



The University of
Nottingham

UNITED KINGDOM • CHINA • MALAYSIA

van der Eijk, Cees and Marsh, Michael (2007) Don't expect me to vote for you just because I like you, even if you do make me feel warm inside: a comparison of the validity of non-ipsative measures of party support. In: Annual Meeting of the American Political Science Association, Chicago, 29 August - 1 September 2007, 29 August - 1 September, Chicago. (Unpublished)

Access from the University of Nottingham repository:

http://eprints.nottingham.ac.uk/1251/1/Cvde-MM_APSA_paper_-_F4.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see:
http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

**Don't expect me to vote for you just because I like you,
even if you do make me feel warm inside**

—

**A comparison of the validity of non-ipsative
measures of party support**

by

Cees van der Eijk (University of Nottingham, UK)

and

Michael Marsh (Trinity College, Dublin, Ireland)

Paper presented at Annual Meeting of the American Political Science Association,
Chicago, 29 August – 1 September 2007

Panel: Methodological Advances in the Study of Elections and Voting Behavior

Division 8 – Political Methodology , Division 36 – Elections and Voting Behavior

Email addresses of the authors:

cees.vandereijk@nottingham.ac.uk

mmarsh@tcd.ie

Abstract

The inclusion of non-ipsative measures of party preference (in essence ratings for each of the parties of a political system) has become established practice in mass surveys conducted for election studies. They exist in different forms, known as thermometer ratings or feeling scores, likes and dislikes scores, or support propensities. Usually only one of these is included in a single survey, which makes it difficult to assess the relative merits of each. The questionnaire of the Irish National Election Study 2002 (INES2002) contained three different batteries of non-ipsative party preferences. This paper investigates some of the properties of these different indicators. We focus in particular on two phenomena. First, the relationship between non-ipsative preferences and the choices actually made on the ballot. In Ireland this relationship is more revealing than in most other countries owing to the electoral system (STV) which allows voters to cast multiple ordered votes for candidates from different parties. Second, we investigate the latent structure of each of the batteries of party preferences and the relationships between them.

We conclude that the three instruments are not interchangeable, that they measure different orientations, and that one –the propensity to vote for a party– is by far preferable if the purpose of the study is the explanation of voters’ actual choice behaviour. This finding has important ramifications for the design of election study questionnaires.

Authors:

Cees van der Eijk, Professor of Social Science Research Methods
and Director of Social Sciences Methods and Data Institute
University of Nottingham, Nottingham NG7 2RD, UK
Email: cees.vandereijk@nottingham.ac.uk

Michael Marsh, Professor of Comparative Political Behaviour
and head of the School of Social Sciences and Philosophy, Trinity College Dublin
http://www.tcd.ie/Political_Science/Staff/Michael.Marsh
email: mmarsh@tcd.ie

Introduction and theoretical framework

Election study surveys commonly include some kind of question about the respondent's preference for each of the various parties in a country. One of the best known examples of such a question is the thermometer scale pioneered in the American National Election Studies –and subsequently used in election studies in many other countries. The respondent is invited to indicate for a set of political parties, and often also for sets of other stimuli such as candidates or social groups, how favourable or unfavourable their feelings are, and to express this on a scale from 0 to 100 where 0 indicates unfavourable or 'cold' feelings and 100 favourable or 'warm' ones. Similar questions have emerged in other surveys, apparently trying to elicit the same kind of preferences by way of somewhat different phrasing of the question, or by different response formats. In the absence of comparative studies of the merits of different forms of measuring such non-ipsative preferences,¹ the choice of which one to use in a particular study seems to be largely a matter of personal taste or happenstance. In some instances, choice of instrument is even dictated by external 'authorities', as in the case in the CSES project that imposes a particular format on its participants.² Yet, there too, no firm evidence about properties and performance of various instruments underlies that decree. This is unfortunate, as numerous studies demonstrate that variations in the wording and response formats of survey questions can have considerable effect on characteristics of the resulting measurements in terms of ease of response, bias, reliability and validity. Not every variation on a common theme is equally 'good' as a measurement instrument. This paper investigates the 'performance' of three different formats in which non-ipsative party preferences are asked in election studies, in the hope of making choice of measurement instruments less a matter of taste and more a matter of evidence.

What makes a measurement instrument a 'good' one, or a 'better' one when compared to a different instrument purportedly measuring the same phenomenon? One can consider in this respect two kinds of criteria: technical and substantive. 'Technical' criteria to evaluate the merits of survey questions relate to properties such as

- Coverage: the extent to which respondents in a representative sample are willing and able to give valid responses.

¹ Preferences are called ipsative if they are constrained to a fixed-sum rule. In most countries, electoral choice is ipsative, as the voter is only allowed to express a single preference on the ballot –voting for one party thus implies not-voting for each of the other options. Ranked preferences are ipsative in a different way: having indicated which of a series of stimuli is preferred most necessitates that other stimuli are ranked lower, etc. In all these cases, the preferences for the various stimuli or choice options are thus not mutually independent, because of their fixed-sum constraint. In the case of non-ipsative preferences, a respondent expresses a degree of preference for each of the options, where the response given for one of the options does not in any way constrain the responses allowed for other options. Most often, non-ipsative preferences are elicited in the form of preference ratings for the various options.

² The CSES –Comparative Study of Electoral Systems– is an invaluable and ambitious attempt to increase comparability of elections by insertion of a standardized block of questions in the election studies of all participating countries. The project is currently running its 3rd module. In modules 1 and 2 the number of participating countries was 34 and 38 respectively, while 56 countries have indicated their intention to participate in round 3. For more details about the project and the data that it has generated, see <http://www.cses.org>. The particular form of non-ipsative party preferences imposed by the CSES in Module 2 asks respondents to rate parties on a scale from 0 ('strongly dislike') to 10 ('strongly like').

- Discrimination: the extent to which the distribution of responses allows the comparison of different groups of sufficient size. Heavily skewed response distributions thus permit little discrimination, irrespective of the number of response options, while uniform or even normal distributions permit, *ceteris paribus*, more discrimination.
- Reliability: the response stability over-time under conditions of no change in the true scores.
- Equivalence of meaning of the instrument across stimuli and across groups of respondents

In this paper we will look at coverage and equivalence of meaning; we lack the necessary over-time data for adequate reliability assessment, and the relevance of the discrimination criterion is conditional on the assessment of substantive criteria have been assessed.

Substantive criteria for evaluating measurement instruments relate to concept validity. They thus require a conceptual explication of what the instrument was intended to measure and, subsequently, a theoretical explication of what observable differences would be between more and less valid instruments. This brings us to the question what it is that non-ipsative party preferences are supposed to measure. In view of the widespread and varied usage that is made of election study data, there is, of course, no single authoritative account of the concepts that the various non-ipsative party preference measures are purported to measure. One could, of course, revert to a form of operationalism that states that questions that ask about ‘feelings’ measure just that, while other questions measure something else, such as the extent to which one likes or dislikes parties. That is unsatisfactory and tautological, however, unless the concepts of feelings or likes/dislikes are theoretically explicated in such a way as to allow an empirical assessment of the extent to which specific instruments do tap the concept in question.

