Kroh, Martin

Measuring Left-Right Political Orientation: The Choice of Response Format

Public Opinion Quarterly

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Kroh, Martin (2007) : Measuring Left-Right Political Orientation: The Choice of Response Format, Public Opinion Quarterly, Oxford University Press, Oxford, Vol. 71, Iss. 2, pp. 204-220,
https://doi.org/10.1093/poq/nfm009 ,
http://poq.oxfordjournals.org/content/71/2/204.abstract

This Version is available at:
https://hdl.handle.net/10419/74462

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in Public Opinion Quarterly following peer review. The definitive publisher-authenticated version "Measuring Left-Right Political Orientation / Martin Kroh In: Public Opinion Quarterly 71 (2007), 2, pp. 204-220" is available online at: http://dx.doi.org/10.1093/poq/nfm009.

Surveying the Left-Right Dimension:
The Choice of a Response Format

By:

Martin Kroh

DIW Berlin – German Institute for Economic Research

Socio-Economic Panel Study (SOEP)

Contact:

Martin Kroh

DIW Berlin

Königin-Luise-Straße 5

14195 Berlin, Germany

Phone: +49 30 89789 678

Fax : +49 30 89789 109

Email: mkroh@diw.de

Acknowledgements

I would like to thank Willem E. Saris and three anonymous reviewers for their valuable comments and suggestions on earlier drafts.

Abstract

Although left-right items are a standard tool of public opinion research, there is little agreement about the optimal response format. Two disputes can be identified in the literature: (a) whether to provide respondents with a small or large number of answer categories and (b) whether or not to administer the response scale including a midpoint. This study evaluates the performance of the 101, 11 and 10-point left-right scales, which directly speak to the two disputed aspects of measuring the left-right dimension. Drawing on data from a split ballot multi-trait multi-method experiment carried out in a methodological pretest to the German Socio-Economic Panel (*SOEP*), the analysis shows that the choice of a response format makes a difference in terms of data quality: The 11-point left-right scale produces the highest validity of left-right data closely followed by the 10-point scale. Moreover, an application from electoral research illustrates that the choice of response formats affects substantive interpretations about the nature of the left-right dimension. Since all three scales perform about equally well in terms of reliability and the ease of administration, the findings suggest that the 11-point left-right scale should be used in survey research.

1. Introduction

Social cognition enables us to orient ourselves within a complex social environment (e.g., Kunda 1999). In a political context, the left-right dimension represents a typical form of social cognition: When people reflect on politics, they tag themselves and others (people, groups, institutions, etc.) as 'left' or 'right'. Acting on this understanding, they derive policy views from their left-right position and vote for parties and candidates which they perceive as being close to their own left-right placement (Downs 1957; Butler and Stokes 1969; Inglehart and Klingemann 1976; Conover and Feldman 1981).[1] Many political systems share this, or functional equivalent, conceptualization of politics such as the differentiation between liberal and conservative.

For more than thirty years, left-right scales –the empirical operationalization of the left-right dimension– have been widely used survey instruments in pubic opinion research (e.g., Deutsch et al. 1966; Barnes 1971; Inglehart and Klingemann 1976). Survey data on the left-right dimension and its functional equivalences have proven to be especially valuable sources for studying party competition (e.g., Robertson 1976; Budge and Farlie 1977), coalition formation (e.g., Laver and Budge 1992), policy representation (e.g., Huber and Powell 1994; Thomassen and Schmitt 1997), opinion change (e.g., Smith 1990; Noelle-Neumann 1998; Knutsen 1998) as well as electoral research in general and spatial models of voting in particular (e.g., Stokes 1963; Enelow and Hinich 1984; Merrill and Grofman 1999). The notion that voters may reasonably use general heuristics instead of single political issues to orient themselves politically has further increased scholars' interest in the left-right dimension in recent years (e.g., Popkin 1991; Sniderman et al. 1991).

Although applied social science research often draws on left-right survey data, few studies examine the data quality of different instruments for surveying the left-right dimension. This is particularly surprising if one considers the variety of alternative instruments often used to measure it. This article seeks to fill this gap by rigorously testing for differences between typical left-right survey instruments. While the wording of the survey question is very similar in many studies,[2] there are substantial differences in the response format used. Many of these differences cluster around two disputed aspects of questionnaire design: the number of response categories and the (non)existence of a midpoint of scales. The number of answer categories with which the left-right dimension is surveyed ranges from three-point scales (Butler and Stokes 1969) to very detailed answer categories on a 101-point scale (Converse and Pierce 1973). The issue of whether to administer the left-right scale with or without a midpoint is reflected in the distinction between the two most frequently applied response scales: the 10- and the 11-point left-right scales. An inspection of codebooks collected by data repositories like ICPSR and the Central Archive at the University of Cologne reveals that the 10-point scale without a midpoint is used, for instance, in the Eurobarometer Studies, the European Election Studies, the World Value Surveys and the Dutch and other national election studies. The 11-point left-right scale, which provides a neutral point to respondents, is applied in the European Social Surveys and in national election studies such as the British or the Swedish ones. In some cross-national surveys, one even finds a mixture of scales provided by national collaborators. For example, 10 and 11-point scales are included in the Comparative Study of Electoral Systems and 7 and 10-point scales in the International Social Survey Programme. Finally, even within single countries, researchers will encounter polling institutes using different instruments for surveying the left-right dimension and its functional equivalences (e.g., Robinson and Fleishman 1988).

