

Evaluation of the Feeling Thermometer\*

A Report to the National Election Study Board  
Based on Data from the 1979 Pilot Survey

Herbert F. Weisberg  
Ohio State University

Arthur H. Miller  
University of Michigan

\*We thank Phil Kramer of Ohio State University for his assistance in conducting the data analysis.

The research was carried out as part of the NES developmental work in preparation for the 1980 election surveys.

## THE FEELING THERMOMETER

### I. BACKGROUND

The CPS election studies have had respondents rate possible presidential candidates on a thermometer measure since 1968. The pilot study included three methodological studies of the thermometer question.

First, the calibration of the ratings was tested. As a visual aid, respondents have always been shown a card (Figure 1) which labels nine specific scores (0, 15, 30, 40, 50, 60, 70, 85, and 100). For example, 70" is **labelled** "fairly warm or favorable feeling." While the intention of this labelling is to make the scale more understandable to respondents, the effect has been to convert a scale that was designed to be interval into an ordinal g-point rating scale. As a case in point, in the 1978 CPS study, 96% of the ratings of Carter, Ford, Kennedy, Reagan, and Brown were one of the nine **labelled** categories. Not only would it be desirable to make the scale more interval as originally desired, but making the scale less dependent on the labels would make the use of the thermometer in personal interviews more comparable to its use in telephone interviewing. Art Miller has already shown in a report to the Board of Overseers that thermometer responses in phone interviews (where a card cannot, of course, be shown to the respondent) are very different from those obtained in personal interviews with the same households.

To measure the effect of the thermometer card in the pilot study, a new thermometer card was drawn up by Art Miller and Herb Weisberg, with heavy consultation with Center for Political Studies personnel. The new card (Figure 2) gives verbal labels only to the middle 50" ("neither warm nor cold") and to the two anchors 0" ("very unfavorable feeling") and 100" ("very

Figure 1

"FEELING" THERMOMETER

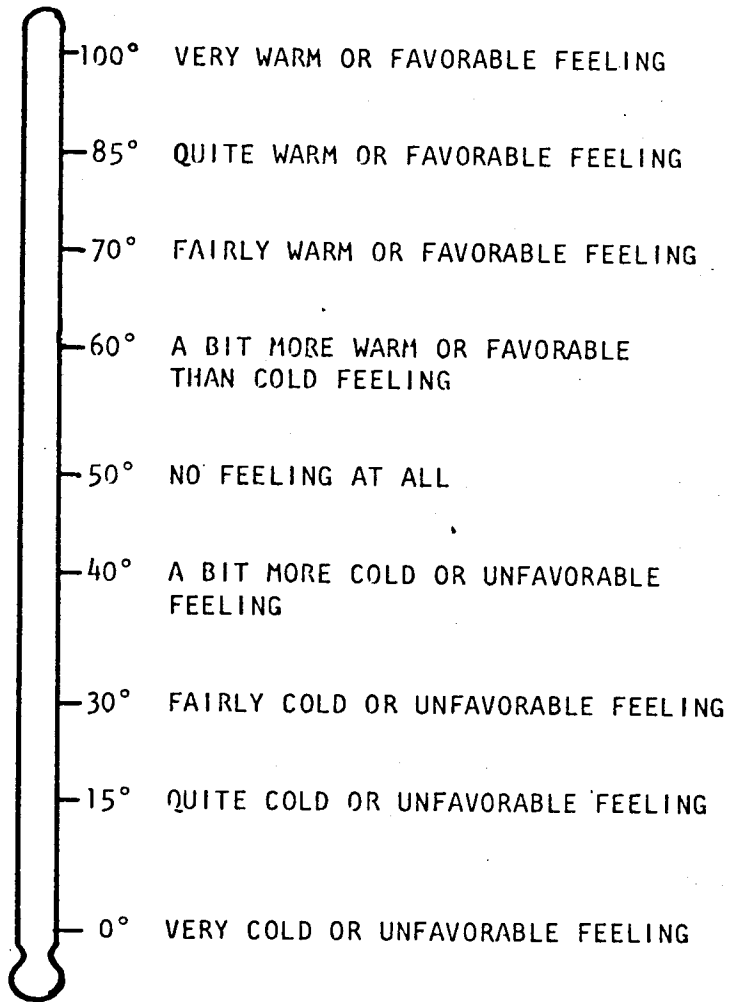
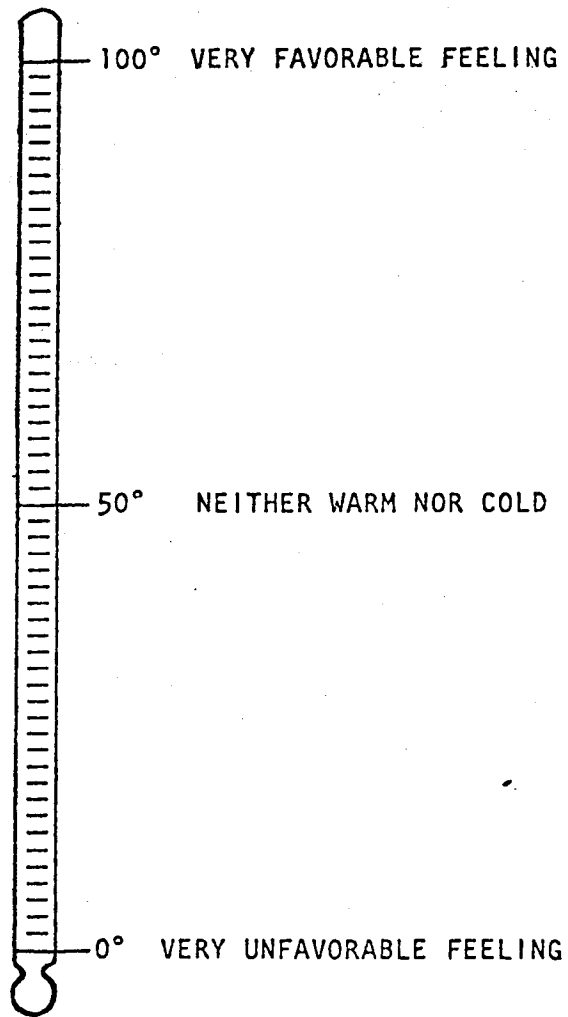