When looking at the analytical usage of non-ipsative party preferences in the literature, it is quite clear that most authors –usually implicitly– interpret them as expressions of the over-all electoral attractiveness of parties. The work by Rabinowitz, McDonald and their various collaborators (cf. Rabinowitz and McDonald 1989, MacDonald, Listhaug and Rabinowitz 1991), for example, derives conclusions about directional voting from analyses of thermometer ratings. The logic of their analysis rests on the assumption that –apart from random error– the party that is given the highest preference score is the one that will actually be voted for (voting itself being an ipsative form of preference).³ Although –again– usually not explicated, this interpretation of the meaning of non-ipsative preference measures coincides with Downs’ (1957) concept of ‘party utility’ (the benefits – subjective or objective – that a voter gets from voting for a particular party). Indeed, the term party utility, in this sense of the word, is often used interchangeably with that of (non-ipsative) party preferences (how strongly parties are preferred in relation to each other). Downs argues that voters cast their ballot for the party they prefer most, i.e., the party that is to them the most attractive to vote for, and which, in other words, has the highest utility.⁴ This implies a tautological relationship between

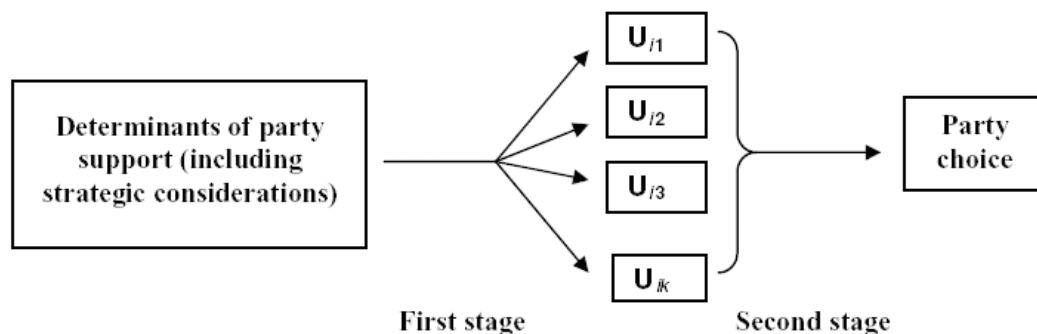
³ Surprisingly, this assumption is never tested by these authors, although the information to do so is directly available to them.

⁴ Downs’ conceptual elaboration of the utility concept is particularly of definitional relevance. It does not lead, however, to falsifiable theories about voting unless no restrictions are placed on the kind of

the most preferred party (the party yielding the highest utility) on the one hand and party choice on the other. This tautology is actually quite useful. As Bingham Powell (2000: 160) argues so eloquently, choice –an ipsative expression of preferences– does not tell us enough about voter preferences to adequately understand electoral behaviour, but the additional information that is required for that purpose is contained in non-ipsative preferences (utilities). This Downsian conceptual foundation yields an immediate observable criterion by which different operationalisations can be distinguished in terms of their concept validity: the extent to which the party chosen is also the party a voter prefers most strongly.

The distinction between utilities (non-ipsative preferences) and discrete choice implies a two-stage model of electoral choice, as illustrated in Figure 1, below. The first

Figure 1: The two-stage model of electoral choice



stage involves the development of utilities; the second stage involves the transformation of non-ipsative utilities in ipsative choice, generally by utility-maximization. The unique advantages of such a two-stage model in the analysis of a variety of electoral phenomena have been elaborated elsewhere (van der Eijk and Franklin 1996; van der Eijk et al. 2006; van der Brug et al. 2007a). In the present context it suffices to derive from this conceptual foundation the (empirically assessable) criterion for concept validity that actual choice coincides with highest preference. We will focus in detail on the performance of three different measures of non-ipsative preferences in this respect. But before we can present those analyses, we first need to introduce the data, and the specific context in which they have been collected.

Data, measures and context

The source of our data is the Irish National Election Study (INES) of 2002 (see Marsh et al. 2008). The study was funded under the Programme for Research in Third Level Institutions (PRTL). This project was co-directed by Michael Marsh and Richard Sinnott (UCD). The survey was conducted by the Economic and Social Research Institute (ESRI)

considerations that are admissible as determinants of utility [insert references]. It will be evident that in our case the tautological perspective is of relevance and that we have no stake whatsoever in the kind of restrictions on utility determinants that many analysts have used in order to arrive at falsifiable theories.

after the election using face-to-face interviews. A three-stage clustered sampling approach was used for sample selection. In the first instance a random sample of 220 PSUs (Primary Sampling Units) was selected from 3,400 District Electoral Divisions (DEDs). All 42 constituencies were sampled as part of this process. In the second, a random sample of 25 households was selected from within each of the PSUs. In the third a random person within the household was selected using the 'next-birthday' rule. No non-sample replacement methods were used. No letter was sent to the respondent prior to the study and there were no payments or gifts used. However, as an incentive, respondents were entered into a lottery, which had eight prizes. Each household was contacted up to five times in order to obtain a respondent and interview. Interviews started a few days after the election and 20 percent were completed within the first month, and two-thirds within two months, 90 percent were completed within two and a half months and almost every interview was carried out within 4 months. For full details of the methodology see Marsh et al 2008; but see also Laver 2005; Marsh 2004, 2006b, 2007.

The INES data are of particular interest for at least two reasons. First, the questionnaire that was used contains three different non-ipsative preference measures, thus allowing rare possibilities for a comparative assessment of the concept validity of each. Second, the Irish electoral system offers more, and more demanding elaborations of the utility-maximizing criterion than most other electoral systems.

Three measures of non-ipsative party preferences

The questionnaire of the INES contains 3 sets of non-ipsative preference questions. One of these is the 'classic' feeling thermometer that derives from the tradition of American National Election Studies. The question is used for politicians, parties, as well as for the candidates in the voter's constituency:

I'd like to ask you how you feel about some Irish politicians, using what we call the 'feeling thermometer'. The feeling thermometer works like this: If you have a *favourable feeling* (a warm feeling) towards a POLITICIAN you should place him/her somewhere between 50 and 100 degrees; if you have a *unfavourable feeling* (a warm feeling) towards a POLITICIAN you should place him/her somewhere between 0 and 50 degrees; if you have a *don't feel particularly warm or cold* (have no feeling towards the politician at all) then you should place him/her somewhere at 50 degrees; where would you place these Irish politicians? And where would you place each of the following PARTIES; and where would you place these CANDIDATES who ran in your constituency in the general election in May?

The second set of non-ipsative preference questions is also asked for politicians (i.e., the national leaders of the major parties) as well as for parties. These questions follow the formulation and format prescribed by the CSES (see also footnote 2). For the political parties this question reads as follows:

I'd like to know what you think about each of our political parties. Please rate each party on a scale from 0 to 10, where 0 means you STRONGLY DISLIKE that party and 10 means that you STRONGLY LIKE that party. If you come to a party you haven't heard of or you feel you do not know enough about, just put a tick (a) in the 'Don't know' box on the right.

This question was not included in the questionnaire for the main face-to-face interview component of the INES, but in a drop-off self-completion questionnaire that was given to respondents after the main interview, with the request to complete it and mail it in.