2. Hypotheses

The choice of particular response formats rests – at least implicitly – on assumptions about differences in data quality that are generated by these survey instruments (cf. Schumann and Presser 1981; Tourangeau et al. 2000; Presser et al. 2004). According to cognitive models of survey response, interviewees formulate their answers in a multi-stage process consisting of (a) the comprehension of the question, (b) the retrieval of relevant information from memory and context, (c) the judgment based on these sampled considerations and (d) the translation of the latent answer into the choice of a response category provided (e.g., Zaller 1992; Tourangeau et al. 2000). Measurement error may be a function of deficiencies at each stage of the process. Debates on different response formats measuring the left-right dimension are mostly associated with the last two stages of the cognitive process of forming a survey response. These debates concern the question of which response format permits respondents to map their latent opinions with regard to the left-right dimension most adequately; thus which response format can best minimize measurement error.

The Number of Scale Points

Ideally, survey formats permit respondents to convert their latent answers one-to-one into a response category provided. Respondents may experience difficulties in mapping latent answers on response formats when, for example, response categories are too broad or too specific. Measurement error becomes a function of the mismatch between the gradation of latent answers and the gradation of answer categories provided.

According to a common assumption of questionnaire design, a greater number of response options permits respondents to convey more information than few answer categories (for a review see Alwin 1997). If, for instance, respondents are provided with a three-point response scale (left, center, right), those with a moderate leaning to one side of the scale are unable to

express both the direction and strength of their latent answer at the same time and may randomly select the 'center' category, or one of the extreme categories, 'left' or 'right'.
Too many scale points may also reduce data quality: If, for instance, respondents' latent answers distinguish among five categories (left, moderate left, center, moderate right and right) but the provided response scale allows for a more detailed gradation, respondents may be uncertain about the difference in meaning between adjacent categories. Another drawback of many answer categories is that they increase the cognitive burden of respondents and thus their tendency to shortcut answers by accepting the first response category that fits more or less well (Alwin 1997; Krosnick and Fabrigar 1997).

Midpoints

The left-right dimension is bipolar with two opposing alternatives (completely left/right) and therefore with a theoretically defined midpoint (neither left/right, center). There has been a debate as to whether or not the midpoint of the scale should be represented by a response category, i.e. if the left-right scale should be administered with an unequal or equal number of scale points. Two opposing positions can be identified in the literature.
First, concerns have been voiced as to whether respondents who select the midpoint of a scale provide an accurate report of their latent opinion or whether they hide the absence of attitudes by not taking sides (Inglehart and Klingemann 1976; Schumann and Presser 1981, 162). Indicative of such a conjecture, Deutsch et al. (1966) and van der Eijk and Niemöller (1984) report that non-response increases if no neutral point is provided and that particularly respondents with low levels of political sophistication use the midpoint of the left-right scale.
The counter argument in favor of a middle category is that it provides an additional –and possibly crucial– gradation for mapping latent opinions. This may be of particular relevance in case of the left-right dimension: Single parties or candidates often portray themselves as being completely independent, open to voters and coalition partners on both sides of the left-

right spectrum. Hence, omitting the 'center' category in left-right scales may force respondents to use scale values for describing their own position or parties' positions that do not accurately reflect their latent opinion. Moreover, a midpoint may serve as an anchor for respondents' answers that may increase the derived data quality (Saris 1988).

3. Analysis and Data

The 11 and 10-point scales are used most frequently to measure the left-right dimension. Their key difference speaks to the issue of midpoints. The 11-point scale provides a neutral point, whereas the 10-point scale forces respondents to take sides on the left-right scale. Both scales are, however, limited in their number of answer categories. Among the traditional categorical scales easily administrable in survey research, the 101-point scale allows for the most detailed gradation of expressing opinions (Converse and Pierce 1973).[3] The choice of these three instruments is not meant to cover the most frequently used response scales of political research in general (which would probably involve 5 to 9 point scales) but it permits us to test for differences in quality of common left-right data with variation in the number of response options and the representation of midpoints. In the methodological pretest to the German Socio-Economic Panel Study (*SOEP*),[4] respondents were therefore asked to report their own views and the positions of all major German parties by placing them along these different left-right scales. The parties included the Christian democratic *CDU* and its Bavarian counterpart, the *CSU*; the social democratic *SPD*; the liberal party, the *FDP*; the environmentalist party, *B90/Die Grünen*; the former single party of East Germany, the *PDS*; and the far right party, the *Republikaner*.[5]

This study investigates two aspects of data quality: measurement error, i.e. validity and reliability of survey responses, and problems of data administration, i.e. non-response, the

elapsed time of interviews, respondents' willingness to provide answers and respondents' comprehension of their task.

For the estimation of measurement error, Saris et al. (2004) suggest a design that combines two of the classic approaches: an experimental design and statistical modeling. The multi-trait multi-method (MTMM) approach was suggested first by Campbell and Fiske in 1959 and has since then attracted much attention in survey research (for an overview, see Wothke 1996).[6] The basic idea of the MTMM approach is that by repeatedly observing single traits using different methods, the analyst can identify the amount of measurement error in different survey instruments. Figure 1 provides a simplified illustration of how data quality, i.e. validity and reliability, is defined in the MTMM context.

Suppose data are collected on respondents' position on the left-right dimension using a 101 and a 11-point scale. Respondents' observed answers (in bounded boxes) are a function of (a) the 'true score' given the response format and (b) measurement error. The share of variance in the observed data that is attributable to the variance of the underlying 'true score' defines the reliability of the measurement instrument.

