-year correlations for issues in the 1972-74-76 NES panel (Markus and Converse). The noticeably higher coefficient for Racial Prejudice even matches nicely the higher coefficients for moral and civil rights issues.
(2/1/90, 5) hypothesized that one of the Governmental Interventionism items belonged to a distinct dimension. Stoker (9/87) suggested that a scale of "moral conservatism" was better than the measure of Moral Traditionalism. And I assume that everyone would, other things being equal, prefer more items, especially for the Governmental Interventionism scale, which is made up of only three questions.

In short:

1) Equalitarianism, Racial Prejudice, Moral Traditionalism, and Governmental Interventionism are measured quite well, and we have a short time series for each of them.

2) These four scales can almost certainly be improved, at least marginally, by tinkering with the items, but adding even one item, especially to the Governmental Interventionism scale, would probably outweigh most such tinkering.

3) Any "improvements" would destroy the modest continuity that we have thus far achieved. Even adding an item creates problems of comparability.

4) Other measures have been proposed over the years, but none seems to have had the measurement qualities and substantive interest to warrant inclusion in more than one or two regular NES surveys.

I would add one qualification to these up-beat conclusions:

5) If one regards values/predispositions as "fundamental" or more basic than attitudes on current policies and therefore expects them to be nearly invariant over short periods of time, one might interpret the over-time correlations as evidence that we have done a poor job of measuring them. (One might argue that the correlations for three of the scales—even when corrected for attenuation—are "only" on the order of .80 and that the correlations for the individual items are no higher than those for policy questions.) My sense is that most of us have come to expect only moderate correlations for individual items due to measurement error and that the correlations for the scales are high enough for us to say that we have done a good job of measuring these values (especially considering that not all error can be removed). Yet interpretation of the over-time correlations is sufficiently important and contentious that it should be a matter of discussion.

III. SHOULD WE CHANGE RESPONSE FORMATS?

As I reviewed what we have done, the matter of response formats leaped out at me as ripe for discussion. For all of the items in the Equalitarianism, Racial Prejudice, and Moral Traditionalism scales, the response format is Agree strongly, Agree somewhat, Neither agree nor disagree,
Disagree somewhat, and Disagree strongly. Not surprisingly, an "agree" response set occurs. Conover and Feldman (1985), for example, detected it with an unrotated factor analysis in which all of the positively worded items loaded on the first factor and all of the negatively worded items loaded on the second factor (see also Feldman 10/19/83, 2, 4-5). Using all four scales, a response set is easy to detect simply in the pattern of correlations and distributions.

Consider Table 2 of this report. For each of the agree/disagree items, the first column shows the average inter-item correlation when the two items were worded in the same direction—e.g., the first and third items in Table 1 of the Technical Report, where both are worded such that agreement was a pro-equalitarian response. The second column shows the correlation when the two items were worded in the opposite direction—e.g., the first and second items in Table 1 of the Technical Report. For the Governmental Interventionism scale, which was a forced choice type item, the first column shows the correlation when the two items were worded so that the first response meant the same thing (anti-interventionism or pro-interventionism), and the second column shows the correlation when the two items are such that the first response meant opposite positions on interventionism.

As is apparent, the correlations are markedly dependent on the direction in which the items are worded when the response format is agree/disagree. When the items are worded in the same direction, the correlations average .19 higher than when they are worded in the opposite direction. In contrast, whether pro-interventionism is the first or second response makes virtually no difference under the forced choice response format.²

Table 3 of this report contains a similar message. It shows the percentage scoring "high" on each of the four scales when high is the agree response compared to when it is the disagree response (the first compared to the second response in the case of Governmental Interventionism). When the agree/disagree format is used, the percentage scoring high differs, on average, by 31 points. When the forced choice format is used, the difference is only nine points.

But does the existence of a response set make any difference, at least so long as the items are balanced between positive and negative wordings? Perhaps not. Indeed, one could even argue that there is no need for balancing, at least if we do not interpret the responses as some sort of absolute measure of each concept. If there is error in each item due to a response set, the error should be roughly the same each time the item is asked; if our interest is in change over time, the error is merely a constant added to the parameter of interest.

On the other hand, I don't think we know enough about response sets to be sure that this sort of error is unimportant for typical analyses. Indeed, I suspect that this error will sometimes heighten and sometimes diminish relationships of interest and that we cannot always be certain of the direction in which the error is taking us. Moreover, I am generally uncomfortable knowing that:

²These results explain the fact that Governmental Interventionism is more reliable than Moral Traditionalism and almost as reliable as Racial Prejudice despite having three rather than four items.
(a) we are aware of systematic response errors; (b) we know how to avoid them; and yet (c) we do not.