The third set of questions probing non-ipsative preferences for parties derives was originally developed by van der Eijk and Niemöller in the early 1980s for the Dutch national election studies (van der Eijk and Niemöller 1984; DPES 1982). It has been used in all Dutch national election studies since 1982, in all European Election Studies since 1989, and in a growing number of election studies in European countries, including Britain, Germany, and Spain. This formulation of this so-called ‘propensity to vote’ (PTV) question is:

We have a number of political parties in Ireland each of which would like to get your vote. How probable is it that you will ever give your first preference vote to the following parties? Please use the numbers on this scale to indicate your views, where ‘1’ means ‘not at all probable’ and ‘10’ means ‘very probable’.

Each of these questions about party preferences was asked for the 6 main Irish parties: Fianna Fáil (FF), Fine Gael (FG), Greens (Gr), Labour (Lab), Progressive Democrats (PDs) and Sinn Féin (SF). In addition, the propensity to vote question was also asked for ‘an Independent candidate’. The three sets of questions were separated by significant intervals so as to minimize risk of response set. Nevertheless, it is not unlikely that respondents are increasingly primed into a party evaluation frame of mind as the interview progresses, which should make it easier for them to respond (and respond consistently) to later party preference questions. It is therefore relevant to mention that the propensity to vote questions were the first non-ipsative preference questions in the questionnaire with the thermometer questions—on party leaders, on parties and then candidates – shortly after.⁵ The final set of likes/dislikes question came much later, in the final drop-off section of the questionnaire. This section was completed by 89 per cent of respondents.⁶

The INES is unique in having included all three different forms of these questions in the same study. Most election studies select only one of them for inclusion, which makes it virtually impossible to make a comparative assessment of the ‘performance’ of each.⁷

Institutional context: voting under STV

Ireland uses the single transferable vote system. Voters are presented with a ballot that contains the names –and photographs– of the candidates in their constituency, and the party affiliation of those candidates. Voting consists of indicating a preference ranking for as many candidates as one wants. Figure 2 illustrates this in the form of a completed ballot on which 4 preferences are marked out of a total of 16 candidates listed on the

⁵ The questionnaire was a split one. For a random half of the sample, the intervening questions were open ended ones about what each party stood for; the other half were asked to place each party’s position on both a scale of attitudes to the EU and a scale on environmental protection.

⁶ Many of these were completed as part of the normal interview and other collected later by the interviewer. A number were also posted back to the ESRI.

⁷ The Dutch national election studies has occasionally included the thermometer scale in addition to the propensity to support question, thus also allowing some comparisons of the characteristics of the respective items.

ballot paper. Voters know that their votes are never ‘wasted’ as the allocation of seats redistributes votes that are not required for winning a seat to other candidates according to the patterns of lower preferences marked on the ballot. Additional complexity is generated by the fact that parties may field more than one candidate in a given constituency, while at the same time not always having candidates on the ballot in all constituencies. As a case in point, in 2002 the PDs did not field a candidate in 25 of the 42 constituencies, which were populated by almost 60% of the Irish electorate. Obviously, such no-shows occur particularly in areas where a party expects its support to be too small to warrant investment of scarce resources. As the geographical distribution of party support is different for the various parties, the particular combination of choice options available to voters may differ from constituency to constituency. Just under one third of all voters could choose between candidates of all parties and a further third again were faced with this set, minus a PDs candidates. Apart from these two combinations, the options were varied, with SF fielding candidates in only 34/42 constituencies, Greens in 30/42, Labour in 39/42. A final relevant detail is that also independent candidates compete for votes in all but three constituencies, and not without success.⁸

The variety of availability, and the variety of ways in which voters may use their preferences has to be kept in mind when analyzing and interpreting our survey data. In contrast to election surveys from many other countries we cannot simply look at party choice, but rather we have to distinguish 1st preference votes from 2nd preference votes and from lower ones. Even when examining simple party choice in most political systems, the question arises whether that single expression of preference is a sincere one. This is also the case with the more sophisticated expression of preference under STV. However, the most common reason for casting an insincere preference in other electoral systems – the fear of ‘wasting’ a vote – is not a rational one within STV, as unused votes are transferred to other candidates, as indicated by lower preference votes. Arguably the need for coalitions could undermine sincere preferences, since the coalition partners a party might prefer, or require, might not be those that would be chosen by all of its supporters. However, it could be argued that the uncertainty about post-coalition options is normally so great as to make it very difficult for any voter to know how to use lower preferences to maximise the chance of her preferred party being in government (See Laver 2000 for some discussion of this.)

Finally, it is to some extent an open question whether Irish voters actually vote for parties (who field candidates), or for candidates (most of whom happen to be associated with a party). It is widely assumed that candidate factors are of great importance in Ireland, yet exactly how important candidate and party factors are vis-à-vis each other is still very much a matter of debate (for a more elaborate discussion, see, e.g., Marsh 2007; Marsh et al 2008 Chapter 8). However, it does appear that, to a considerable degree, candidate and party loyalties coexist: if a voter likes a candidate, she likes that party and vice versa. The debate is essentially about how often the party is the chicken and how often it is the egg in this process. Even when a voter indicates support for multiple candidates of the same party, the order of preferences for these candidates cannot simply be assumed to reflect her candidate preferences. Parties who run more than one candidate often try to ensure support is balanced across these candidates, and will sometimes ask

⁸ In 2002 the number of Independent candidates was 95 nationwide, who amassed 9.5% of first preference votes nationwide, and 13 of whom were elected (total size of the Dáil is 166 representatives).




Figure 2: Irish 2002 ballot with 4 preference votes marked

supporters in an area to vote for particular candidates. Therefore they instruct their followers in great specificity as to how to order their preferences over candidates, as illustrated in Figure 3.

For a winning Fine Gael team in Galway East, I ask you to

Vote 1-2 in the order of your choice for
Paul Connaughton
and
Cllr. Tom McHugh

IN THE NORTHERN HALF OF THE CONSTITUENCY
and continue your preferences for
Dr. John Barton and
Ulick Burke.



Enda Kenny, T.D. Leader of Fine Gael

Vote 1-2 in order of your choice for
Dr. John Barton
and
Ulick Burke

IN THE SOUTHERN HALF OF THE CONSTITUENCY
and continue your preferences for
Paul Connaughton and
Cllr. Tom McHugh.

Together we're better – Vote Fine Gael

Figure 3: Segment from Fine Gael election flyer, instructing voters how to order their votes (see areas indicated by ellipses)

Performance of different non-ipsative preference measures

Our discussion of the institutional context within which voters make their choices in Ireland may help to illuminate some of the complexities involved in analysing electoral choice in Ireland. These complexities relate particularly to the interpretation of choices as driven by candidate or by party factors, and thus to our analysis of the construct validity of the three different preference measures by way of the criterion of maximum-preference choice. Before we engage in those analyses, however, we first report a much simpler comparison of the performance of the three measures: the extent to which our interviewees are willing and able to respond to them.

