Measuring Attitudes: Using Branching and Numerical Scales

Davide Lubian [1]
[1]Social Research and Sociological Methodology Department,
Sapienza University of Rome, Corso Italia 38/a, 00198 Rome, Italy
davide.lubian@uniroma1.it

Abstract

Past research has recognized the effectiveness of verbally labelled branching questions with two aims: to simplify the judgement task, and to explore and optimize the midpoint answer. However, I recommend branching since positive and negative extremity answers are asked in the second step, both with positive labels, resulting in the same unit of measurement.

Numerical scales are usually inappropriate for measuring attitudes because they exclude the negative part of the continuum. Accordingly, I conducted a split-ballot experiment (SAQ; N=146) to study the effect of branching with numerical scales, comparing 1-10 scales of trust toward 16 different institutions with their branching versions (i.e. dichotomous question followed by 5-point scale of extremity). Only a few studies have attempted to disambiguate the meaning of numerical scales using negative numbers. Their findings showed a positivity bias effect, which significantly shifts distributions to the positive side. I argue that there is also an avoidance effect with negative numbers, because respondents do not use them in everyday life. The solution that includes negative answers, but not negative numbers with rating scales, is branching.

Results of an MTMM analysis, including four Guttman scales, showed higher convergent and discriminant validity for the branching format. Although branching scales had an average mean higher than the unipolar scales (5.6 vs. 4.6), respondents did not avoid negative answers (46% in the first step) as happened elsewhere with negative numbers. Furthermore, branching allows a better use of the scale gradation (SD 2.7 vs. 2.4), especially for the positive side.

This study warns researchers against employing unipolar rating scales for measuring bipolar constructs, as attitudes are, suggesting that negative answers, either verbally or numerically labelled, have a different unit of measurement than positive answers. This asymmetry implies a comparison problem between the two directions of the evaluative continuum, which branching can reasonably resolve.

Key Words: Attitude measurement, branching, numerical scales, bipolar scales, bipolar constructs, rating scales

1. Why branching bipolar questions

Scientific debate concerning rating scales in survey research is not yet arrived to an end and I will present some further arguments that required to be explored. The most

discussed topics have been related to the length and to the labelling of rating scales: how many points are optimal, is it better to adopt a numerical or a verbally labelled scale, with partial or full labelling? It is acknowledged that it mainly depends on many practical conditions: what is going to be measured, in what context questions will be asked, and among what type of respondents. Nevertheless, the preference is accorded to 7 point (Miller, 1956; Masters, 1974; Matell and Jacoby, 1971; Birkett, 1986; Wedell and Parducci, 1988; Alwin and Krosnick, 1991) fully verbally labeled scales (Alwin and Krosnick, 1991; Schwarz and Hippler, 1991) to reduce risks of measurement bias. The goal of this study is, however, to redirect the discussion toward a more fundamental and important issue. I want to focus analysis on bipolar answer scales, which are used to measure, explicitly and directly, constructs with two opposite sides, usually one positive and the other negative, and I will present some reasons to contemplate the application of branching, a two-step format, instead of the typical one-step strategy. With this goal, I will consider attitudes as one of the best example of bipolar constructs and I will offer some evidence of the effectiveness of using branching numerical scales for measuring attitudes of trust toward institutions.

Branching technique has been suggested because it simplifies the judgment task (Armstrong, Denniston, and Gordon, 1975) with its two steps, asking first the direction, e.g. “Are you satisfied or dissatisfied?,” and after the extremity, e.g. “How much?.” It has also been argued that is less burdensome for respondents to a telephone interview answer branching questions (Groves and Kahn, 1979; Miller, 1984). Another reason that has been proposed for using branching is its apparent ability to explore the midpoint information (Miller, 1984; Krosnick and Berent, 1993; Malhotra, Krosnick, and Thomas, 2009). In addition, the technique of branching bipolar questions takes advantage of the effectiveness of two-category scales to reliably capture the direction and also of longer scales to measure the extremity (Alwin, 1992).

Krosnick and Berent (1993) found that the branching format has higher validity and reliability than the non-branching format. I maintain that the effectiveness of branching is not mainly an effect of simplification. The hypothesis is that the principal cause of the higher validity and reliability of branching is due to the presence of a second step with positive labels either for positive or negative answers, resulting in the same unit of measurement. Saris and Gallhofer (2007a; 2007b) found that symmetric bipolar scales, with the same number of categories in each direction, were less valid and reliable than asymmetric scales, in which the negative part had fewer answer categories. Consequently, it must be highlighted that the power of discrimination of the positive side is different from the negative one because the distance between the categories in one side is not the same of that in the other side; accordingly, the two sides have two different units of measurement (see Figure 1).