<Figure 1>

The 'true score' of respondents' left-right position given a particular survey method is a function of (a) the underlying 'left-right factor', i.e. the latent answer, and (b) the method used, i.e. the 101 and the 11-point scales. Validity means the importance of the 'left-right factor' in the 'true scores'. The identification of validity and reliability parameters in the classical MTMM approach requires observations on at least three traits which have to be measured with three different methods. In other words, respondents would have to provide answers to the same set of three items (e.g., left-right position of political objects *a*, *b* and *c*) with some variation in the response format only (e.g., using a 101, 11 and 10 point scale).

The repeated surveying of the same items in the classical MTMM context means not only a burden for respondents but also bears the risk of memory and order effects. The combination of the MTMM approach with a split-ballot design reduces the number of necessary repetitions. Since each of the randomly drawn groups is presented with a different combination of two response formats as illustrated in Table 1, one requires only one instead of two repetitions of traits. Memory effects are therefore less likely to occur since a considerable time elapses between both observations of the same traits and the design allows to control for order effects by placing each method once at the beginning and once at the end of the interview. [7] Even though not all combinations of traits and methods are observed for all respondents, validity and reliability parameters can nonetheless be identified by normal theory maximum likelihood in multiple groups, assuming a common model, i.e. with equality constraints of all parameters across random groups (for a discussion of the estimation technique see Saris et al. 2004).

<Table 1>

4. Findings

Standardized parameters of Table 2 vary between 0 and 1.[8] The squared validity and method parameters denote the share of variance in the true scores attributable to the left-right factor and the method factor respectively. For instance, 42% (= $.65^2$) of the variance in the true left-right self-placements on a 101-point scale is due to the particular scale used and just 58% (= $.76^2$) is due to the latent left-right placement. Similarly, the squared reliability parameters indicate the share of variance in the observed left-right placements attributable to their true scores. For instance, 77% (= $.88^2$) of the variance in the observed left-right self-placements on a 101-point scale reflects true score variation; 21% of the variance reflects the unreliability of the measure.

<Table 2>

The entries of Table 2 indicate in line with previous research that left-right measures produce for the most part rather high data quality as compared to other attitude questions (e.g., Alwin and Krosnick 1991). This can be inferred from the validity and reliability parameters which are –apart from the validity parameters of the 101-point scale– above .85.

But are the reliability and validity estimates significantly different between response formats? Eyeballing the reliability parameters suggests only moderate differences between response formats. On average, reliability hovers around .90 for all scales. If differences in reliability estimates across scales are truly ignorable, than a model that sets these parameters to be equal across scales should perform about equally well to a model that allows these parameters to vary across scales. Table 3 reports the model fit in terms of χ^2 values and degrees of freedom of such a nested model relative to the basic model reported in Table 2 (Steiger et al. 1985). With respect to reliability, one cannot reject the null hypothesis that estimates are equal across all three scales (χ^2 difference of 18.26 and a difference in degrees of freedom of 16).

In terms of validity, differences between response formats are more pronounced: validity estimates reported in Table 2 hover around .97 for the 11-point scale, .93 for the 10-point scale and .76 for the 101-point scale. Table 3 suggests that constraining validity parameters to be the same across all three methods reduces the model fit significantly (χ^2 difference 258.12 and a difference in degrees of freedom of 18). Also each null hypothesis of equality in validity between two of the three response formats is rejected. Particularly equality constraints on validity estimates between the 101-point scale on the one hand and the 11 and 10-point scale on the other hand lead to a reduced model fit.

<Table 3>

The validity parameters suggest that left-right scales with many scale points, such as the 101-point scale tested here, may be too detailed for many respondents. Respondents' frequent choice of exposed values confirms this view: 86% of all respondents who report their own

position on the 101-point scale use integers which are multiplies of 10 (for problems of rounding, cf. Tourangeau et al. 2000, 232ff). The consequence of this behavior is a strong method effect of the 101-point scale that reduces the validity of responses on the left-right scale.[9]

The difference between the 10-point scale and the 11-point scale lies in the exclusion/inclusion of a midpoint. The analysis suggests that the 11-point scale generates somewhat more valid left-right data in mass surveys than the 10-point scale omitting a central response option. The absence of such a neutral point may force respondents to systematically deviate from their latent answer, which becomes particularly evident when the positioning of the *FDP* is considered. In the German party system, the Liberals are close to the political center, which is acknowledged by experts (Huber and Inglehart 1995) and the electorate (e.g., Dalton 2002, 202). While the mapping of the *FDP* on scales with a midpoint (101-point scale and 11-point scale) leads to an average or even weaker method effects for the *FDP* (101-point scale .64; 11-point scale .23) as compared to the method effect of the same scales for other parties (101-point scale around .64; 11-point scale around .25), the location of the Liberals on a scale without midpoint (10-point scale) leads to a higher method effect for *FDP* (.39) as compared to method effects of the same scale for other parties (around .36). The same is true for reliability estimates for *FDP* placements across response formats.

Besides the design of the study that allows to estimate the reliability and validity of different left-right scales, the *SOEP* pretest also provides four indicators of problems during the administration of interviews. A first indicator is the refusal to give left-right answers. The time necessary to conduct all eight questions on the left-right scale operates as a second indicator for the ease of administration. Immediately after the administration of the left-right scale, interviewers are asked to grade respondents' participation using a six-point school grading system. The third indicator of the ease of administration is interviewers' grade of

respondents' willingness to provide answers on the left-right scale and the fourth indicator is the interviewers' perception of respondents' comprehension of their task.