Inability or unwillingness to respond

When looking at the content of the three sets of questions, we see no *a priori* reasons for differences in the percentages of respondents who are unwilling or unable to provide an answer. If we were to have any expectations at all, it would be that questions asked later in the interview would yield lower non-response, on account of the respondents having been longer socialised into their role, and having been more extensively primed into a mindset of evaluating parties. Yet, other studies often do reveal systematic differences in non-response that could not be deduced on the basis of first principles from a visual or linguistic inspection of the questions involved. As a case in point, a comparison of thermometer ratings and propensity to vote scores in the Dutch National Election Study yields a considerably higher incidence of nonresponse in the former (cf. Kroh 2001). When assessing the three instruments in the INES in this respect we limit the comparison to only those respondents who completed the drop-off questionnaire (n=2367) –as the like-dislike instrument was part of that questionnaire. Table 1 provides the proportions of nonresponse for each of the parties separately, as well as in a listwise fashion.

Table 1: Proportion of nonresponse to three measures of non-ipsative party preferences.
Data: INES2002, n=2367

	FF	FG	Green	Labour	PDs	SF	Indep	Listwise
PTV	0.01	0.02	0.03	0.02	0.02	0.02	0.02	0.04
Thermometer	0.02	0.04	0.03	0.03	0.04	0.04	inap	0.05
Like/Dislike	0.06	0.08	0.16	0.12	0.11	0.15	inap	0.21

Table 1 shows that the propensity to vote measure (PTV) generates the least nonresponse, even listwise, while it is asked for 7 stimuli, in contrast to the other two measures which are only asked for 6. The feeling thermometer yields consistently more nonresponse than the PTVs, although the difference is small (and smaller than in the case of a similar comparison in the Dutch National Election Study). The largest difference we find, however, between PTVs and thermometer scores on the one hand, and the like/dislike scores on the other. It is not entirely clear whether this difference can only be attributed to the instrument itself. Part of it may be caused by a different mode of data collection (face-to-face for PTVs and thermometer, self-completion drop-off questionnaire for the likes/dislike scale). Additionally, we see a slight tendency for smaller parties to generate somewhat higher proportions of nonresponse, but, over-all, the main conclusion must be that non-ipsative preference questions do not pose insuperable cognitive or emotive barriers to interviewees.

A more challenging –and substantively more interesting– assessment of the relative merits of these questions concerns their relationship with the actual choices made by voters on the ballot. That is what we turn to now.

Relationships between non-ipsative preferences and actual choice

As discussed earlier, non-ipsative preferences for parties are used in the literature as measures of the electoral attractiveness of parties or of what Downs (1957) calls ‘electoral utilities’ – although it has to be emphasised that most authors are hardly explicit with respect to their conceptual interpretation of these instruments. In the case of valid measures of that concept, knowledge of which party received a voter’s highest preference score does not yield anything new over and above his/her report of which party was actually supported on the ballot.⁹ The two are –conceptually– identical. The major advantage of valid non-ipsative preference questions is something else: they shed light on a structure of preferences that remains in the dark if the only information at our disposal is ipsative preference in the form of vote choice (cf Bingham Powell 2000:160). And that information is invaluable for improving explanatory models of party choice (cf van der Brug et al. 2007a), for studying party competition (cf van der Eijk and Oppenhuis 1991; Kroh et al. 2007) and the determinants of the support of small parties (cf. Van der Brug et al 2000) and various other phenomena in the field of electoral studies (for a fuller discussion, see van der Eijk et al. 2006).

From this conceptual background we can deduce a simple and straightforward criterion for construct validity: highest preferences should coincide with choice. When

⁹ The only additional information that can be derived from a valid highest preference score is how someone who did not vote would have voted had he/she turned out to vote.

applying this criterion to the Irish case a number of complications arises that derive from the specific institutional setting of Irish elections, as summarised earlier. Should we look at first preference vote or at all preference votes? How do we take the differences into account in the availability of choice options on the ballot? How do we distinguish between instances of candidate voting and instances of party voting that are indistinguishable in terms of their manifestation on the ballot? And, finally, how can we prevent the results from being influenced artefactually by differences in response format –particularly the difference between a 0-100 scale in the thermometer instrument, and a much shorter scale for the other instruments (1-10 for PTVs, 0-10 for likes/dislikes).

To keep the degree of complexity manageable we decided to investigate this criterion first with respect to the first preference vote, and only thereafter for any lower preference vote that involves a different party than the one getting the first preference vote.¹⁰

Secondly, we decided to present the results of our analyses using two different ways of handling the differences in the electoral supply side. On the one hand we disregard the problem and conduct our analyses on all respondents who completed the drop-off questionnaire, and on the other hand we select a relatively small segment of the sample consisting of the voters in those constituencies where a full slate of choice options was available (i.e., the 6 major parties listed earlier and one or more independent candidates). This segment consists of 845 respondents (slightly more than 30% of the realised sample).¹¹ We did look into all intermediate situations, i.e., constituencies where only PDs were missing, or only Labour, or where the ballot contained neither a Green nor an Independent, etc. Presenting results for all those conditions separately serves no purpose as the results turn out to be exceedingly similar, as is illustrated by the two extreme cases that we do present.

Third, in order to tighten the interpretation of the results in terms of whether they are reflect candidate voting or party voting, we differentiate voters in that respect. We make use of a 3-fold typology that distinguishes candidate-oriented voters, party oriented ones, and a group with mixed orientations. This typology was developed by Marsh et al (2007, Chapter 8) and is based on responses to 2 questions, one whether respondents consider party or candidate to be the most important factor determining their choice, and the second whether or not they would still vote for the same candidate if (s)he were to switch to a different party.¹² We expect that the most-preferred party will more often be

¹⁰ Keep in mind that the second party voted for does not have to be same as the party of the second preference vote. Parties can have more than one candidate in the same district, which makes it possible that first and second preference vote involve the same party, in which case the results in terms of relationship between choice and non-ipsative preference will be the same. Instead, we are interested in the first other party that is supported after that of the first preference vote (if any).

¹¹ Not all of these are useful observations for the analyses to follow. A sizeable portion of this group did not vote; a small number did vote but for one of the few minor parties (Christian Democrats, Workers Party, etc.) for which we have no preference ratings, some voted for independent candidates for whom we have no thermometer or like/dislike scores. Owing to these factors, the number of valid observations in Table 2 varies somewhat for the different instruments.

¹² We do not believe that self-attributions of motivations constitute a fully adequate basis for classifying people into party- or candidate-oriented groups (see, e.g., Kaplan 1964 on the distinction between act meaning and action meaning). Yet we think it plausible that to the extent that such differences in orientations exist amongst voters, they will be correlated with these self-attributions.

different from the one actually chosen for candidate-oriented voters when compared to party-oriented voters.

Finally, we decided to look at the thermometer scale in its original (0-100) form, as well as in a recoded 10-category form, in order not to bias our assessment of the thermometer instrument downwards as minor random differences are more likely to result in a deviation from maximum-preference voting when using a much larger number of categories.

The results of our analyses can be summarised in a set of percentages that display how often the most preferred party is the same as the party of the first preference vote. This is reported in Table 2.

Table 2: Percentages of highest-preference voting for different measurement instruments and different groups of voters
Data: INES2002; valid N: 1466-1643 (left panel), 450-502 (right panel); see also footnote 9.