Employing asymmetric scales, as Saris and Gallhofer suggested, is not the best solution, because the two directions do not have the same probability to be chosen. The bias resulting from the different way respondents give negative evaluations might be reduced by removing semantic reference to either the positive or negative direction (e.g. no reference to satisfaction or dissatisfaction), offering the same extremity answer categories as like “a little,” “quite a bit,” “a lot,” for either side. Thus, for example, categories like “a little satisfied,” “quite a bit satisfied,” “a lot satisfied,” “a little dissatisfied,” “quite a bit dissatisfied,” “a lot dissatisfied,” should not be used in a branching rating scale. In fact, branching gives us the possibility to put this into practice: in the second step it is not necessary to use negative or positive semantic elements.

To the best of my knowledge, branching has never been used with the just mentioned purpose. The reason of that could be seen in the fact that this idea contrasts the common sense, following which a graphic line, visual or hypothetical, from an extreme point to the opposite one is the best representation in order to help respondents express their position along an underlying bipolar dimension. It is probably for this embedding truism that branching technique has been only rarely and marginally applied.

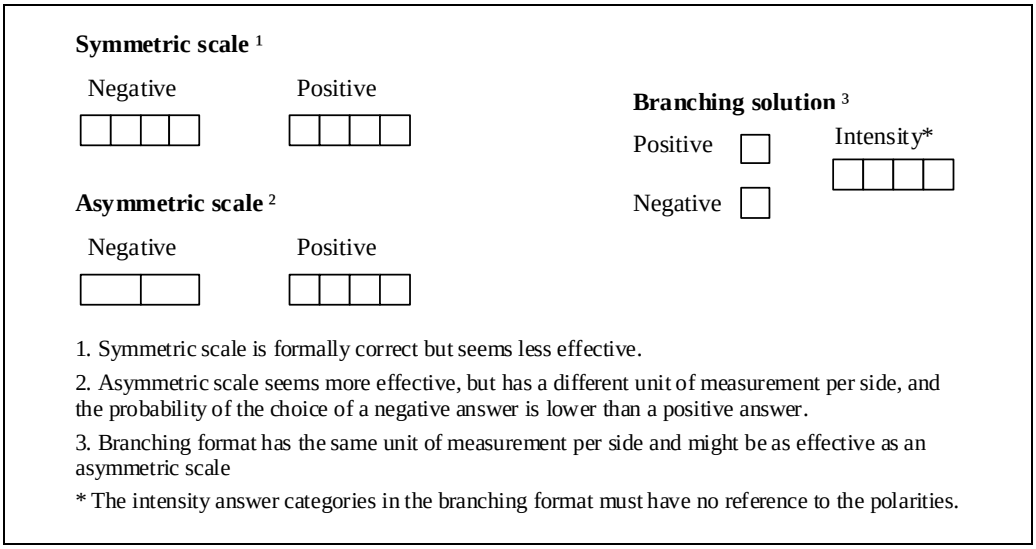

Figure 1: Branching format compared with symmetric and asymmetric scales

2. Numerical scales

Fully labeled numerical scales, as well as scales numerically labeled only at the extremities, are characterized by the presence of numbers to convey the ordinal sense of the categories. Numerical scales may be more accurate than verbally labeled scales, because they suggest the idea of equal distance between the categories (Krosnick and Fabrigar, 1997). The basic consequence is that respondents are free to assign a personal meaning, based on their personal assumptions, perceptions, aims, and values, to each non-verbally labeled point, taking into account only the equal distance between the points (Kilpatrick and Cantril, 1960). Therefore, this kind of rating scales could be more effective for cross-cultural research (Cantril and Free, 1962), and, because for numbers there are no problems of translation, the numerical format might be more effective for interlingual research too (Bernheim et al., 2006).

Despite these positive aspects, there are also some limits typical of the use of numerical scales. The first can be recognized in a possible cognitive divide among respondents: the abstract nature of numbers might be a problem for those respondents who are more comfortable to express evaluations with a verbal semantics and are not used to employ numerical categories for judgments. Secondly, the meaning of numbers may be context sensitive (Mohler, Smith, and Harkness, 1998), which means that people give different meanings to numbers in different cultural contexts (e.g. numerical preference, lucky and unlucky numbers, numerology, use of numerical scales at school or in other contexts).

2.1 Problems of unipolar numerical scales

For measuring bipolar constructs, like attitudes, researchers should use numerical scales with a bipolar format (e.g. -5 to +5), instead these constructs are usually measured with unipolar numerical scales presenting only a series of positive numbers (e.g. 1-10, 1-7, 3-7, 1-100, etc.), which is conceptually and formally wrong, and also conveys the confounding idea that in every point there is presence of the attribute (e.g. in a 1-10 scale of satisfaction, the numerical value 1, mathematically speaking, means the lowest level of satisfaction, and not complete dissatisfaction, as assumed by researchers who use this format for measuring satisfaction). Thus, the lower part of the scale has not negative values, but low values. In addition, zero means, in a unipolar numerical scale, absence of the attribute. Zero neither is a negative answer; it could be used only in scales measuring unipolar constructs, or at worst as a midpoint in a bipolar scale, but never as its lowest category. In survey practice, instead, zero is widely used for measuring bipolar constructs with unipolar numerical scales, without paying necessary attention to this caveat.