In contrast to Deutsch et al. (1966) and van der Eijk and Niemöller (1984), who find that non-response increases if no neutral point is provided, the *SOEP* pretest does not show that the 10-point scale significantly inflates non-response. Irrespective of the response format, about eight percent of all respondents fail to report at least one out of eight left-right items. Moreover, interviewers' ratings of respondents' participation both in terms of willingness to provide answers as well as comprehension of their task are generally positive. On scales ranging from 1 (very good participation) to 6 (very poor participation), the average score hovers around 2 for all response formats. Only in terms of the elapsed time of surveying left-right positions two significant mean differences can be observed across response formats: Conducting an interview with 10 and the 11-point left-right scales takes less time than interviews with 101-point left-right scales.

5. Consequences for Applied Research

Left-right scales perform differently well in terms of their data quality. As a practitioner in the social sciences one may nonetheless ask whether the choice of a response format makes a difference for applied research. To better understand the costs of different left-right scales, consider the following example: Respondents' vote intentions are regressed on their left-right self placements on different scales. Fitting the data to a multinomial logit model, Figure 2 reports the estimated probability of vote intention for the *SPD* as a function of respondents left-right self placements.[10] The curves for the 11 and 10-point scales are much steeper than the curve for the 101-point scale. Moreover, the maximum probability of voting for the Social Democrats is somewhat more to the center of the left-right dimension if the 11-point scale is used instead of the 101 and the 10-point scales.

<Figure 2>

Different response formats produce different predictions about the nature of respondents' left-right positioning (see also Van Doorn et al. 1984). One can of course not judge from Figure 2 which scale reveals the 'true' relationship between the left-right dimension and vote choice. However, the estimated differences in validity documented in Table 2 and 3 may be interpreted indicative of the presumption that the 10 and particularly the 11-point scale produce the more valid picture of the relationship between left-right self placement and vote choice.

6. Conclusions

The results reported in this article underscore the role of the left-right dimension for individuals' ability to orient themselves in the realm of politics: Almost all individuals interviewed are able to locate their own views as well as all relevant parties of the German party system on left-right scales. Very few interviewers report problems when surveying left-right placements. The estimated reliability and validity of respondents' left-right placements is high as compared to other survey items (e.g., Scherpenzeel and Saris 1997). This holds for the self-identification in left-right terms as well as the perception of party positions. This all indicates that most individuals are skilled at using their social cognition of the left-right dimension.

Although the left-right dimension is a standard tool in survey research, so far no agreement has been reached on the choice of a specific response format. As a consequence, applied research is faced with a mixture of different scale formats, particularly in cross-national data sets. In pooled analyses of these data, researchers have to assume equivalence of different scales. The analysis presented in this study suggests that such an assumption does not hold empirically: The most common response scales to the left-right dimension (101, 11 and 10-point scale) produce different data. To facilitate comparative research, differences between

response formats should be avoided in the future by agreeing on a single left-right survey instrument. All things considered, the findings of this study point to the conclusion that the 11-point scale should be used for surveying the left-right dimension. While all scales tested perform equally well in terms of reliability and ease of administration, the 11-point scale generates the highest validity of left-right data. The strong method effect of the 101-point scale affects all left-right placements and the moderate method effect of the 10-point scale appears to increase for political stimuli close to the center of the dimension. This does of course not imply that the 11-point scale is the optimal response format for measuring attitudes under all circumstances. Rather, the appropriateness of a survey measure is a function of characteristics of the concept under study, such as domain and complexity, characteristics of respondents, such as their familiarity with the concept, and characteristics of the interview setting, such as mode of data collection (e.g., Andrews 1984; Alwin and Krosnick 1991; Scherpenzeel and Saris 1997).

This article attributes the moderately inflated method effect in case of the 10-point scale to the lack of a midpoint. In case of the 101-point scale it is likely that ambiguity in the meaning of adjacent scale points reduces the validity of respondents' answers. Both interpretations may, however, be called into question.

The difference in data quality between the 10 and the 11-point scale may not necessarily be due to the issue of midpoints but to the difference in the total number of response options. Two arguments speak in favor of the first interpretation. As outlined before, the center of the left-right dimension not only has a distinct function in models of survey response (e.g., Schumann and Presser 1981) but also for political competition (Downs 1957) that provide plausible explanations for significant variation in data quality between the otherwise minor difference between 10 and 11 answer categories. Moreover, several scholars suggest a curvilinear relationship between the number of scale points and the quality of survey data

with a minimum of measurement error around 7 to 9 answer categories (e.g., Alwin and Krosnick 1991; for a related argument see Miller 1956). If this holds true, additional gradation of the 11-point scale as compared to the 10-point scale should *decrease* the quality of left-right data. Since the 11-point scale *increases* data quality as compared to the shorter 10-point scale, such a difference may not be attributed to the absolute number of scale points but to the inclusion of a midpoint.

The second issue in the interpretation of findings relates to the invalidity of the 101-point scale. Does this finding imply that people reflect upon the left-right cognition as a rough ordinal differentiation rather than a continuous dimension? Putting it in terms of cognitive models of survey response: Does the invalidity originate from the stages in whch respondents form latent judgments based on available information, or is it due to the last stage, where they translate these judgments into a response category on the 101-point scale? Previous research lends support for the latter hypothesis: Survey instruments drawing on numbers eventually reach the abilities of respondents' to express their opinions accurately (Presser et al. 2004). In other words, it is maybe respondents' inability to ascribe meaning to more than a certain number of integers that diminishes the data quality of very detailed response scales but not necessarily their inability to reflect upon the left-right cognition as a continuous dimension.