	Disregarding differences in party supply				Only constituencies with full party supply			
	PTV	Thermo- meter	Therm. recoded	Like / dislike	PTV	Thermo- meter	Therm. recoded	Like / dislike
Candidate-oriented voters	82.0	59.6	60.7	73.4	81.2	57.3	58.9	72.4
Voters with mixed orientations	87.5	66.3	67.4	86.1	88.3	67.0	70.0	83.9
Party-oriented voters	94.5	70.5	70.7	92.1	90.6	66.7	67.7	89.5
TOTAL	87.1	64.9	65.7	82.9	86.5	63.6	65.6	81.6

Inspection of Table 2 shows, first of all, that it makes very little difference whether or not we take account of the supply of parties from which voters can choose. Disregarding this aspect (left panel of Table 2) unavoidably leads to a number of false negatives: voters seen as not voting for their most preferred party only because that party was not present on their ballot. The minimal differences between the left panel and the right one (where no false negatives can possibly occur, as voters can choose from all parties for which they expressed preferences – see also footnote 9) indicates on the one hand that parties are quite astute in not fielding candidates in areas where their support is minimal. On the other hand, it also may indicate a tendency of voters to develop preferences only for choice options that are actually available to them, which would imply that – in the long run – parties stand to gain from being present on the ballot in as many constituencies as they can manage.

The next conclusion that can be drawn from Table 2 is that there are considerable differences in the extent to which the three instruments stand up against the criterion that the most preferred party should be the same as the one actually chosen. For the entire groups of voters for whom we can make this assessment, as well as for each of the three subgroups, we find that PTVs perform in this respect much better than the other two

instruments. The like/dislike format does considerably better than the thermometer scores. When comparing the three instruments, it should be kept in mind that the criterion of highest preference voting becomes – *ceteris paribus* – more demanding as the number of options increases. In the case of the PTVs this criterion is based on 7 options (6 parties plus Independents), whereas only 6 options underlie this test of the other two instruments.

The incidence of highest preference voting is –as expected– highest among party-oriented voters. Yet, the PTV measure, in particular, shows that even candidate-oriented voters behave mostly (in excess of 80%) in a party-driven fashion: voting for the *party* that has their highest preference. This is not to deny that candidates may also be important, but it illustrates the weakness of typologies of candidate – versus party-orientation that are based on self-assessments (see also footnote 10). The question remains, of course, whether party preferences are driven by candidate preferences, or the other way around, or whether party and candidate preferences affect each other reciprocally. Yet, interesting as these observations may be, they are somewhat beside the remit of this paper, which is to make a comparative assessment of three different instruments for measuring the electoral attractiveness of parties by way of non-ipsative preference measures. As far as that central question is concerned, the results so far point unequivocally to the PTV as the most valid of the three instruments. This assessment is based, however, on first preference votes only. Earlier we allowed that the STV system offers the possibility of rational, party-driven preference voting that does not manifest itself in first preference votes, but rather in lower preference votes. Our assessment of the performance of the three instruments is therefore not yet complete, and we will turn to lower preference votes next.

Non-ipsative preferences and lower preference votes

Comparing lower preference votes with party preferences only makes sense when the lower preference vote involves a different party than the first preference vote, as otherwise we simply duplicate earlier findings. This restriction leads to the loss of some cases: respondents who did not give any of their lower preference votes to a candidate of a different party, or if they did, voted for one of the minor ‘other’ parties for which we have no preference ratings.¹³

In view of the great similarity of the findings in the left and right panels of Table 2 we limit the presentation of findings only to those respondents who are offered a full party supply in their constituency. Moreover, we cannot restrict ourselves just to reporting the percentage of cases where the relevant higher preference vote is for the party that was given the highest rating score. That would misrepresent the relationship between the non-ipsative preferences (ratings) and actual lower preference choices. Obviously, when the first preference vote is for the party that was given the highest rating score, we cannot find the same relationship anymore for a lower preference vote – except when the two parties involved are tied at highest rating score. We rather have to assess in that case whether or not the next party chosen is the one that was given the next highest rating, and so on. In view of this we have to distinguish a number of possible relationships between preferences and choice, as displayed in Table 3.

¹³ Only 1 per cent of or sample supported such a party as their second (party) choice and less than 1 per cent did so as a first choice.

*Table 3: Relationship between non-ipsative preference and lower preference vote while taking into account the same relationship for first preference vote (percentages)
Data: INES2002*

	PTV	Thermometer	Like/dislike
Category 1: 2nd party best	5.5	3.2	6.1
Category 2: 2nd party tied for best with first	28.5	10.4	13.1
Category 3: 2nd party next best after 1st	33.0	28.0	36.9
Category 4: 2nd party better than first but not best	0.00	9.3	6.6
Category 5: 2nd party not best after 1st	33.0	49.1	37.3
Total	100.0	100.0	100.0
Sum of categories 2 and 3	61.5	38.4	50.0
valid N	397	279	244
Selected sample section (N)	401	401	401
valid N as % of selection	99.0	69.6	60.8

Category 1 in Table 3 consists of those respondents for whom their second party vote for is cast for the party to which they gave the highest preference, while the party of the first preference vote was given the next highest preference rating. At first sight this seems to defy the notion of voting according to highest preference, but it is not necessarily the case. Earlier we mentioned that desires additional to who should get elected may in the STV setting lead to strategically giving the first preference vote to a party that is not the most preferred, and a higher preference vote to the party that is best in terms of preference. Yet, whether or not this is actually the case is unknown to us, so that we have to consider this category as equivocal in terms of our assessment of highest utility voting for lower preference votes.

Category 2 describes respondents who had given their highest preference rating to both parties, so that both manifest highest preference voting. The magnitude of this category indicates that ties at highest preference are relatively common, particularly in the PTV measure.

The third category in Table 3 is the straightforward situation where preference votes follow the order of non-tied preference ratings: the second party chosen is the next highest in terms of rankings. A clear and unequivocal manifestation of highest utility voting.

Category 4 manifests a clear refutation of highest-rating voting. Neither of the two parties chosen is best in terms of non-ipsative preference ratings. It is conceivable that these respondents have given a yet lower preference vote to a third party that is the best in this respect. In that case they do not refute highest preference voting. Yet, as we will see, the incidence of highest-preference voting seems to be lower in lower preference

votes than in first preference votes, which makes it unlikely that this category would contain many false refutations, particularly also in view of the small size of this category. Category 5, finally, manifests situations where the second party voted for defies the utility-maximization criterion, although this does not imply that the same holds for the party of the first preference vote.

In our view, categories 2 and 3 both reflect situations in which a second party chosen follows the utility-maximization criterion. The total of these 2 categories is indicated in Table 3 for each of the three instruments. The PTV measure here also performs considerably better in this respect than either of the other two measures. Yet, even when using the PTV measure, and disregarding the ambiguous category 1, in more than one third of the instances in which a second party is chosen in a lower preference vote the party utility-maximising rule is violated, a much higher percentage than we found for first preference votes. This may indicate different things. One possibility is that the first preference vote is the one that really expresses party preference, while subsequent choice of parties represents something else

A second possibility, already mentioned, is that apparent refutations of highest-utility voting are caused by the fact that we have not included all subsequent preference votes in our analysis. We think that it is unlikely that extending the analyses to further parties chosen would change the picture over much. One reason for this is the decreasing number of voters who express on the ballot support for a third, or even a fourth or yet further party. The second reason is that extending the analysis to third and further parties chosen carries an increasing risk of non-falsifiable, hence trivial propositions. Those few voters who rank all candidates (and thus also all parties in their constituency) on their ballot will by necessity have voted for the parties that yield them maximum utility, second-best utility, etc.