Figure 2

FEELING THERMOMETER



favorable feeling"). This labelling is essentially similar to that given in the explanation of the thermometer in phone interviewing. All Wave I respondents were asked to rate Carter, Ford, Kennedy, Reagan, and Brown on the new thermometer card, while form B respondents in Wave II were also asked to rate the same five candidates on the old thermometer card. The comparison is designed to show the effects of the thermometer card. If this experiment proves successful, the new card might replace the old card in future CPS studies, so that personal and phone interviews would be more comparable. This is the main part of the thermometer experiment in the pilot study, but before turning to the results the additional methodological studies of the thermometer question will be briefly described.

The second methodological study has to do with the existence of a positivity effect in the thermometer question. That is, do respondents give ratings to the candidates that are more favorable than their actual feelings? There is evidence of such a positivity effect when comparing thermometer ratings with open-ended like-dislike questions, though that discrepancy is often explained in terms of open-ended probes ("Anything else?") which make the respondents give more negative comments than they actually feel.

The candidate approval experiment in the pilot study was designed to measure this effect. All Wave II respondents were asked for which of the five candidates listed above they would be willing to vote. Thermometer ratings will be compared with the candidate approval to test for a positivity effect.

The third methodological study is a test of the use of thermometer ratings as preference measures. Many researchers have assumed that respondents who rate one candidate above a second on the thermometer prefer the first

candidate to the second. This assumption is critical in the many spatial analysis studies on the thermometer question, but the correspondence has never been directly confirmed.

The preference order experiment in the pilot study was designed to test this equivalence. Preference order data on the five candidates listed above were obtained for all Wave II respondents. Thermometer ratings will be compared to the preference orders to check the equivalence of the two procedures.

The remainder of this report will describe the results of these three methodological studies. The appendix will focus on the relative success rates of the thermometer question, the candidate approval question, and the preference order question in case there is interest in switching from the thermometer question to one of the other two formats in future CPS studies.

## II. THE THERMOMETER CARD EXPERIMENT

The thermometer card experiment was originally designed to compare ratings of the five candidates using the old and new thermometer card formats. However, there was a confounding effect in the experiment which requires the introduction of an additional control before any analysis results can be reported. The new card was used on all Wave I respondents, so when form B respondents used the old card in Wave II, they had experience just 3 weeks earlier in using the new card. That could have produced a learning effect that would confound the comparison of the two cards. Indeed, Wave II respondents were much more likely to use categories other than the nine labelled categories than were respondents in the 1978 CPS study (15% of the Wave II ratings were other than the nine labelled categories contrasted to

only 4% in the 1978 study). This strongly suggests a confounding effect. Fortunately, the 1978 CPS election study had respondents rate the same five candidates on the old thermometer card just a few months before the pilot study. (The interviews for the 1978 study were conducted from November 8, 1978 to January 31, 1979 while the pilot study first wave was March 5 to March 16 and the pilot study second wave was March 26 to April 13). Making the assumption that ratings of these candidates changes only minimally through this period, it seems safe to use the 1978 study as an additional control group in this report. In fact, the 1978 study also provides a further set of comparisons, discussed later, involving thermometer data obtained through telephone interviews with a subset of all the respondents interviewed.