[1] The substantive meaning of the left-right dimension is traditionally associated with the socio-economic cleavages of equality (Lipset et al. 1954) and government intervention in the economy (Downs 1957). But the interpretation of the left-right semantic is variable across individuals (e.g., Klingemann 1979; Van der Eijk and Niemöller 1983, 225-247; Fuchs and Klingemann 1990), countries (e.g., Huber and Inglehart 1995; Knutsen 1995) and periods (Inglehart 1984; Fuchs and Klingemann 1990; Knutsen 1995). Some scholars interpret respondents' inability to uniformly describe the substantive meaning of the left-right or liberal-conservative semantic as indicative of the dimension's weakness as a cognitive structure – at least for the unsophisticated part of the population (e.g., Converse 1964; Luttberg and Gant 1985; Jacoby 1986). Others call this conclusion into question and maintain the notion of a left-right dimension as organizing element of the shared political consciousness of individuals in a given society: Even if voters disagree on what 'left' exactly means, they agree to a large extent, for instance, on classifying a socialist party as leftist (Laponce 1970; Klingemann 1972; Van der Eijk 2001; Kroh 2003). Moreover, panel data show high within-person stability in people's own left-right positioning over time, which suggests that the left-right cognition is part of individuals' political identity (e.g., Sears and Funk 1999; Zuckerman et al. 1998) and numerous studies demonstrate that people's left-right positioning uniformly affects all sorts of political attitudes and behaviors (e.g., Nie et al. 1976; Van der Eijk and Franklin 1996).

[2] Most versions of the measure contain wording like the following: "In politics people sometimes talk of 'left' and 'right'. Where would you place yourself [party *a*] on a scale from *n* to *m* where *n* means extreme left and *m* means extreme right?" Similarly the liberal-conservative scale is often surveyed by asking "We hear a lot of talk these days about 'liberals' and 'conservatives'. Here is a *m*-point-scale on which people's political views are arranged from extremely liberal to extremely conservative. Where would you place yourself [party *b*] on this scale?" Following common practice, all left-right scales analyzed in this article are labeled at their endpoints only.

[3] Magnitude scales that do permit continuous responses (cf. Lodge and Tursky 1981; Wegener 1982) are more difficult to administer than traditional categorical scales and often generate lower levels of data quality (for an overview see Krosnick and Fabrigar 1997). A magnitude estimation of the left-right dimension seems therefore a less practical alternative to the 101, 11 and 10-point scales in common day survey research.

[4] The population underlying the in-person household survey are persons above the age of 16 living in Germany. The response rate of the survey is 50% (response rate 1, see AAPOR standard definitions). All 772 realized interviews were conducted by way of computer-assisted personal interviewing (CAPI) in April and May 2004.

[5] The dominant parties of the German political arena are the *CDU/CSU* and the *SPD*. The *FDP*, *B90/Die Grünen* and the *PDS* each receive about five to ten percent of the votes in national parliamentary elections; the *Republikaner* never passed the electoral threshold in elections to the federal parliament (Bundestag).

[6] There is much debate on the choice of a formal model for the analysis of multi-trait multi-method matrices. Many parameterizations have been suggested in the literature, such as the correlated uniqueness model (Marsh 1989), the direct product model (Browne 1984) or the true score model applied here (Saris and Andrews 1991). A full discussion of the merits and caveats of these different models is beyond the scope of this study. Note that this study draws on the ´true score model` proposed by Saris and Andrews (1991) as it represents one of the accepted and frequently applied parameterizations.

[7] In the empirical data analyzed in this article, traits were repeated on average 50 minutes after the first round of left-right placements. Van Meurs and Saris (1990) show that 20 minutes are sufficient to obtain independent measures.

[8] Since the scales investigated differ in terms of the number of response categories, polychoric correlations are estimated, which are suited for data with different levels of measurement (Olsson 1979).

[9] Collapsing the 101-point scale into an 11-point scale does produce very similar estimates of the data quality to the original 101-point scale. For lack of space, this analysis is not reported in form of a table.

[10] For the lack of space, estimates of the multinomial logit model are not reported in form of a table. Note, however, that differences in the effect magnitude are significant between 101-point scale on the one hand and the 11 and 10-point scales on the other hand the at $p < .05$ but not between the 11 and 10 point scales.

6. References

Alwin, Duane F. 1997. Feeling Thermometers Versus 7-Point Scales. Which are Better? Sociological Methods and Research 25: 318-340.

Alwin, Duane F., and Jon A. Krosnick. 1991. The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes. *Sociological Methods and Research* 20: 139-181.

Andrews, Frank. 1984. Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach. *Public Opinion Quarterly* 48: 409-442.

Barnes, Samuel H. 1971. Left-Right and the Italian Voter. *Comparative Political Studies* 4: 157-175.

Browne, Michael W. 1984. The Decomposition of Multitrait-Multimethod Matrices. *British Journal of Mathematical and Statistical Psychology* 37: 1-21.

Budge, Ian, and Dennis Farlie. 1977. *Voting and Party Competition*. New York: Wiley.

Butler, David, and Donald E. Stokes. 1969. *Political Change in Britain*. New York: St. Martin's.

Campbell, Donald T., and Donald W. Fiske. 1959. Convergent and Discriminant Validation by the Multitrait Multimethod Matrices. *Psychological Bulletin* 56: 81-105.

Conover, Pamela Johnston, and Stanley Feldman. 1981. The Origins and Meaning of Liberal-Conservative Self-Identifications. *American Journal of Political Science* 25:617-645.

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Public. In *Ideology and Discontent* edited by David E. Apter. New York: Free Press, 206-261.

Converse, Philip E., and Roy Pierce. 1973. Basic Cleavages in French Politics and the Disorders of May and June, 1968. In *Political Behavior in Western Societies* edited by Richard Rose. New York: Wiley.