But the most problematic aspect is that respondent's interpretation of the scale points might be different from the meaning previously assigned by the researcher, so when these scales are applied for measuring bipolar constructs, they suffer of a measurement bias consisting in the fact that respondent's indifference point might not be in the middle of the scale (e.g. in a 0-10 scale a respondent might assign positive meaning only to the numerical values 7, 8, 9, 10, considering the remaining categories, from 0 to 6, negative answers). Furthermore, the respondent's indifference point might be different from those of other respondents. It is impossible to assume that the logical midpoint of the scale is the same of every respondent's midpoint, which instead is influenced by the respondent's personal use and interpretation of the scale.

2.2 Problems of bipolar numerical scales

Bipolar numerical scale, employing both positive and negative numbers, seems to be the most appropriate numerical format for bipolar questions. But this format suffers higher positivity bias than other scales. The positivity bias is a psychological propensity to express positive answers and it is explained by the preference of people to be positively oriented toward the others and to think, conversely, that also the others have this positive attitude toward them. Consequently, respondents are more reluctant to give negative evaluations. This effect is stronger when negative numbers are present in numerical scales (Tourangeau, Rips, and Rasinski, 2000), as it was empirically demonstrated by investigators who attempted a comparison between unipolar and bipolar numerical scales (Schwarz et al., 1991; O'Muircheartaigh et al., 1995). A remarkable refusal of expressing evaluations with negative numbers emerged from their experiments and that sort of refusal was even underestimated by the authors.

The studies of Schwarz et al. (1991) and O'Muircheartaigh et al. (1995) investigated the effects of negative numbers in rating scales, comparing bipolar numerical scales with unipolar numerical scales. In both studies the authors read the results seeing a difference of 10-15% in the negative answers between the formats. I argue, instead, that negative answers in the 0-10 scales, ranging from 0 to 5, were 25-30% (and not 10-15%) higher than the negative answers in the -5 to +5 scales, ranging from -5 to -1 (34% vs. 4% in Schwarz et al.; 64.5% vs. 37.9% in O'Muircheartaigh et al.). The authors underestimated this discrepancy because they didn't notice that the two formats were not properly comparable. In fact, they were using formats with 11 points each, but without taking into account the different meaning of zero in a bipolar or unipolar setting. So the scale from 0 to 10 has six negative and five positive answers, whereas the scale from -5 to +5 has five

negative, one neutral, and five positive answers. Schwarz and colleagues overestimated the negative answers in the bipolar scales (six vs. six points); because they treated the zero in the bipolar format as a negative answer, instead of an indifference point. In the study of O'Muircheartaigh and colleagues the negative answers of the unipolar scales were underestimated (five vs. five points), because the sixth category (number 5) in the unipolar format was wrongly treated as an indifference point, whereas in a 0-10 scale it represents a negative answer and despite its position in the middle of the scale it is not a logical midpoint.

There is another intervening factor besides the positivity bias. The higher positivity bias of bipolar numerical scales is here interpreted with the hypothesis of an interaction of a specific "avoidance effect," which is different from avoiding negative answers. People simply avoid minus numbers, because they don't use negative numbers in everyday life, in fact, negative numbers are used only within arithmetic operations, and also in these cases they don't have any autonomous semantics.

2.3 Branching format for numerical scales

For measuring bipolar constructs, numerical scales can be presented also with a third format, employing two steps: the branching format. In the first step respondents are asked to give the direction of their answer and in the second its extremity, using only positive numerical labels for both sides. Branching format has always been used with verbal labels only, so this could probably be the first attempt to propose branching scales also with numerical labels. Branching numerical scales maintain the positive characteristics of numerical scales (no ex-ante definition of the categories, and apparent equal distance between the categories) without the negative aspects typical of the one-step numerical scales either unipolar or bipolar.

Compared to the unipolar, the branching numerical format is (a) appropriate for measuring bipolar constructs, and (b) without the bias resulting from the researcher's misunderstanding of the answer direction, prevented by the first step asking that information directly, and consequently without misinterpretation of the extremity answers too. As a result, with the branching format the logical midpoint coincides with each respondent's midpoint. In addition, compared to the bipolar numerical format, the branching version (c) is not vulnerable to the respondents' avoidance of negative numbers, simply because there are no negative numbers in the second step.

Besides these advantages, the main problem of using the branching numerical format is the higher number of missing values along the two steps (Miller, 1984), but it could be prevented by ad-hoc strategies and with a gradual familiarization of respondents with this technique. Finally, I suggest to use the branching format with batteries of items so that the answering task is explained only once, allowing the data collection procedure to run faster.