Table 1 summarizes the scores given by respondents to the five candidates. There is an obvious tendency for respondents to use the labelled categories in their ratings, whether 9 or 3 categories are labelled. The new thermometer card does yield a scale which is more interval in its properties. One cost is an increase in the number of 50° ratings--from 17% in the 1978 study to 24% with the new card. Another indication of this tendency to use the labelled categories is that 9% of the respondents gave only scores of 0, 50, and 100 when using the new thermometer card in Wave I, while only 1% of the form B respondents only gave those scores when using the old card in Wave II.

The most dramatic change when using the new card is the drop in the proportion of responses using the other 6 labelled categories (15, 30, 40, 60, 70, and 85)--from 66% in the 1978 study to 23% with the new card. Overall, 96% of the ratings of the candidates in the 1978 study were one of the

Table 1  
Rating of the Five Candidates

	<u>Wave I New Card</u>	<u>Wave II Old Card</u>	<u>1978 Study Old Card</u>
50°	24.2%	19.7%	17.1%
0°, 100°	13.2%	7.0%	12.9%
0°, 50°, 100°	37.4%	26.7%	30.0%
15, 30, 40, 60, 70, 85	23.2%	58.4%	66.1%
0, 15, 30, 40, 50, 60, 70, 85, 100	60.5%	85.1%	96.1%

9 labelled categories, compared to only 60% of the ratings with the new thermometer card.

What intermediate ratings do respondents use when not given the suggestions on the old thermometer card? Nearly all ratings (96%) are still multiples of 5 (the exceptions were predominantly ratings of 1, 2, 48, 49, 51, 52, 98, and 99). Table 2 shows the distribution of ratings across scores that are multiples of 5. Note particularly the increased use of 75°, and also the increases for 25°, 80°, 90°, 45°, 65°, 55°, 10°, and 20°.

It was originally intended to use the panel aspect of the pilot study to determine the conversion rules between old and new thermometer scores. Unfortunately, there are not enough cases to explore these rules with any real confidence. Scores do tend to shift from adjacent categories, as one would hope. For example, the new popular 75" category tended to come from responses of 85, 60, and 70 with the old card. At least, the ratings on the old and new cards correlate highly for the form B respondents who rated candidates on both cards. The correlations between the old and new scales range from .61 for Ford to .73 for Carter, with an average correlation of .66.

A comparison of thermometer ratings obtained through personal and telephone interviews reveals differences similar to those observed for the old and new thermometer cards. Two data sets are available for making these comparisons. The first derives from a planned methodological experiment conducted by Robert Groves and Robert Kahn of the Survey Research Center during the spring of 1976. This study involved personal interviews with a representative sample of adults and a comparable set of telephone interviews with a sample of individuals living within the same PSU where the personal interviews were conducted. The thermometer was employed to measure ratings

Table 2

Ratings of the Five Candidates (all in percentages of all valid ratings)

<u>Rating</u>	<u>Wave I New Card</u>	<u>Wave II Old Card</u>	<u>1978 Study Old Card</u>
0	7.5	3.3	4.4
5	.2		
10	1.9	.6	.1
15	.5	3.7	3.0
20	1.8	.8	.3
25	4.0	.4	.1
30	3.8	4.1	4.5
35	.5		.1
40	4.6	8.1	7.2
45	2.8	1.2	.1
50	24.2	19.7	17.1
55	2.3	1.0	.1
60	8.4	16.4	17.0
65	2.7	1.9	.4
70	4.0	16.1	19.1
75	10.8	4.8	.8
80	4.0	1.5	1.0
85	1.9	10.1	15.3
90	3.5	1.9	.6
95	.9		.1
100	5.7	3.7	8.5
non-multiple of 5	4.1	.8	.2
.# of responses	(1274)	(517)	(10188)
Not recognize name	3.1%	2.1%	5.3%
Can't judge	5.9%	6.4%	5.7%