A final aspect of relevance in Table 3 concerns the last rows, which indicate the extent to which the relevant segment of the sample is ‘covered’ by each of the three preference measures. The large difference that we see here is partly attributable to the fact that the PTVs cover 7 choice options, and the other two measures only 6 options. But in addition they also reflect differences in non-response. It is important to keep track of this coverage, as good performance of a measure –a high incidence of utility-maximizing voting— loses much of its relevance if the measure is inapplicable to large groups in a population (or representative sample of that population).

Reviewing the findings of this together with the previous section in the light of our main concern – how do the three non-ipsative preference measures fare in terms of construct validity— we can only conclude that the PTV measure outperforms thermometer scores as well as like/dislike scores by a wide margin. Does this imply that the other two measures are less valid as measures of the electoral attractiveness of parties? That question cannot be answered on the basis of our analyses so far. It is quite conceivable that, for example, all three measures are equally valid yet differ in terms of the amount of noise they contain. This question is partly one of reliability, but also partly one of latent structure. As stated earlier, we cannot unequivocally address reliability issues because of lack of replications under circumstances of no latent change.¹⁴ However, if all three

¹⁴ The sample of the INES2002 study has been re-interviewed almost annually up to an including 2007. The long period between repeated observations makes the assumption of no latent change –a necessary assumption to deduce reliability from observed stability— implausible. Elsewhere we will use this panel

instruments are valid measures of the same concept, they are *congeneric* in the psychometric sense of the word. And the available data allow us to test whether or not this is the case, by using latent structure analysis.

Latent structure analysis of non-ipsative preferences

If the three rating measures of non-ipsative preferences measure the same phenomenon, they should be congeneric or homogenous. Although this is a common notion – widely applied in various forms of latent structure analysis of, e.g., attitude items – it is less self-evident what this implies for our current measures. In contrast to many other applications of the same reasoning, we are here interested in the homogeneity of instruments that are not single items, but sets each consisting of multiple items. When considering each of the instruments in isolation, we can analyse its items in order to diagnose the latent structure underlying them. That latent structure is usually interpreted in a dimensional sense. If all items of an instrument can be shown to express a single latent variable, then they can all be considered to be equally valid as expressions of – in our case – over-all electoral attractiveness, or electoral utility in the Downsian sense of the word. If one or several items do not (or do not sufficiently strongly) reflect the same latent trait as the others, then the non-ipsative instrument in question may be valid for some parties, but not for others. This does not necessarily lead to serious violations of the utility-maximizing rule used as criterion for construct validation in the previous sections, particularly not when the ‘deviant’ party is small.

We can thus extend our assessment of construct validity of the three measurement instruments by analysing each in terms of the commensurability of the observations with a single latent trait. But, more interestingly, we can use latent structure analysis also to investigate whether the three instruments all measure the same underlying structure, by jointly analysing the items of each. If the three instruments measure the same thing, then we should find that they all express the same underlying dimension. In that case, the differences in performance according to the utility-maximization criterion used in the previous sections would seem to be mainly attributable to differences in ‘noise’ or reliability.

What model to use for assessing latent structure depends on the character of the items. In our case, each item expresses a similarity relationship between a voter and a party. In other words, they constitute a particular kind of single stimulus data (Coombs 1964). Moreover, as we cannot expect the items to be located at the same position in the latent space, factor analytic methods are inappropriate (cf. van Schuur 1993; van Schuur and Kiers 1994). Models of the IRT (Item Response Theory) tradition are more suitable for these data. We will use the nonparametric unfolding model MUDFOLD, as it does not require us to buy into various kinds of assumptions that are either implausible or that appear arbitrary in the context of our data.

As do all unfolding models, MUDFOLD attempts to construct a single representation of both persons and stimuli in a latent space. Both should be located in this so-called joint space so that the distances between persons and stimuli reflect (inversely) the empirically

information for specifying models that estimate change and reliability simultaneously, but such analyses are beyond the scope of the present paper.

observed preferences of each person individually: the more a respondent prefers a party, the smaller the distance between them should be in the spatial representation. Formally, this amounts to the assumption of single-peaked preferences. These relations should hold to a satisfactory degree for all voters and for all parties. An interesting aspect of such representations –if they exist– is that they demonstrate that all respondents use largely the same criterion in rating all of the parties.

MUDFOLD provides a goodness-of-fit measure – the coefficient of homogeneity H – which can be used to compare the strength of a scale across instruments and subgroups. H attains an upper limit of 1 if the constructed scale represents the data perfectly, without any violations. If, on the other hand, a proposed scale yields as many violations with empirical observations as would occur in the case of statistical independence of the stimuli, H is 0. Generally, the lower boundary for acceptable scalability of survey data is set at $H=0.30$. H -values surpassing 0.40 indicate a medium-strong scale, and above 0.50 a strong scale. In addition to the H coefficient for a scale in its entirety, item-specific coefficients – H_i – indicate how well each of the items fits in a given scale.

Unfolding PTV ratings of parties

The unfolding analysis of the 7 PTV items yields a medium strong scale, as reported in Table 4.¹⁵ The table reports the order of the parties in the underlying latent dimension, and the degree of scalability of the set of items as a whole ($H=0.40$) and for each of the parties as part of this scale (H_i).

The order that was found to represent the observed preferences best is:

FF / PDs / FG / Independent / Labour / Greens / SF¹⁶

Voters are located on the same dimension, and as the model is based on the assumption of single-peaked preference functions, this order implies voters who have expressed a

*Table 4: Unfolding scale of PTV ratings of Irish political parties and Independents
Data: INES2002*

	H_i (parties)
	H (scale)
FF	0.38
PDs	0.44
FG	0.36
Independent	0.37
Labour	0.44
Greens	0.44
SF	0.41
Scale	0.40

¹⁵ In order to avoid comparability problems owing to different rates of nonresponse, we conducted the all unfolding analyses on the group of respondents who gave valid responses to all items of all three non-ipsative party preference instruments (N=1801).

¹⁶ This is the ordering also found by analysis of PTVs in another survey conducted in 2001-2002. For this and orderings from earlier years see Marsh 2006a.

high preference for a particular party tend to also have relatively high preferences for parties that are adjacent in this ordering, and increasingly lower preferences for parties that are further removed. The fit of the unfolding model for these data is quite acceptable, although the H_i 's indicate that preferences for FF, FG and Independents generate somewhat more violations of the model than preferences for the other parties do. We find, however, that these perturbances are to some extent caused by the inclusion of the 'Independent' stimulus. This is not surprising in view of considerable differences in outlooks and policy emphases between independent candidates. Replicating the analysis for only the 6 parties –leaving out Independents– yields a scale with the same ordering, and slightly stronger scalability, as reported in Table 5.