3. Comparison of unipolar numerical scales with branching numerical scales

3.1 Experimental design

I carried out a split-ballot experiment aiming to compare unipolar numerical scales and branching numerical scales. The experiment was conducted during April-May 2009 and involved 146 members of 21 cultural and recreational associations using self-administered paper and pencil questionnaires, exploring the characteristics of their

membership, their system of values, attitudes toward institutions, political orientation, religious practice, and trust toward the future. The selected associations were representative of the universe of cultural and recreational associations in the metropolitan area of Rome (Italy) and the two subsamples obtained were both heterogeneous for sex, age, education, and socio-professional status. Only one manipulation was implemented for the experiment through the creation of two versions of the questionnaire, differing from each other for only the kind of rating scale used for measuring attitudes toward some institutions. Within each association respondents were randomly assigned to answer to the one or the other version of the questionnaire.

In synthesis, I compared two formats of a battery of items measuring trust toward 16 different institutions with the hypothesis of finding a significant difference in the results. The first format was constituted by 1-10 scales, using a typical version of the Cantril scale (n = 75), while the second format employed instead branching numerical scales (n = 71), which were presented with a dichotomous question in the first step, measuring the attitude direction (trust or distrust), and a 5-point scale in the second step, measuring the attitude extremity (from 1 to 5 for both directions).

The second hypothesis was that branching should be more valid because it is a bipolar format and because it has no problems in the interpretation of the answer direction. A Multitrait-Multimethod (MTMM) survey design was adopted to evaluate and compare the convergent and discriminant validity of the two formats. It was, therefore, introduced another instrument into the questionnaire, an inferential scale, with the aim of measuring, in a different way, the attitudes toward a limited amount of institutions that emerged as most representative of aggregates of institutions in a previous study conducted by Biorcio and Diani (1993) among members of associations in Lombardy (Northern Italy). The selected institutions were: Judiciary, Parliament, Catholic Church, and Banks. The second method chosen for measuring these 4 traits was the Guttman scale and this because of its complete difference from a rating scale: for each trait it was constructed a battery of 10 dichotomous items, each one presenting a sentence, evaluating an aspect of the considered institution, to be approved or disapproved. From each Guttman scale it was possible to reproduce the underlying construct, the attitude toward the institution, through an index that was obtained with an inferential procedure, elaborating the scale with the Rasch model (Rasch, 1960/1980).

3.2 Results

3.2.1 Answers comparison

The first observation that must be pointed out resulted from the data collection and was related to the ease of administration of the branching numerical scales. I found that, with the mode of the self-administration of questionnaires, 10 respondents out of 71 produced missing values in the second step, four of which gave extremity answers only when indicating a positive attitude direction in the first step, while no extremity answer was selected by them when their attitude was of distrust. On the whole, the difficulties of comprehension of the response task should not be considered a problem, because 60% of the respondents, helped only with written instructions, understood exactly how the technique works. Nevertheless, it must be acknowledged that people need to get accustomed to this new format of rating scale.

Looking at the comparison of the two formats, after a standardization of data into a 1-10 scale, I found that the average mean was 1 point lower for the unipolar scales (M = 4.6) compared to the branching scales (M = 5.6) (Table 1). That was in line with the

difference in the means resulting from the new interpretation of the experiments comparing unipolar and bipolar one-step scales (M = 6.4 vs. M = 7.8 in Schwarz et al.; M = 4.6 vs. M = 5.6 in O'Muircheartaigh et al.). However, it is particularly interesting to note that using another bipolar format, with two steps and without negative numbers, the avoidance of the negative side is not a problem anymore. The amount of negative answers reported with the unipolar numerical format (61%) was only 15% higher with respect to the bipolar branching format, which registered almost a half of the respondents (46%) not avoiding the negative side.

Table 1: ANOVA test and mean comparison of branching and Cantril scale

	Branching		*Cantril*		*Difference*			
Institutions	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*Sig.*
Parliament	5.0	2.75	3.6	2.23	1.39	.52	1.6	**.001**
European Union	6.8	2.49	5.5	2.22	1.35	.27	11.32	**.001**
Voluntary associations	8.6	1.66	7.5	2.23	1.09	-.57	1.55	**.001**
Trade unions	4.6	2.72	3.3	2.17	1.23	.55	8.78	**.004**
Catholic Church	5.9	3.28	4.5	3.09	1.46	.18	7.31	**.008**
Political Parties	3.6	2.51	2.6	1.88	1.01	.63	7.34	**.008**
Municipality of Rome	5.4	2.85	4.3	2.44	1.15	.41	6.51	**.012**
Associations of companies	5.0	2.74	3.9	2.35	1.09	.39	6.35	**.013**
Judiciary	6.0	2.83	4.9	2.65	1.15	.18	6.23	**.014**
Banks	4.0	2.55	3.0	2.18	.97	.37	5.92	**.016**
Environmental associations	6.8	2.58	5.8	2.57	.97	.01	4.95	**.028**
Army	6.8	2.87	5.8	3.04	1.04	-.17	4.34	**.039**
Media	5.1	2.71	4.2	2.10	.84	.61	4.24	**.041**
Government	4.4	2.88	3.5	2.54	.88	.34	3.64	**.058**
Lazio Region	5.0	2.81	4.3	2.28	.73	.53	2.93	**.089**
Police	7.2	2.87	6.5	2.85	.68	.02	1.97	**.163**
Average	**5.6**	**2.69**	**4.6**	**2.43**	**1.07**	**.26**	-	-

The analysis of the variance (ANOVA) showed that the difference between the unipolar and the branching scales was statistically significant, with a margin of error of 5%, for even 13 items out of 16 (Table 1). The means were always higher with the branching scales, so it can be noticed that there is a tendency to give less negative evaluations with the branching format.