Dalton, Russell J. 2002. *Citizen Politics. Public Opinion and Political Parties in Advanced Industrial Democracies*. 3rd Edition. New York: Chatham.

Deutsch, Emeric, Denis Lindon, and Pierre Weill. 1966. *Les Familles Politiques*. Paris : Les Editions de Minuit.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.

Enelow, James, and Melvin Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. New York: Cambridge University Press.

Fuchs, Dieter, and Hans-Dieter Klingemann. 1990. The Left-Right Schema. In *Continuities in Political Action: A Longitudinal Study of Political Orientations in Three Western Democracies* edited by M. Kent Jennings, Jan van Deth, Samuel H. Barnes, Felix J. Heunks, Ronald Inglehart, Max Kaase, Hans-Dieter Klingemann and Jacques J. Thomassen. Berlin: deGruyter, 203-234.

Huber, John D., and Ronald Inglehart. 1995. Expert Interpretations of Party Space and Party Locations in Forty-Two Societies. *Party Politics* 1: 73-111.

Huber, John D., and G. Bingham Powell. Congruence between Citizens and Policy-Makers in 2 Versions of Liberal Democracy. *World Politics* 46: 291-326.

Inglehart, Ronald. 1984. The Changing Structure of Political Cleavages in Western Society. In *Electoral Change in Advanced Industrial Democracies: Realignment or Dealignment?* edited by Russel J. Dalton, Scott C. Flanagan, and Paul A. Beck. Princeton: Princeton University Press, 25-69.

Inglehart, Ronald, and Hans-Dieter Klingemann. 1976. Party Identification, Ideological Preference and the Left-Right Dimension among Western Mass Publics. In *Party Identification and Beyond: Representations of Voting and Party Competition* edited by Ian Budge, Ivor Crewe, and Dennis Farlie. London: Wiley, 243-273.

Jacoby, William G. 1986. Levels of Conceptualization and Reliance on the Liberal-Conservative Continuum. *Journal of Politics* 48: 423-431.

Klingemann, Hans-Dieter. 1972. Testing the Left-Right Continuum on a Sample of German Voters. *Comparative Political Studies* 5: 93-106.

Klingemann, Hans-Dieter. 1979. Measuring Ideological Conceptualisations. In *Political Action: Mass Participation in Five Western Democracies* edited by Samuel H. Barnes, and Max Kaase. Beverly Hills: Sage, 215-254.

Knutsen, Oddbjørn. 1995. Value Orientations, Political Conflicts and Left-Right Identifications: A Comparative Study. *European Journal of Political Science* 28: 63-93.

Knutsen, Oddbjørn. 1998. Europeans Move Towards the Center: A Comparative Longitudinal Study of Left-Right Self-Placement in Europe. *International Journal of Public Opinion Research* 10: 292-316.

Kroh, Martin. 2003. *Parties, Politicians, and Policies: Orientations of Vote Choice Across Voters and Contexts*. Ph.D. diss., University of Amsterdam.

Krosnick, Jon A., and Leandre R. Fabrigar. 1997. Designing Rating Scales for Effective Measurement in Surveys. In *Survey Measurement and Process Quality* edited by Lars E. Lyberg, Paul Biemer, Martin Collins, Edith De Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin. New York: Wiley, 141-164.

Kunda, Ziva. 1999. *Social Cognition: Making Sense of People*. Boston: MIT Press.

Laponce, Jean A. 1970. Note on the use of the left-right dimension. *Comparative Political Studies* 3: 481-502.

Laver, Michael J., and Ian Budge, eds. 1992. *Party, Policy, and Government Coalitions*. London: Oxford University Press.

Lipset, Seymour M., Paul F. Lazarsfeld, Allen H. Barton, and Juan Linz. 1954. The Psychology of Voting: An Analysis of Political Behavior. In *Handbook of Social Psychology (Vol. 2)* edited by Gardner Lindzey. Cambridge: Addison-Wesley, 1124-1175.

Lodge, Milton, and Bernard Tursky. 1981. On the Magnitude Scaling of Political Opinion in Survey Research. *American Journal of Political Science* 25: 376-419.

Luttberg, Norman R., and Michael M. Gant. 1985. The Failure of Liberal/Conservative Ideology as a Cognitive Structure. *Public Opinion Quarterly* 49: 80-93.

Marsh Herbert W. 1989. Confirmatory Factor Analysis of Multitrait-Multimethod Data: Many Problems and Few Solutions. *Applied Psychological Measurement* 13: 335-361 .

Merrill, Samuel III, and Bernard Grofman. 1999. *A Unified Theory of Voting: Directional and Proximity Spatial Models*. Cambridge: Cambridge University Press.

Miller, George A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review* 63: 81-97.

Nie, Norman H., Verba, Sidney, and Petrocik, John R. 1976. *The Changing American Voter*. Cambridge: Harvard University Press.

Noelle-Neumann, Elisabeth. 1998. A Shift from the Right to the Left as an Indicator for Value Change: A Battle for the Climate of Opinion. *International Journal of Public Opinion Research* 10: 317-334.

Olsson, Ulf. 1979. Maximum Likelihood Estimation of the Polychoric Correlation Coefficient. *Psychometrika* 44: 443-460.

Popkin, Samuel L. 1991. *The Reasoning Voter: Communication and Persuasion in Presidential Elections*. Chicago: University of Chicago Press.

Presser, Stanley, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer. 2004. *Methods for Testing and Evaluating Survey Questions*. New York: Wiley.

Robertson, David. 1976. *A Theory of Party Competition*. New York: Wiley.

Robinson, John P., and John A. Fleishman. 1988. A Report: Ideological identification: Trends and Interpretations of the Liberal-Conservative Balance. *Public Opinion Quarterly* 52: 134-145.