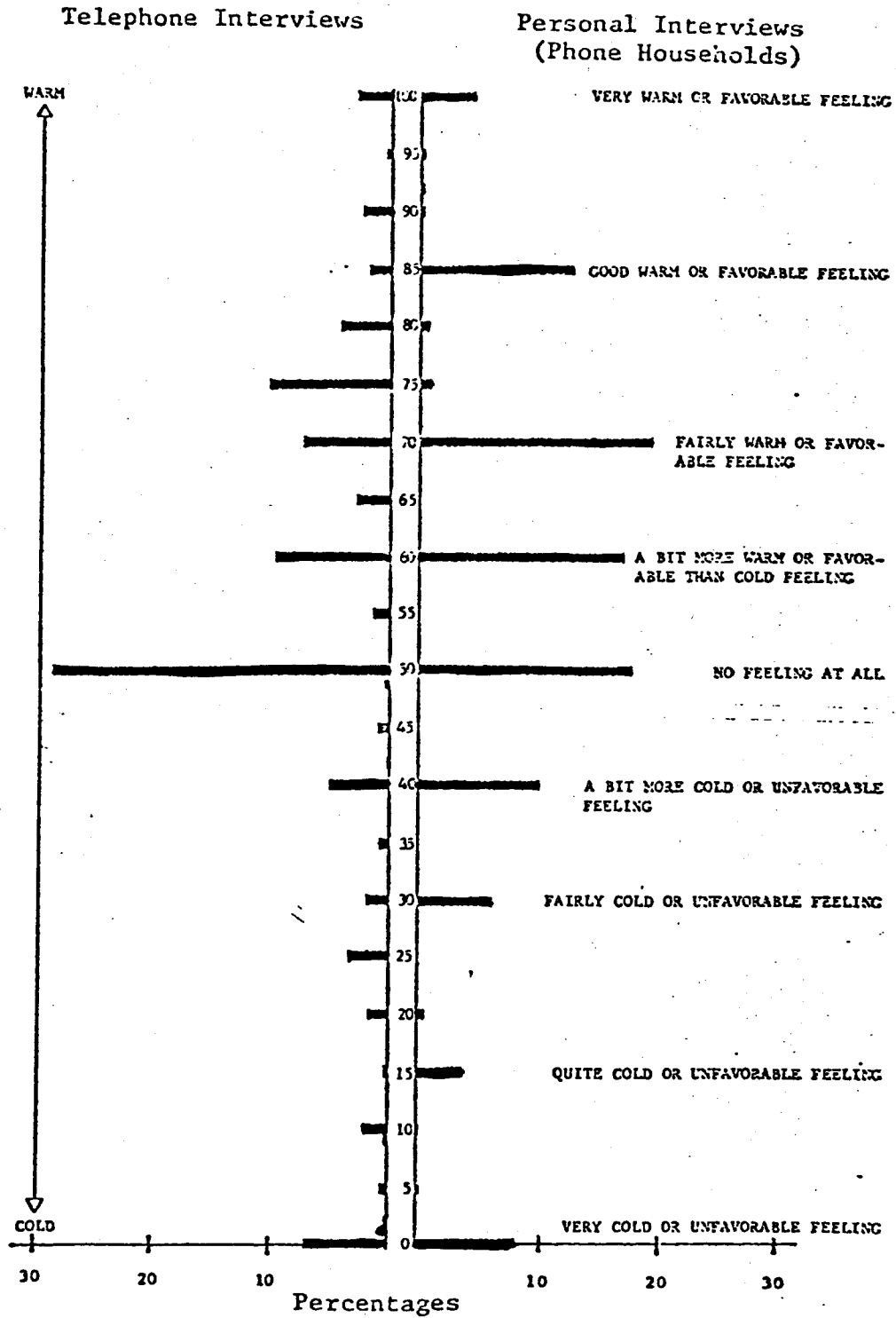
of then President Ford in both the personal and phone interviews. The old thermometer card was used in the personal interviews.

The second data set available for a comparison of personal and phone applications of the thermometer is the 1978 national election study. Because the "Democratic" and "Republican party" occurred in a somewhat obscure place among the list of various political leaders the respondent was asked to rate with the thermometer, several interviewers inadvertently skipped them during the interview. Some 227 respondents, thus, were not asked to rate the parties on the thermometer during the personal interview. After this oversight was discovered the study staff phoned the respondents and obtained ratings of the parties. The telephone subset of respondents do not differ significantly from the other respondents in terms of socio-economic or partisan characteristics, thus allowing direct comparison of the thermometer ratings obtained with the personal and telephone interviews.

The relevant data are summarized in Tables 2a and 2b for the 1978 comparison and Figure 3 presents the ratings of Ford from the SRC experiment. The major differences between the ratings from the phone and personal interviews correspond closely to those found with the old and new thermometer card. The phone interviews show a sharp increase in 50° ratings as a consequence of using labels for only the 0, 100 and 50 degree positions. Figure 3 reveals a 50% increase in 50° ratings of Ford in the phone interviews compared to the personal and there was a similar 30% increase in 50° ratings of the parties (see Table 2b). The phone interviews also reveal a much lower use of the other rating categories that are labeled on the show card (15, 30, 40, 60, 70 and 85 degrees)--from 57.6% in personal interview ratings of the parties to 26.5% on the phone. The labeled categories were

Figure 3

Histogram of Responses for Ford  
Feeling Thermometer Item by Mode  
of Interview<sup>a</sup>



<sup>a</sup>Data weighted by reciprocals of selection probabilities

Table 2a

Ratings of the Democratic and Republican Party Combined  
(percentage of all valid ratings), Comparing  
Personal and Phone Interviews