In addition, the average standard deviation that was 0.26 higher with the two-step format is a cue that respondents tend to use a wider range of the scale with the branching format, especially for the less trustworthy institutions (Parliament, Trade unions, Political Parties, Associations of companies, Banks, and Government) (Table 1). With the unipolar format the intermediate category of the positive side (category 8) was used less, relatively to the positive answers, than with the branching format (22.4% of all positive answers with the unipolar format vs. 33.6%) and the second most extreme positive category (category 9) was quite unused (8.7% of all positive answers with the unipolar format vs. 19.8% with

branching), while the least extreme positive category (category 6) was highly used (32.7% of all positive answers with that format vs. 10.9% with branching). Both formats displayed a relatively high use of the most extreme negative category (36.4% of all unipolar negative answers vs. 34.9% with branching), but, considering the whole amount of answers per format, that phenomenon was more evident with the unipolar format (22.4% of all answers with the unipolar format vs. 15.5% with branching).

By means of the generalized trust index, i.e. the average score given by each respondent to 14 institutions (following the item analysis, trust toward Voluntary associations and Environmental associations were not included in this index), it is possible to see a strong central tendency with the unipolar format (see Figure 2) showing that respondents answering 1-10 scales tend to balance their positive and negative evaluations around a central position, in this case a low generalized distrust (category 5). With branching numerical scales results are completely different: the generalized distrust is less marked, and respondents are less reluctant to express on average an intermediate or high positive attitude toward institutions.

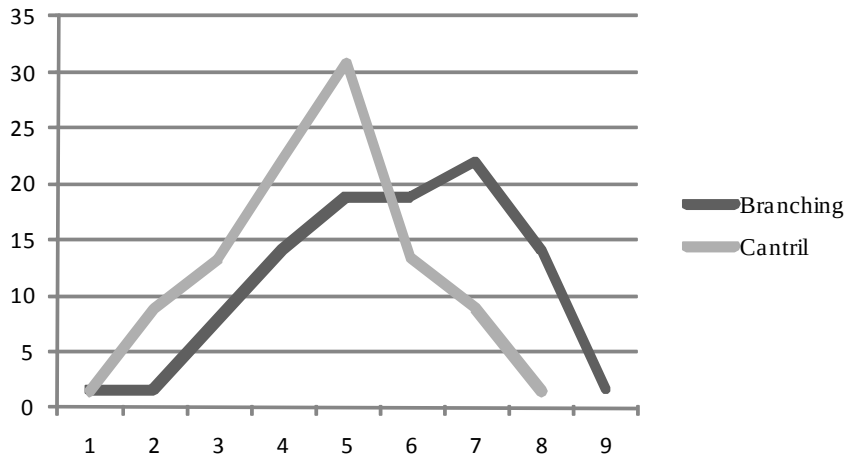

Figure 2: Distribution of generalized trust toward institutions (average score of 14 items)

A plain evidence of a method effect emerged from all these comparisons, individuating a significant difference in the answer configurations when using unipolar or branching numerical scales. As the first hypothesis has been confirmed, I will try now to confirm or disconfirm also the second hypothesis, i.e. that branching numerical scales are more valid for the reason that in the unipolar format there is a systematic error, preventing the correspondence of its results with the latent construct.

3.2.2 Convergent and discriminant Validity

Before proceeding with the assessment of the convergent and discriminant validity of the four selected items, it is important to control the construct validity of the whole battery of items. Exploring the correlations of the generalized trust, it was found a significant correlation with the political orientation, which was the best predictor indicator of this attitude: the higher correlation was individuated with the index obtained from branching scales ($r = .448$; $p < .001$, two tailed), while with unipolar scales this correlation was less evident ($r = .293$; $p < .02$, two tailed). The generalized trust resulting from unipolar scales

had instead a higher correlation with a factor of distortion for the technique of attitude measurement: the respondents' age (r = .346; p < .003, two tailed). While with branching format there was not a significant correlation with that variable (r = .154; p < .22, two tailed). Thus, with 1-10 scales the elderly tended to use only the positive side, while young respondents tended to use only the negative side.

The positive correlations between the numerical scales of trust toward Judiciary, Catholic Church, Banks, and Parliament and the Guttman scales measuring the same traits (see the two main diagonals in Table 2) were always higher with the branching format than with the unipolar format. This is indicative of higher convergent validity for the branching scales.