Saris, Willem E., ed. 1988. *Variation in Response Functions: A Source of Measurement Error in Attitude Research*. Amsterdam: Sociometric Research Foundation.

Saris, Willem E., and Frank M. Andrews. 1991. Evaluation of Measurement Instruments Using a Structural Modeling Approach. In *Measurement Errors in Surveys* edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. New York: Wiley, 575-599.

Saris, Willem E., Albert Satorra, and Germà Coenders. 2004. A New Approach to Evaluating the Quality of Measurement Instruments: The Split-ballot MTMM Design. *Sociological Methodology* 34: 311-347.

Scherpenzeel, Annette, and Willem E. Saris. 1997. The Validity and Reliability of Survey Questions: A Meta-Analysis of MTMM studies. *Sociological Methods and Research* 25: 341-383.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Order and Context*. New York: Academic Press.

Sears, David O., and Carolyn L. Funk. 1999. Evidence of the Long-term Persistence of Adults' Political Predisposition. *Journal of Politics* 61: 1-28.

Smith, Tom W. 1990. Liberal and Conservative Trends in the United States Since World War II. *Public Opinion Quarterly* 54: 479-507.

Sniderman, Paul M., Richard Brody, and Philip Tetlock. 1991. *Reasoning and Choice: Explorations in Political Psychology*. Cambridge: Cambridge University Press.

Steiger, James H., Shapiro, Alexander, and Browne, Michael W. 1985. On Multivariate Asymptotic Distribution of Sequential Chi-square Statistics. *Psychometrika* 50: 253-264.

Stokes, Donald E. 1963. Spatial Models of Party Competition. *American Political Science Review* 17: 368-377.

Thomassen, Jacques J., and Hermann Schmitt. 1997. Policy Representation. *European Journal of Political Research* 32: 165–184.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Van der Eijk, Cees. 2001. Measuring Agreement in Ordered Rating Scales. *Quality and Quantity* 35: 325-341.

Van der Eijk, Cees, and Mark N. Franklin, eds. 1996. *Choosing Europe? The European Electorate and National Politics in Face of a Union*. Ann Arbor: University of Michigan Press.

Van der Eijk, Cees, and Broer Niemöller. 1983. *Electoral Change in the Netherlands: Empirical Results and Methods of Measurement*. Amsterdam: CT Press.

Van der Eijk, Cees, and Broer Niemöller. 1984. *Theoretical and Methodological Considerations in the Use of Left-Right Scales*. Paper presented to the ECPR Joint Sessions, Salzburg.

Van Doorn, L. Jonny, Willem E. Saris, and H. Komen. 1984. Verschillende procedures om partij-identificatie te meten en hun consequenties. *Acta Politica* 2: 211-238.

Van Meurs, A., and Willem E. Saris. 1990. Memory Effects in MTMM Studies. In *Evaluation of Measurement Instruments by Meta-Analysis of Multitrait-Multimethod Studies* edited by Willem E. Saris, and A. van Meurs. Amsterdam: North Holland, 134-146.

Wegener, Bernd, ed. 1982. *Social Attitudes and Psychophysical Measurement*. Hillsdale: Erlbaum.

Wothke, Werner. 1996. Models for Multitrait-Multimethod Matrix Analysis. In *Advanced Structural Equation Modeling: Issues and Techniques* edited by George C. Marcoulides, and Randall E. Schumacker. Mahwah: Lawrence Erlbaum, 7-56.

Zaller, John R. 1992. *The Nature and Origin of Mass Opinion*. Cambridge: Cambridge University Press.

Zuckerman, Alan S., Laurence A. Kotler-Berkowitz, and Lucas A. Swaine. 1998. Anchoring Political Preferences: The Structural Bases of Stable Electoral Decisions and Political Attitudes in Britain. *European Journal of Political Research* 33: 285-321.

List of Figures

Table 1 Design of the 3-Group Split-Ballot MTMM Experiment.

	Beginning of Interview	End of Interview
Group 1	101-Point Scale	11-Point Scale
Group 2	11-Point Scale	10-Point Scale
Group 3	10-Point Scale	101-Point Scale

Table 2 Split Ballot Multi-Trait Multi-Method Experiment of Left-Right Placements.

		Validity								Method Effect			Reliability
		Ego	CDU	SPD	FDP	Green	CSU	PDS	Reps	101-point scale	11-point scale	10-point scale	
Ego	101	0.76								0.65			0.88
CDU	101		0.75							0.67			0.91
SPD	101			0.73						0.68			0.82
FDP	101				0.77					0.64			0.92
Green	101					0.76				0.65			0.89
CSU	101						0.76			0.65			0.93
PDS	101							0.78		0.63			0.85
Reps	101								0.80	0.60			0.92
Ego	11	0.97									0.25		0.85
CDU	11		0.97								0.25		0.85
SPD	11			0.96							0.29		0.73
FDP	11				0.97						0.23		0.94
Green	11					0.97					0.25		0.86
CSU	11						0.97				0.24		0.88
PDS	11							0.97			0.26		0.83
Reps	11								0.97		0.26		0.82
Ego	10	0.94										0.35	0.90
CDU	10		0.93									0.36	0.89
SPD	10			0.93								0.36	0.88
FDP	10				0.92							0.39	0.82
Green	10					0.93						0.37	0.85
CSU	10						0.94					0.35	0.92
PDS	10							0.93				0.36	0.88
Reps	10								0.93			0.37	0.86

Note. All estimates are standardized and significant at $p < 0.01$. $\chi^2 = 1054.84$; df = 821.
Data Source. SOEP-Pretest 2004.