But suppose we do want to eliminate the error? Do we simply switch to the forced choice format of the Governmental Interventionism scale? There are two problems. First is the obvious but vexing matter that we are tossing away up to 10 years of comparable measurement. Second is the fact that the forced choice format was itself abandoned quickly (for issue questions) when it was realized that it provided only two responses (and usually, as here, a small percentage of volunteered "middle" responses). When scaled, a dichotomous item provides only as many positions as there are items (e.g., the pro-interventionist position on 0 to n items). In the case of issue questions, the solution was to move to seven-point scales, which contain both the forced choice feature and a relatively large number of response alternatives. Should we do any less for measuring values?³

Surely this point is worth brief consideration.

IV. HOW MANY VALUES, AND WHICH ONES, SHOULD WE MEASURE?

For purposes of discussion, let me suggest the following as guidelines for future NES surveys.⁴

(1) Include as a regular, continuous item no value or predisposition that is measured with fewer than four items. Rationale: a) items sometimes change meaning, marginals become skewed, and fit with other items changes; b) scale reliability is enhanced by a larger number of items (assuming reasonable inter-item correlations); c) it is more likely that multiple questions will capture multiple components of the value or predisposition. Exceptions would have to be made in extraordinary circumstances (Materialism/Post-Materialism is a case in point), but any such exceptions should be rare.

(2) The scale properties of a value have to be good over the entire period for which they have been included in Pilot or regular NES studies. Rationale: Most measures have been used over a fairly small number of election studies. If a given question or scale performs poorly one or more of these times, this suggests that it will not be good over the long haul. Comments: Criteria 1-2, 4-5, and 8-9 cited above speak most clearly to the validity of the scale. Qualities 6-7 are important, but our theories usually do not pinpoint appropriate criterion variables, and it is likely that one can easily find at least some variables with which any reasonable scale is correlated and

³If one takes Krosnick and Berent's (AJPS, 1993) recent work at face value, one might even wonder whether we should skip over seven-point scales and use a branching format.

⁴My thinking is based on the assumption that NES will continue very much in its present format--pre- and post-election interviews with a representative cross-section of adults, with a heavy emphasis on voting. Were one to focus very heavily on values/predispositions, one might consider major design modifications, such as oversampling young adults, starting from scratch in the search for which values to include, consideration of different auxiliary samples, and so on.
for which it has predictive power.

(3) The values or predispositions that are retained should be those that arguably measure
important, long-term characteristics of American culture. Rationale: The most valuable feature of
values and predispositions will come from long-term measurement—especially from being able to
detect changes in such characteristics. Simply confirming that Americans express a belief in, say,
equalitarianism, is not very insightful. Finding that beliefs in equalitarianism are related to various
individual characteristics and ideas probably confirms our expectations but does little else. But
detecting a persistent change in the level of some long-held value of Americans, and then trying to
decipher the causes and consequences of that change, could be an important contribution to our
understanding of American public opinion and culture.

There is, of course, no statistical test or other objective way of determining which values and
predispositions are important and long-term characteristics of American culture. This will have to
be judged by: a) inquiring of NES users; b) thoughtful consideration by the Board.

(4) Values and predispositions should be chosen with the expectation that they will be included in
NES studies for the foreseeable future. If one accepts point #3, then precisely because they are
long-term qualities of American culture, significant changes are not likely to occur in the short
run, suggesting the need for long-term inclusion.

(5) Immediate consideration (and possibly further testing) should be given to the response format
used with value questions.

(6) Consideration should be given to inclusion of one or more of the chosen values in
comparative studies—precisely because they are thought to be American values. It would be
useful to determine the degree to which such values are distinctively American.

V. HOW WE SHOULD RECORD WHAT WE DO

Values chosen for inclusion should be labelled as such in ICPSR codebooks. This may seem an
obvious point, but I call attention to the following listings:

EQUALITARIANISM:

1984 codebook (2nd ICPSR ed.): Not all asked together; listed under "R's Opinions on
Equal Opportunity" and "R's Position on Equal Opportunity" (along with additional items
under each heading).

1986 codebook (2nd ICPSR ed.): Listed under "R's Opinion on Equal Opportunity."

1988 codebook (2nd ICPSR ed.): Listed under "Equal Rights - R Agree/ Disagree."
1990 codebook (2nd ICPSR ed.): Listed under "Equal Opportunities/ Rights."

1992 codebook (enhanced 1st ICPSR ed.): Listed under "R Agrees/Disagrees" (along with additional items).