Considering the intra-method discriminant validity (Numerical x Numerical), 4 correlations out of 6 were lower with branching (Ch x Jud; Ban x Ch; Parl x Ch; Parl x Ban) (Table 2). Looking, instead, at the inter-method discriminant validity (Numerical x Guttman) 9 correlations out of 12 were lower with branching (Jud x Ch; Jud x Ban; Ch x Jud; Ch x Ban; Ch x Parl; Ban x Ch; Ban x Parl; Parl x Ch; Parl x Ban). Therefore, also the discriminant validity was higher with the branching format.

Table 2: MTMM Matrix of Guttman and numerical scales

Guttman									
Jud.	Branch.	-							
	Cantril	-							
Ch.	Branch.	-.235*	-						
	Cantril	-.232*	-						
Ban.	Branch.	.196	.252*	-					
	Cantril	-.083	.235*	-					
Parl.	Branch.	.288*	.347**	.449**	-				
	Cantril	-.255*	.348**	.234*	-				
Numerical									
Jud.	Branch.	**.606****	-.248*	.023	.028	-			
	Cantril	.552**	.061	.026	-.120	-			
Ch.	Branch.	-.320*	**.779****	.102	.207	-.095	-		
	Cantril	-.096	.697**	.177	.285*	.224*	-		
Ban.	Branch.	.079	.134	**.368****	.139	.271*	.308*	-	
	Cantril	-.017	.443**	.212	.179	.230*	.530**	-	
Parl.	Branch.	.161	.336*.	.151	**.430****	.253*	.373**	.343**	-
	Cantril	-.051	.410**	.257*	.330**	.209	.492**	.592**	-
		Jud.	Ch.	Ban.	Parl.	Jud.	Ch.	Ban.	Parl.
		Guttman				*Numerical*			

*p < .05. **p < .01.

Through a confirmatory factor analysis (CFA), it is possible to reach a more accurate interpretation of the MTMM matrix. With the CFA model both traits and methods are considered latent factors, individuating parameters between traits (the lower the better) and between methods (more approximate to zero is better). It is also possible to see the contribution of each single variable to its latent trait, i.e. the convergent validity (the higher are the contributions of the two related variables, the better), and to the variance due to the method (the lower the better).