Table 3 Chi-Square Difference Tests Between Nested MTMM-Models (with Equality Constraints) and the Basic Model (without Equality Constraints).

	Model Fit		Difference in Model Fit		
	χ^2	df	χ^2	df	p
Basic Model[†]	1054.84	821	-	-	-
Equality-Constraints on Reliability Estimates					
Between 101, 11 & 10-Point Scale	1073.10	837	18.26	16	
Between 101 and 11-Point Scale	1045.08	829	9.76	8	
Between 101 and 10-Point Scale	1089.93	829	35.09	8	***
Between 11 and 10-Point Scale	1058.81	829	3.97	8	
Equality-Constraints on Validity Estimates					
Between 101, 11 & 10-Point Scale	1312.96	839	258.12	18	***
Between 101 and 11-Point Scale	1274.60	830	219.76	9	***
Between 101 and 10-Point Scale	1185.77	830	130.93	9	***
Between 11 and 10-Point Scale	1070.63	830	15.79	9	*

Note. [†] Model allows for variation in reliability and validity estimates across all methods. See Table 2 for respective parameter estimates. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. *Data Source.* SOEP-Pretest 2004.

Figure 1 Conceptual Diagram of MTMM Experiment.

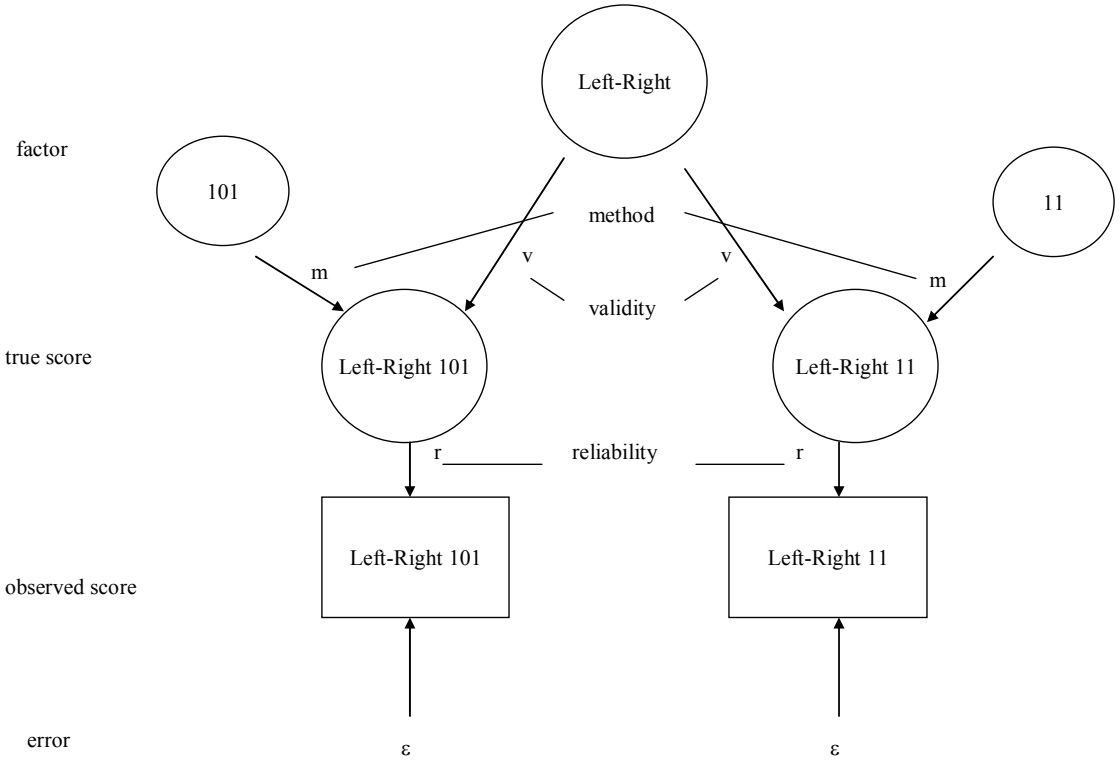
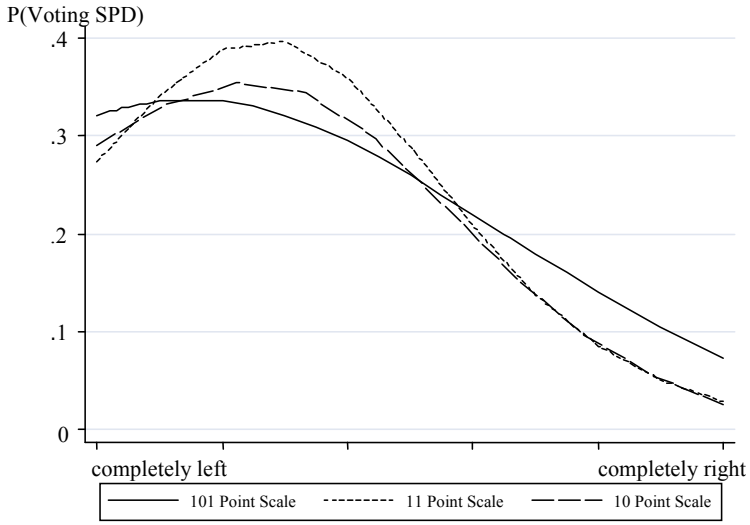


Figure 2 The Predicted Probability of Voting SPD as a Function of Respondents' Left-Right Position.



Note. The graph displays post estimation results of a multinomial regression model. *Data Source*. SOEP-Pretest 2004.