<u>Rating</u>	<u>Personal Interview</u>	<u>Telephone Interview</u>
0°	2.9%	3.0%
5	0.0	.2
10	.1	.2
15	1.6	.7
20	.1	.9
25	.1	2.1
30	3.3	1.6
35	.1	.9
40	7.9	7.1
45	.3	.9
50	29.3	37.9
55	.3	2.0
60	17.1	9.4
65	.2	2.5
70	16.7	5.7
75	.8	8.5
80	1.2	4.3
85	10.9	2.0
90	.5	3.4
95	.1	.2
100	6.2	4.6
Non-multiple of 5	0.1	1.8
# of responses	(3636)	(438)

Table 2b

1978 Election Study Comparison of Party Thermometers

	<u>Personal Interview</u>	<u>Telephone Interview</u>
50°	29.3%	37.9%
0°, 100°	9.1%	7.5%
0°, 50°, 100°	38.4%	45.4%
15°, 30°, 40°, 60°, 70°, 85°	57.6%	26.5%
0°, 15°, 30°, 40°, 50°, 60°, 70°, 85°, 100°	96.0%	71.9%
Below 50°	16.5%	17.6%
50°	29.3%	37.9%
Above 50°	54.2%	45.5%
Mean Ratings		
Democratic Party	63.8	60.6
Republican Party	56.5	53.8
Standard Deviation		
Democratic Party	20.8	21.4
Republican Party	20.2	19.6

used by 96% of the respondents among the personal interview subset when rating the parties as compared with only 71.9% of the phone interview respondents.

Perhaps the most noteworthy increase in category use occurred for ratings of 25 and 75 degrees (see Figure 3 and Table 2a). Apparently people tend to think more in terms of 25 and 75 than 15 and 85 when they are not confronted with constraining cues. This is particularly evident from the SRC experiment where the phone respondents had never seen the show card and would not be subject to any learning contamination. Nevertheless, the fact that the increased use of the unlabeled categories is so large for the 1978 respondents demonstrates the strong tendency to use these other numbers even after having seen the old card and been "trained" to think in terms of 15 and 85 degrees.

One further difference, which needs only brief mention here because it is discussed at length in the next section, is the slightly more positive ratings obtained in the personal interviews (see Table 2b). The differences are not great, but the phone interviews tend to produce mean ratings and distributions which are less positive than those obtained through personal interviews. As will be demonstrated below, this difference does not appear to be an artifact of phone versus personal interviews. Rather, it seems to be a reflection of the meaning conveyed by the labels used on the show card.

To sum up, the new thermometer card is successful in making the thermometer a more interval measure. The main drawback reported so far is the increase in neutral 50° ratings. The increase may only be by 7%, but having one-quarter of the ratings at 50° begins to seriously detract from the thermometer measure. One clear advantage to the new card, however, is

increased comparability between personal and phone interview results. Removing the discrepancy between phone and personal interview data will make for wider application of the measure and provide data that is comparable despite the method of interviewing. The fact that a large difference in ratings does exist for respondents interviewed in person and by phone suggests that CPS might consider including a variable designating which interview method was employed. Such a variable would allow users to sort respondents by interview technique and do further methodological work. Before reaching any final conclusions about the merits of the new thermometer card, an important result on positivity should be reported.

### III. POSITIVITY EFFECTS

An unexpected side-effect of the new thermometer card is to make the candidate ratings less positive than they were with the old card. This can be shown in several ways.

Table 3 summarizes the proportion of thermometer ratings that were above and below 50°. Comparing the 1978 study with the new card yields 17% less use of scores above 50° with the new card and 10% greater use of scores below 50°. Realize that these results are aggregated across five leading politicians, both Democrats and Republicans, so the results are not likely to be due to political events between the two surveys. While the Wave II responses may be confounded by the respondents previous experience with the old thermometer card, the double control they provide further reduces the possibility of this being a spurious result. Ratings of the politicians were less favorable in the Wave I interviews than they were in the 1978 study that was conducted about three weeks later. The most reasonable

Table 3

Positivity of Ratings of the Five Candidates

	<u>Wave I New Card</u>	<u>Wave II Old Card</u>	<u>1978 Study Old Card</u>
Below 50°	29.5%	22.2%	20.0%
50°	24.2%	19.7%	17.1%
Above 50°	46.3%	58.0%	63.0%

interpretation of this sequence is that the new thermometer card yields less positive ratings than does the old.

If this interpretation is correct, it should hold for all five candidates. Table 4 gives the mean ratings of the 5 candidates for each study. In all but one case, the candidate received his lowest mean rating with the new thermometer card. The one exception is Ted Kennedy whose average rating with the new card (55°) was slightly higher than his Wave II average rating (54°), though it was still much below his 1978 study average rating (62°). The mean ratings of the Republican and Democratic parties also reveal lower ratings with the new thermometer card. (As a technical note, the several changes reported here yield higher standard deviations with the new card by an average of about 10%, as shown in Table 5.)