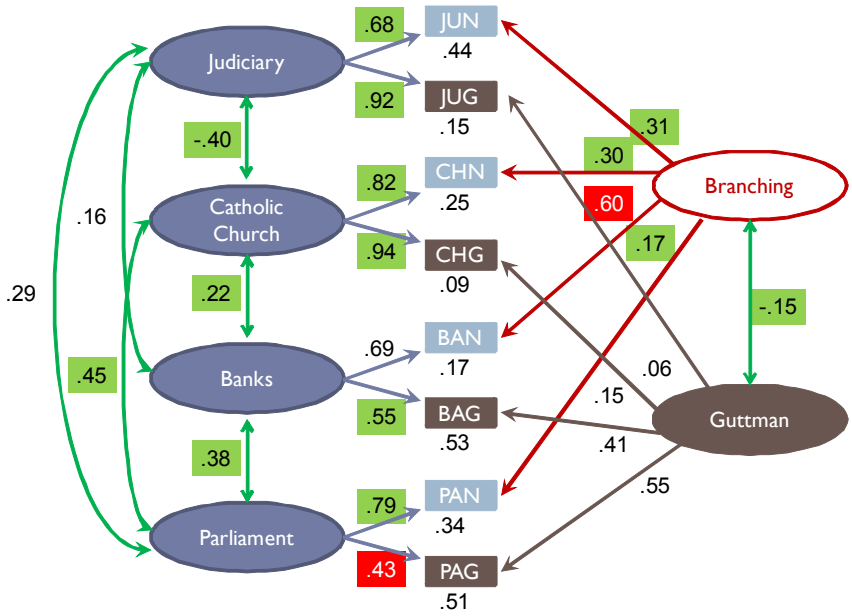

Figure 3: CFA Model of the MTMM with branching numerical scales

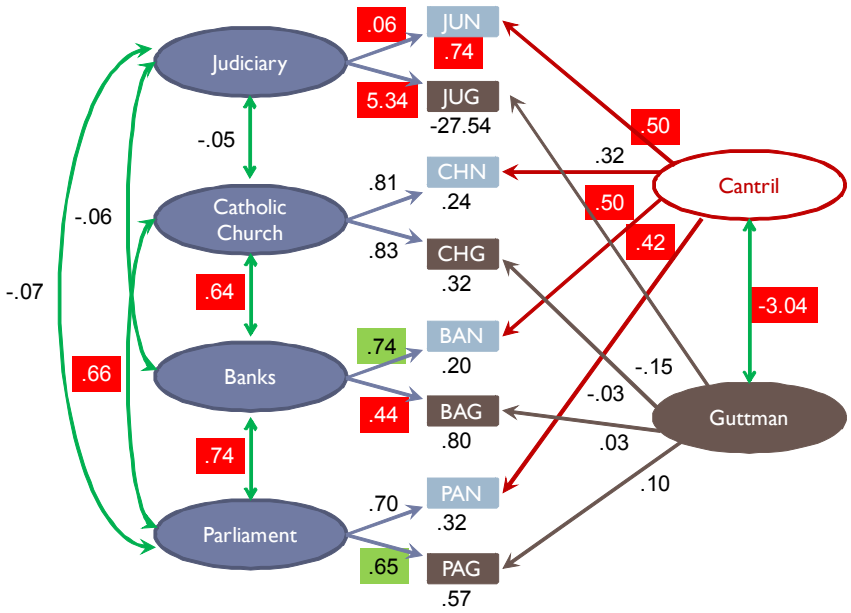

Figure 4: CFA Model of the MTMM with Cantril scales

The discrimination between traits was plainly higher for the numerical scales with the branching format (see Figure 3 and 4), where attitude toward Judiciary, as emerged also from a principal component analysis, resulted more distant from the other traits. The discrimination between methods was less distant from zero in the branching format, indicating similarity in the results generated by the rating scale and the Guttman scale. Even with this analysis, the convergent validity resulted higher with the two-step numerical scales. Finally, the errors due to the technique effect were lower with the branching scales than with the unipolar scales.

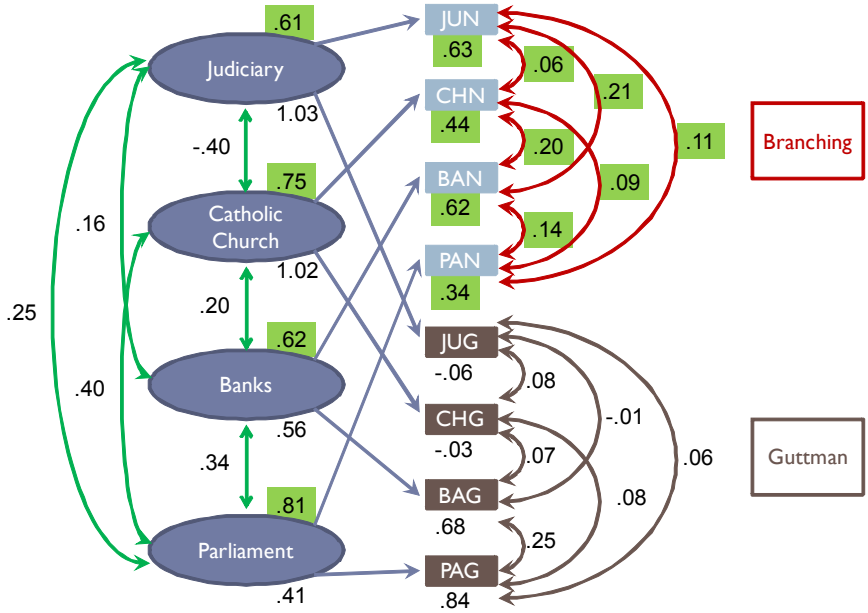

Figure 5: CU Model of the MTMM with branching numerical scales

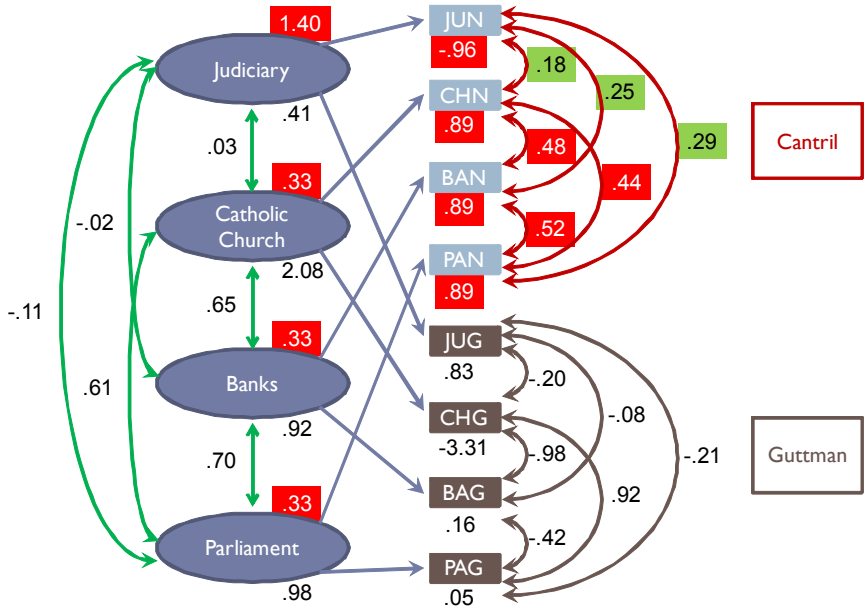

Figure 6: CU Model of the MTMM with Cantril scales

Since CFA has the limit of interpreting the method effects as one-dimensional and risks confusing the variance due to the traits with the variance due to the methods, especially when traits are highly correlated with each other, the MTMM was analyzed also with the Correlated Uniquenesses Model (CU) (Kenny, 1979; Marsh, 1989). The difference is that now the methods are not considered latent factors and that their effects are interpreted with the correlations between the uniquenesses of those variables using the same method. So the attention could be directed only to the convergent validity and the internal errors within the two formats of numerical scales.