*Table 5: Unfolding scale of PTV ratings of Irish political parties
Data: INES2002*

	H_i (parties H (scale))
FF	0.38
PDs	0.46
FG	0.40
Labour	0.45
Greens	0.48
SF	0.34
Scale	0.42

Table 5 shows that SF is the weakest item in terms of scalability, although it still fits acceptably in the scale. This suggests that PTV-preferences expressed for SF either contain somewhat more random noise than is the case for other parties, or express, to a small extent, something other than (or additional to) electoral utility. The drop in scalability is not, however, of such a magnitude that we should conclude that PTV-preferences expressed for SF have to be interpreted in a different light as those given to other parties.

Unfolding of Thermometer and Like/Dislike ratings of parties

Similar analyses as for the PTV ratings were done for the thermometer and like/dislike ratings. The results are remarkably similar to those reported in Table 5. The ordering in

Table 6: *Unfolding scale of Thermometer ratings and Like/Dislike ratings*
 Data: INES2002

	Thermometer H_i (parties H (scale))	Like/Dislike H_i (parties H (scale))
FF	0.38	0.33
PDs	0.45	0.42
FG	0.39	0.36
Labour	0.45	0.42
Greens	0.37	0.41
SF	0.30	0.35
Scale	0.39	0.38

which parties are unfoldable is the same, and both the thermometer and the like/dislike ratings are scalable at an acceptable level. Yet, each of these sets of preference ratings is slightly weaker in its structure than the PTV unfolding scale, and the H_i of the thermometer rating of SF is only borderline acceptable.

Unfolding all non-ipsative party preference ratings jointly

The somewhat weaker scalability of the thermometer and the like/dislike ratings – when compared to the PTV ratings – can conceivably be caused by higher levels of random error or noise. If that were the case, it would also explain the poorer performance of those two measurement instruments in terms of reflecting utility-maximising choice behaviour (see previous section, Tables 2 and 3) and it would indicate that, although they are not all equally accurate, all three instruments for measuring non-ipsative preferences are valid and that they measure the same phenomenon: the electoral attractiveness of parties.

A test of the hypothesis that all three measurement instruments measure the same thing consists of jointly unfolding the items of PTV, thermometer and like/dislike ratings. The results of a large series of variations of such a test can be summarised extremely concisely: the hypothesis that these measurement instruments are congeneric is resoundingly rejected. Using any of the three scales reported in Tables 5 and 6 as starting point and attempting to extend that scale by acceptable items from the other instruments fails. Sometimes one or two items can be added to an original scale –at the cost of a considerable loss of structure, as expressed in the H coefficient– but never more. Forcing the items of any two (let alone of all three) of the instruments into a single scale results invariably in dramatically low H coefficients.

The conclusion of these failed attempts at finding a shared latent structure across the three measurements can only be that they do not measure the same thing. Warm and cold feelings as expressed in a thermometer scale do not express the same thing as do expressed propensities to vote for a party, nor as expressed like/dislike ratings.

Conclusion and implications

We argued that non-ipsative preference ratings are of great importance for the development of better explanatory models of electoral behaviour and choice. But this holds only if such preference ratings are valid operationalisations of the electoral attractiveness of political parties, i.e. of what Downs refers to as electoral utilities.

Different traditions of measuring non-ipsative preferences exist. Yet, because most studies reflect only a single one of these, comparative assessments of their respective merits have been hampered. The Irish National Election Study 2002, however, includes three different of such preference questions: thermometer ratings, like/dislike ratings, and propensity to vote for party ratings. We use these data for a comparative analysis of the validity of these measures as operationalisations of the concept of electoral utility. The application of these measures in the Irish context also allows more stringent assessments of their performance than would be possible in many other countries, as the STV electoral system permits voters to cast multiple, ordered choices, and thus allows more sophisticated comparisons between actual choices on the one hand and preference ratings on the other.

Considerations of construct validity require that, for a measure to be valid as an operationalisation of electoral utility, it must display the tautological relationship between highest preference rating (in conceptual terms: highest utility) and actual party choice. This ‘highest utility’ criterion was clearly best fulfilled by the propensity ratings, with a success rate of 86.5% (using party chosen in STV first preference vote), versus 81.6% for the like/dislike ratings and only some 65% for the thermometer ratings. When looking at the correspondence between non-ipsative preference ratings and other parties chosen in lower preference votes under STV we find again that the propensity scale is much better able to capture utility-maximisation in choice than either of the other two measures. Yet, the success rate is considerably lower than in the case of first preference voting, partly because of unavoidable ambiguity in the interpretation of some patterns of STV’s first and lower preference votes.

Each of the measures of non-ipsative preferences was asked for all 6 of the Irish parties represented by more than one deputy in the 2002-2007 parliament, while the propensity scale also includes a rating for ‘Independents’. When investigating the internal structure of the responses to each of these batteries of party preferences separately, we find that they express underlying uni-dimensional structures, each of which orders the parties in the same way. This desirable property of each of the batteries is slightly more pronounced for the propensity items than for the thermometer or like/dislike ratings. Yet, when all non-ipsative preference ratings from all three instruments are subjected to a common analysis, they defy being combined in an acceptable unfolding scale. Evidently, although the underlying factors that are expressed in the responses to each of the batteries of items are related, they are not identical.¹⁷ This leads, in combination with the

¹⁷ From the unfolding algorithm a score can be calculated that represents respondents’ positions on the underlying continuum. Unsurprisingly, the scores on these latent variables are significantly correlated. The strongest correlations are found between the PTV-based score and each of the other scores (0.66 and 0.62 respectively), whereas the scores on the latent dimensions expressed in thermometer and like/dislike ratings

performance of the three batteries on the ‘highest utility’ criterion to the conclusion that the propensity ratings are the most valid measurements of the intended concept –the electoral attractiveness of parties or, in Downsian terms, electoral utility. The other measures express something that is related to this, yet different from it. One could, of course, ask what it is that is expressed by those other instruments. We will not engage in that question for two reasons. First, the purpose of this paper was to assess the relative merits of three different operationalisations of the same concept. That we have done. Second, we feel it is not rewarding to engage in a social science dating enterprise where “nice operationalisation seeks compatible concept”. In explicatory analyses of electoral choice, there is a strong need for valid non-ipsative measures of electoral utility. The analyses in this paper have indicated that such a measure exists. Having established that, it is time to leave the field of instrumentation and enter that of instrument-application, i.e., substantive explanations, of electoral choice.

The excellent performance of the propensity measure does not come as a total surprise, as it has earlier been demonstrated to generate a success rate in excess of 90% on the highest utility criterion in most European countries (cf. Van der Eijk et al 1996; Van der Eijk et al. 1999; Van der Brug et al. 2007b). Most of those analyses pertained to the second-order national elections for the European Parliament, but Dutch national election data showed that the measure also performs very well in first-order contexts (Van der Eijk and Niemöller 1984; Tillie 1995). Those earlier findings, however, were mostly unable to assess the merits of the other non-ipsative preference measures that were analysed here. The only earlier comparison between instruments that we are aware of was done by Kroh (2001), who systematically compared the performance of thermometer ratings and propensity ratings in the Netherlands. His conclusions mirror ours: thermometer ratings are much more loosely connected to actual party choice than are propensity ratings, and therefore are much less adequate as measures of the electoral attractiveness of political parties.¹⁸ It might be more surprising that such a strong link is found between propensity rating and party choice in the Irish context, as there is so much evidence that Irish voters are apparently less party centred than voters in most other countries. However, these findings make it clear that even if voters are drawn to a party via a candidate, they still tend to vote for that party which has the highest utility rating, and moreover, such voters tend to give their second (party) preference to the next most highly rated party.