The less positive ratings with the new thermometer card is also evident in the panel part of the pilot study. Table 6 compares the ratings of the five candidates by the Form B respondents. Overall, the ratings given the candidates were 15% less positive with the new card--10% more negative and 5% more neutral. The biggest shift in the internal cells is the change from positive with the old card to neutral with the new (11% of the ratings shift in this direction, compensated by only 4% in the opposite direction). There is also a sizeable shift from positive with the old card to negative with the new (7% of the ratings shift in this direction, compensated by only 1% in the opposite direction).

The original question about the positivity of the thermometer ratings was whether thermometer ratings were more positive than actual feelings toward the candidate. It is impossible to ever answer this question definitively, since there is no assurance of the greater validity of the other questions with which the thermometer ratings can be compared. However,

Table 4

Mean Ratings of Candidates and Parties

	<u>Wave I New Card</u>	<u>Wave II Old Card</u>	<u>1978 Study Old Card</u>
Carter	53.6	57.4	64.4
Ford	56.0	63.4	61.5
Kennedy	54.9	54.0	62.5
Reagan	52.8	57.6	57.3
Brown	51.3	56.1	57.2
Republican Party	53.3 (form B only)	not asked	55.2
Democratic Party	55.9 (form B only)	not asked	62.5

Table 5

Standard Deviations of Ratings of Five Candidates and Parties

	<u>Wave I New Card</u>	<u>Wave II Old Card</u>	<u>1978 Study Old Card</u>
Carter	27.3	25.5	22.3
Ford	22.6	19.8	22.5
Kennedy	29.0	23.7	27.4
Reagan	27.1	20.9	24.7
Brown	23.5	17.1	21.7
Republican Party	25.4 (form B only)	not asked	20.1
Democratic Party	25.1 (form B only)	not asked	20.8

Table 6

Changes in Ratings of Five Candidates Over Pilot Study Panel

<u>Wave I, New Card</u>	<u>Wave II, Old Card</u>			<u>Total</u>
	<u>Below 50°</u>	<u>50°</u>	<u>Above 50°</u>	
Below 50°	18%	7%	7%	33%
50°	4%	8%	11%	24%
Above 50°	1%	3%	39%	43%
Total	23%	19%	58%	100%

comparing thermometer ratings with the candidate approval question in the pilot study yields some important results.

Table 7 gives the proportions of people willing to vote for a candidate by their thermometer ratings of that candidate. Few people would vote for candidates they rate below 50°. The 50° category also turns out to be fairly negative, with few respondents being willing to vote for a candidate they rate 50° (and the 50° is particularly negative for the old thermometer card). There are too few cases in key categories to be sure exactly where the ratings really become positive. It appears that ratings of 60° with the old thermometer card are still negative while ratings of 55° and 60° with the new card are somewhat negative. Ratings above 70° with the old card and 65° with the new card are basically positive (with a two-thirds majority of the respondents being willing to vote for the candidate). Ratings of 85° and above with the old card and 80° and above with the new card are strongly positive. These results reveal a positivity effect of the thermometer--neutral and mildly positive ratings do not indicate support of the candidate. And the new thermometer card suffers from this problem somewhat less than the old does. Incidentally, the correlations between thermometer ratings and approvals are reasonably high. The Pearson  $r$  correlations range from .43 (Ford, old thermometer card, form B, Wave II) to .65 (Kennedy, old thermometer card, form B, Wave II), with a median correlation of .55.

By way of summary to this point, the responses to the old thermometer card are affected by the labelling on the card and suffer from a slight positivity effect. Responses to the new card are more interval and less positive, though there is still a concentration of responses around the

Table 7

Proportion Willing to Vote for Candidate by Thermometer Score

<u>Thermometer Scores</u>	<u>New Card</u>	<u>Old Card</u>
0-49	9% (295+)	4% (100+)
50	25% (265)	16% (102)
55-60	46% (114)	32% (85)
65	70% (30)	50% (10)
70-75	70% (145)	68% (108)
80-100	84% (161)	87% (89)

labelled 50° point and most respondents still do not support candidates they rate 50°. These results are supportive of a shift to the new thermometer card in future CPS studies. However, it is still important to examine the equivalence of thermometer ratings and preference orderings for candidates, as described in the next section.

#### IV. THERMOMETER RATINGS AND PREFERENCES

The correspondence between thermometer ratings and preference orders is quite high, though definitely not perfect. A few separate tests will be reported here.

First, respondents with a unique first choice on the thermometer question generally put that same candidate on top of their preference order, as shown by the data in Table 8. The correspondence is best for the old thermometer card, but the reason is likely that the old card was used in the same wave as the preference order question. The lower fit for the new card is likely to be due to mild changes in thermometer preferences between Wave I and Wave II. In any case, only 5% of the respondents put their top thermometer choice below second choice in their preference ordering.