RACIAL PREJUDICE:

1990 codebook (2nd ICPSR ed.): Listed under "Civil Rights/Position of Blacks" (along with additional items).

1992 codebook (Enhanced ICPSR 1st ed.): Listed under "R Agrees/Disagrees."

MORAL TRADITIONALISM

1986 codebook (2nd ICPSR ed.): Listed under "R's Position on Morality in Society."

1990 codebook (2nd ICPSR ed.): Listed under "Moral Values."

1992 codebook (Enhanced, 1st ICPSR ed.): Listed under "R agrees/disagrees" (along with additional items).

GOVERNMENTAL INTERVENTIONISM

1990 codebook (2nd ICPSR ed.): Listed under "Individualism" (along with additional items).

1992 codebook (Enhanced ICPSR 1st ed.): Listed under "R's Opinion on Various Political Issues."

In addition, descriptions in the Variable List differ from year to year.

Finally, of these four values, only Equalitarianism is included in the Cumulative Data File, 1952-1992. (Governmental Interventionism was excluded because it had not yet been asked three or more times.)
Table 1

Summary of Inter-Item Correlations and Scale Reliabilities for Four NES Values/Predispositions

<table>
<thead>
<tr>
<th>Value/Predisposition</th>
<th>Year</th>
<th></th>
<th></th>
<th></th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equalitarianism</strong></td>
<td>.19</td>
<td>.25</td>
<td>.26</td>
<td>.22</td>
<td>.29</td>
</tr>
<tr>
<td>(six items)</td>
<td>.59</td>
<td>.66</td>
<td>.68</td>
<td>.62</td>
<td>.72</td>
</tr>
<tr>
<td><strong>Racial Prejudice</strong></td>
<td>.42</td>
<td>.41</td>
<td>.44</td>
<td>.43</td>
<td>.42</td>
</tr>
<tr>
<td>(four items)</td>
<td>.75</td>
<td>.74</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
</tr>
<tr>
<td><strong>Moral Traditionalism</strong></td>
<td>.30</td>
<td>.27</td>
<td>.26</td>
<td>.33</td>
<td>.29</td>
</tr>
<tr>
<td>(four items)</td>
<td>.63</td>
<td>.59</td>
<td>.58</td>
<td>.65</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Governmental Interventionism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td>(three items)</td>
<td>.70</td>
<td>.72</td>
<td>.71</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* Entries in the first row are the average inter-item correlations for all items in the scale. Entries in the second row are scale reliabilities (alpha).

(Source: NES Technical Report #50 (Shing-Yuan Sheng, 1/23/95).)*
Table 2

Average Inter-Item Correlations When Questions Are Worded in the Same or Opposing Directions

<table>
<thead>
<tr>
<th>Scale</th>
<th>Type of Response</th>
<th>Item Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Same</td>
</tr>
<tr>
<td>Equalitarianism</td>
<td>Agree/Disagree</td>
<td>.37</td>
</tr>
<tr>
<td>Racial Prejudice</td>
<td>Agree/Disagree</td>
<td>.52</td>
</tr>
<tr>
<td>Moral Traditionalism</td>
<td>Agree/Disagree</td>
<td>.42</td>
</tr>
<tr>
<td>Governmental Interventionism</td>
<td>Forced Choice</td>
<td>.44</td>
</tr>
</tbody>
</table>

Note: For the Governmental Interventionism scale, "same" refers to whether the first response was pro- (or anti-) interventionist in the two items.

Table 3

Percentage High on Scale When Questions Are Worded Positively versus Negatively

<table>
<thead>
<tr>
<th>Scale</th>
<th>Type of Response</th>
<th>High on Scale Indicated by Agree (1st response)</th>
<th>Disagree (2nd response)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equalitarianism</td>
<td>Agree/Disagree</td>
<td>70.5</td>
<td>41.9</td>
</tr>
<tr>
<td>Racial Prejudice</td>
<td>Agree/Disagree</td>
<td>62.8</td>
<td>41.5</td>
</tr>
<tr>
<td>Moral Traditionalism</td>
<td>Agree/Disagree</td>
<td>75.6</td>
<td>33.3</td>
</tr>
<tr>
<td>Governmental Interventionism</td>
<td>Forced Choice</td>
<td>71.7</td>
<td>63.1</td>
</tr>
</tbody>
</table>

*Note:* For Equalitarianism, Racial Prejudice, and Moral Traditionalism, high is the average percentage who agree (column 1) or disagree (column 2) "strongly" or "somewhat." For Governmental Interventionism, high is the average percentage selecting the pro-interventionist alternative.