At least two general implications of our findings deserve specific mention, one analytical, and one with respect to data collection. The analytical implication of our findings is that researchers who utilise thermometer ratings (e.g., in ANES data) or like/dislike ratings (in CSES data) must be aware of the considerable slippage between these ratings and actual choice. As a consequence, to the extent that their inferences about

are intercorrelated at 0.57. Although these correlations are of a respectable magnitude, they also indicate that the shared variance of respondents’ positions on these latent variables ranges somewhere between 0.32 and 0.43, which leaves ample room for mutually incompatible meaning components. In a later publication we will use these scores in a more elaborate latent-variable model explicating the relationships between these and other central variables in the explanation of electoral choice.

¹⁸ Using data from the 1994 Dutch Parliamentary Election Study he finds that 72.6% of the respondents vote according to the highest rating on the thermometer measure, while the corresponding percentage for the propensity rating was 94.6.

choice behaviour are based upon analysis of thermometer or like/dislike ratings they run considerable risks.

The implication with respect to data collection is that the decision as to to which of several measures of non-ipsative party preferences to include in a survey cannot anymore be made innocently on the basis of personal taste or even continuity considerations. In the presence of empirical evidence about stark differences in the validity of different measures the only sensible choice is to include the best measure in election surveys. The best –in the present comparison clearly the propensity ratings– is not more difficult or more expensive to implement than inferior instruments. There is no excuse for choosing second rate instruments.

References

- Coombs, Clyde. 1964. *A Theory of Data*. New York: Wiley.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.
- Brug, Wouter van der, Cees van der Eijk, and Mark N. Franklin. 2007a. *The Economy and the Vote*. Cambridge, MA: Cambridge University Press.
- Brug, Wouter van der, Cees van der Eijk, and Mark Franklin. 2007b. "EU Support and Party Choice," in *European Elections and Domestic Politics. Lessons from the Past and Scenarios for the Future*, W. van der Brug and C. van der Eijk (eds.). Notre Dame: University of Notre Dame Press, pp.168-188.
- Brug, Wouter van der, Meindert Fennema, and Jean Tillie. 2000. "Anti-immigrant parties in Europe: Ideological or protest vote?" *European Journal of Political Research* 37: 77-102.
- van der Eijk, Cees, Wouter van der Brug, Martin Kroh, and Mark Franklin. 2006. "Rethinking the Dependent Variable in Voting Behavior – on the Measurement and Analysis of Electoral Utilities," *Electoral Studies* 25: 424-447.
- van der Eijk, Cees, Mark N. Franklin, and Wouter van der Brug. 1999. "Policy Preferences and Party Choice," in *Political Representation and Legitimacy in the European Union*, Hermann Schmitt, and Jacques Thomassen (eds.). Oxford: Oxford University Press, pp.161-185.

- van der Eijk, Cees, Mark N. Franklin, and Erik V. Oppenhuis. 1996. "The Strategic Context: Party Choice", in *Choosing Europe? The European Electorate and National Politics in the Face of Union*, Cees van der Eijk, and Mark N. Franklin (eds.). Ann Arbor: The University of Michigan Press. pp.332-365.
- van der Eijk, Cees, and Erik V. Oppenhuis. 1991. "European parties' performance in electoral competition", *European Journal of Political Research* 19: 55-80.
- van der Eijk, Cees, and Kees Niemöller. 1984. "Het potentiële electoraat van de Nederlandse politieke partijen," *Beleid en Maatschappij* 11: 192-204.
- Kaplan, Abraham. 1964. *The Conduct of Inquiry. Methodology for Behavioral Science*. Scranton: Chandler.
- Kroh, Martin 2001. *An Empirical Analysis of the properties of the Probability of Future Vote Question (PFV) and the Party Sympathy Scale (PSS)*. Amsterdam: Amsterdam School of Communication Research (internal manuscript)
- Kroh, Martin, Wouter van der Brug, and Cees van der Eijk. 2007. "Prospects for Electoral Change," in *European Elections and Domestic Politics. Lessons from the Past and Scenarios for the Future*, W. van der Brug and C. van der Eijk (eds.). Notre Dame, IN: University of Notre Dame Press, pp. 209-225.
- Laver, Michael. 2000. "STV and the politics of coalition", in *Elections in Australia, Ireland and Malta under the Single Transferable Vote: reflections on an embedded institution*, Shaun Bowler and Bernard Grofman (eds.), Ann Arbor: Michigan University Press, pp. 131-152.
- Laver, Michael. 2004. "Analysing structures of party preference in electronic voting data", *Party Politics* 10::521-41
- Laver, Michael. 2005. "The electorate," in *Politics in the Republic of Ireland*, John Coakley and Michael Gallagher (eds). London: Routledge (4th ed.) ,
- MacDonald, Stuart Elaine, Ola Listhaug, and George Rabinowitz. 1991. "Issues and Party Support in Multiparty Systems," *American Political Science Review* 85: 1107-32.
- Marsh, Michael. 2004. "None of your post-modern stuff around here: grassroots campaigning in the 2002 Irish general elections" in *British Elections and Parties Review*, Roger Scully et al (eds.), Volume 14. Frank Cass: 245-67.
- Marsh, Michael. 2006 a. 'Stability and change in structure of electoral competition 1989-2002', in *Facing change in a new Ireland: An analysis of Irish social and political attitudes*, Diane Payne, John Garry and Niamh Hardiman (eds.) Dublin: Liverpool University Press.

- Marsh, Michael. 2006 b “Party identification in Ireland: An insecure anchor for a floating party system” *Electoral Studies* Volume 25(3) 2006, pp. 489-508
- Marsh, Michael. 2007. “Candidates or parties? Objects of electoral choice in Ireland”, *Party Politics* 13 (2) 2007, pp.500-27
- Marsh, Michael, Richard Sinnott, John Garry and Fiachra Kennedy. 2008, *The Irish Voter: The Nature of Electoral Competition in the Republic of Ireland* Manchester: Manchester University Press.
- Powell, G. Bingham, Jr. 2000. *Elections as Instruments of Democracy*. New Haven, CT: Yale University Press.
- Rabinowitz, George, and Stuart Elaine MacDonald. 1989. “A Directional Theory of Issue Voting,” *American Political Science Review* 83: 93–121.
- van Schuur, Wijbrandt H. 1993. Nonparametric unidimensional unfolding for multicategory data. *Political Analysis* 4, 41-74.
- van [Schuur, Wijbrandt H.](#), [Kiers, Henk A. L.](#) 1994. “Why Factor Analysis Often Is the Incorrect Model for Analyzing Bipolar Concepts, and What Model To Use Instead.”, *Applied Psychological Measurement*, 18: 97-110
- Tillie, Jean. 1995. *Party Utility and Voting Behavior*. Amsterdam: Het Spinhuis.