Second, the correspondence between thermometer ratings and candidate approval is on par with the correspondence between preference orderings and candidate approval. Examining only the respondents who rated all five candidates on the thermometer and had a unique first choice, more than 95% were willing to vote for the candidate they scored highest on the thermometer. Using all respondents with a unique first choice on the thermometer regardless of the number of candidates they rated on the thermometer, more than 93% were willing to vote for the candidate they gave their highest thermometer rating. Table 9 gives the comparable statistics for the preference orders,

Table 8

Thermometer Rating by Preference Order Position

	<u>Wave I</u> <u>Full Sample</u>	<u>Wave II</u> <u>Form B Only</u>
% choosing their top thermometer choice as:		
First choice	77%	84%
Second choice	18%	11%
Third choice	5%	2%
Fourth choice	1%	2%
Fifth choice	0%	0%

Table 9

Proportion Approving a Candidate as a Function of  
Preference Order Position

	<u>% Willing to Vote for Candidate</u>
First choice	96.2%
Second choice	87.5%
Third choice	13.0%
Fourth choice	3.4%
Fifth choice	.5%

where 96% of the respondents were willing to vote for the candidate they liked most. Willingness to vote for first choice is just as high for thermometer data as for preference order data. (Table 9 also shows that most respondents would vote for their second choice candidate on the preference order, but few would vote for their third, fourth, or fifth choices.)

Third, the correspondence of ratings of candidate pairs on the thermometer card and on the preference orders is very high. For the 62 Form B, Wave II respondents who gave thermometer ratings to all five candidates and gave a full preference order on all five candidates in the same interview, 90% of the instances in which a higher thermometer score was given to one candidate than a second (tied thermometer ratings are omitted from this analysis) are cases in which that first candidate ranked higher in the person's preference ordering than did the second. This 90% figure should be contrasted to a 50% chance figure. This is not perfect equivalence, but it is reasonably high correspondence.

The point at which the preference ordering question differs most from the thermometer ratings has to do with ties in the orderings. It is easy for respondents to give the same thermometer score to more than one candidate. For the 62 respondents with full Form B data analyzed above, there were 85 ties among candidate pairs. A total of 620 ties would be possible in giving thermometer scores to 5 candidates, so this is a tie rate of 12%. However, the preference ordering question continually asks the respondent which of a set of candidates he/she likes more (or least), without making it easy to indicate a tied ranking. As a result, there are no ties in preference order rankings of these individuals. States in another way, only 14 Wave II, Form B respondents gave complete untied thermometer ratings

and full preference ordering for all five candidates, but another 48 were able to give a full preference ordering on the candidates even though they tied some in their thermometer scoring. (By contrast, noone gave an untied set of thermometer ratings but an incomplete or tied preference ordering.)

Overall, the common use of thermometer ratings to generate pseudo-preference orders seems to be valid. In particular, only 10% of the time are candidate pairs ranked in the opposite manner by the thermometer and preference order questions. However, the questions differ in their propensity to obtain incomplete orders. Ties are much common in thermometer scores than they are in preference orders on five candidate objects.

## V. SUMMARY OF FINDINGS

### A. The new thermometer card as compared to the old

1. greater number of 50° responses
2. less concentration of responses at 15, 30, 40, 60, 70, 85
3. less positive ratings than old
4. greater variance in ratings
5. correlates highly with old thermometer card (.6+)
6. more comparable to results obtained with telephone interviews

### B. Thermometers as compared to new measures

1. neutral and mildly positive candidate ratings generally do not indicate willingness to vote for a candidate (the "positivity effect")
2. same candidate generally rated highest on thermometer and at top of preference ordering
3. candidates ranked higher on thermometer ratings are generally ranked higher in preference ordering
4. ties are much more prevalent in thermometer scores than in preference data

## VI. PERSONAL RECOMMENDATIONS

### A. Thermometer card

We would support the adoption of the new card. There are costs, but we now know how severe they are. If the new card is adopted, we would strongly recommend doing something to decrease the

number of 50° responses. The instructions to interviewers could emphasize that responses other than 0, 50, and 100 degrees are legitimate and desired. Even better, the instructions to respondents should be modified by adding an extra sentence: "Remember you can use any score between 0 and 100 in rating the candidate."

B. Alternative measures

The relative utility of thermometers, preference orders, and approval voting questions is discussed in the appendix. Our feeling is that the thermometer question withstands the comparisons quite well and should be retained as the main measure. However, because of the problem of ties in thermometer scores and because of the common practice of using relative thermometer evaluation of candidates as a vote surrogate, we would suggest adding a brief question at the end of the thermometer series to obtain a clear first choice between the two presidential nominees: "All in all, do you prefer Carter or Ford as president?"

C. Future methodological experimentation

If there is future methodological experimentation, we would suggest examining the utility of the Stony Brook line-length measures. John Kessel has developed a shorter version of the respondent instructions for a Columbus survey, and his impression is that an even shorter version of the instructions would be successful. Now that we know a lot about thermometer ratings, we should see how they stack up against psychophysical measures.

Appendix

COMPARATIVE SUCCESS RATES FOR THERMOMETER, PREFERENCE, AND APPROVAL QUESTIONS

Although incidental to the main purpose of the use of these questions in the pilot study, an important point to consider is which set of questions best measures candidate attitudes. Stated more vividly, should the thermometer question be replaced by either the preference ordering question or the candidate approval question? An answer requires an examination of the comparative success rates for the three sets of questions.

First, the usual success of the thermometer ratings of candidates should be pointed out. As shown in Table 2, only about 9% of the time did respondents indicate either that they did not know a candidate or had no feelings about him. However, as pointed out in section IV of the report, tied responses are common on the thermometer.

The preference ordering was less successful (except for obtaining fewer ties). Nine percent of the respondents would not rank the candidates at all, mostly saying they didn't like any of the candidates (and in one instance saying that he/she liked all of the candidates equally). Note that these are completely missing cases for preference order analysis, whereas the 9% missing data for thermometer ratings generally meant that the respondent could rate all but one candidate (generally Brown).

A special validity problem arises with respect to respondents not knowing a candidate well enough to include him in their preference ordering. The preference order question does not give respondents an easy out when they don't know a candidate. Actually, the question asks who they like most, second, third, and least, which suggests that unknown candidates will amass a huge number of fourth place choices (when there are five candidates to

rank). Indeed Brown (the candidate with the most missing data on the thermometer) received a disproportionate number of fourth place choices on the thermometer--30% of the 4th place choices were Brown compared to a high of 21% for any other candidate. Another way of saying this is that 21 Wave II form B respondents gave thermometer ratings to all candidates but Brown and are recorded as having full preference orders, but fully two-thirds of these respondents (14) have Brown as their fourth-place choice. Obviously, preferences are confounded with missing data in this instance. If the preference ordering question is used in any future studies, a means should be devised to allow respondents to indicate when they do not know enough about a candidate to rank him.

Another problem arises in the coding of the preference order question. The question was coded in the easiest manner possible--as five variables, one for first choice (code values 1 For Brown to 5 for Reagan plus missing data codes), one for second choice, and so on. A more useful coding would be to have a variable for each candidate with the code being where the person puts that candidate in his/her preference order. For example, a person liking Kennedy second would have a 2 on the Kennedy variable. It is difficult to generate this type of coding from that available in the pilot study. The analysis given in this report is based on a very tedious hand compilation of responses that is possible only because of the small number of respondents in the pilot study. If the preference ordering question is used in any future studies, a means should be devised for generating the second type of coding as part of the dataset. (How missing data would be handled in this case is a perplexing issue, so perplexing that it argues against the use of direct preference ordering questions.)

By contrast to the preference order question, the candidate approval question was more successful. Seven percent of the respondents, however, did not approve subsets of the candidates, mostly because they would not vote for any of the candidates (an answer which is so intriguing that it raises fascinating questions about how common such feelings are during presidential elections) or, in a few cases, because they "don't know." Note again that this yields 7% missing data on all the candidates, whereas the 9% missing data for thermometer ratings generally means some valid data for each respondent.

A special validity problem arises with respect to respondents not knowing the candidate. Apparently respondents not knowing a candidate will be marked as not being willing to vote for the candidate (since the question is for which of the candidates the respondent would be willing to vote). (Respondents not willing to vote for any candidate were also counted as not willing to vote for each candidate separately for purposes of this report, so the missing data rate on the approval questions in this report is actually quite low.)

At least, respondents seem to have had little difficulty with the concept of indicating they would be willing to vote for multiple candidates. The distribution of the number of candidates respondents were willing to support is given in Table 10. On the average, respondents were willing to vote for 1.86 candidates. It is not obvious how this figure is affected by the choices made in constructing the survey--to use 5 candidates, 3 Democrats, and 2 Republicans.

By their very nature, approval data are inherently dichotomous. As a result, the approval materials are less rich for scaling purposes than

Table 10  
Number of Candidates Approved

Five	0%
Four	1
Three	9
Two	72
One	13
None	<u>6</u>
Total	100%

thermometer ratings or preference orders are. Switching from thermometer ratings to approval voting in the CPS study would therefore render spatial analysis a virtually crippling blow. That is not to say that the approval voting question would not be an interesting question in its own right, particularly if better means could be devised for handling missing data (such as assigning a candidate a missing data code on the approval measure if the respondent was unable to rate that candidate on the thermometer).

A reading of the relative advantages of these three types of questions is inherently subjective. All have disadvantages. The missing data problems seem more severe with the preference order and approval voting questions than with the thermometer question. Our choice would be to stick with the thermometer question, avoid the preference order question, and possibly experiment further with the approval voting question.