The latter comparison offered another confirmation of the higher convergent validity obtained with the branching numerical format (see Figure 5 and 6). All saturation trait-variable were particularly high with branching, while extremely low with the unipolar format (Judiciary's parameter was even out of scale). The intra-method discriminant validity, also using this model, was higher with the branching format.

According to the MTMM analysis carried out, it is possible to affirm that branching numerical scales have the capability to better reproduce a latent bipolar construct than conventional Cantril scales. Branching scales are more valid not only because they are constructed as bipolar, but also because they don't have the problem of the discrimination of the answer direction.

Conclusions

The effectiveness of the use of bipolar numerical scales with two steps for measuring attitudes was empirically confirmed. As a result, it seems evident that branching could be an optimal solution for the problems of both unipolar and bipolar numerical scales for measuring bipolar constructs. Accordingly, I foster the use of the branching numerical format in all those situations where verbal labeling is not recommended, and I suggest presenting branching questions in batteries of items, so that the comprehension of the response task could be easier for respondents, especially considering that they are not familiarized with numerical scales employing two steps.

Further research is needed, first of all, to replicate this study in other countries, measuring other bipolar constructs, and using other modes of data collection, and, secondly, to assess whether this new numerical format is less valid and reliable than the verbally labeled format, confirming the results obtained with the unipolar numerical format, or not. This evaluation could be designed comparing branching numerical scales either with branching verbally labeled scales or with Likert-type scales.

Finally, I remind researchers to avoid the zero in the unipolar format when comparing unipolar numerical scales with one-step bipolar numerical scales, or with branching numerical scales for testing the unidimensionality error.

Appendix

Table 1: Question format of the branching numerical scales (First 2 items of the original version)

Le presentiamo ora alcune istituzioni. Facendo riferimento all'attuale contesto italiano, indichi per ciascuna di esse se vi ripone fiducia oppure no. Scelga una risposta e poi

specifichi quanto è forte questo suo sentimento di fiducia o di sfiducia dando un voto compreso fra 1 (il minimo di fiducia o di sfiducia) **e 5** (il massimo di fiducia o di sfiducia).
(per ogni riga faccia prima una scelta e poi indichi l'intensità di questa scelta)

		Prova fiducia?	Quanto è forte questo suo sentimento?
a.	Forze armate	□ Sì, ho fiducia □ No, ho sfiducia	1 2 3 4 5
b.	Magistratura	□ Sì, ho fiducia □ No, ho sfiducia	1 2 3 4 5

Table 2. Dichotomous items used for the four Guttman scales (Language: Italian)

Judiciary

a. Ci sono troppi magistrati che utilizzano il loro potere per interferire nella politica **(-)**
b. La Magistratura a volte interpreta le leggi in modo poco corretto **(-)**
c. La Magistratura spesso non è in grado di individuare chi è colpevole e chi innocente **(-)**
d. La Magistratura infligge per lo più pene giuste **(+)**
e. I vari organi della Magistratura garantiscono il buon funzionamento della giustizia **(+)**
f. La Magistratura viene troppo attaccata della politica **(+)**
g. La Magistratura spesso non riesce a punire i colpevoli, che evitano facilmente la condanna **(-)**
h. Con il suo operato la Magistratura favorisce la certezza della pena **(+)**
i. Con il suo operato la Magistratura rallenta il corso della giustizia **(-)**
l. Molti magistrati sono una specie di eroi perché mettono a rischio la loro vita per perseguire la criminalità **(+)**

Parliament

a. Il Parlamento si sofferma troppo sulla discussione di argomenti poco importanti **(-)**
b. L'attività parlamentare è seriamente rallentata dai continui passaggi di legge tra Camera dei deputati e Senato **(-)**
c. La lentezza del Parlamento giustifica il ricorso da parte del Governo ai decreti legge **(-)**
d. I passaggi di legge tra Camera e Senato sono essenziali, perché garantiscono il perfezionamento delle leggi attraverso una riflessione ragionata e plurale **(+)**
e. Essendoci in Italia molte leggi, il Parlamento potrebbe dedicarsi soprattutto ad armonizzare la legislazione già esistente **(+)**
f. I lavori parlamentari sono ostacolati dalle eccessive pressioni esercitate dal Governo **(+)**
g. Il Parlamento sta approvando leggi che favoriscono il buon funzionamento del Paese **(+)**
h. Ci sono troppi parlamentari che non pensano affatto agli interessi della gente **(-)**
i. Il Parlamento sta facendo troppe leggi che tutelano solo gli interessi dei potenti **(-)**
l. Molti parlamentari si impegnano ad accogliere le aspettative dei loro elettori **(+)**

Catholic Church

a. E' un male che le richieste della Chiesa cattolica vengano limitate dalla politica **(+)**
b. L'influenza della Chiesa cattolica rende il Paese più arretrato e conservatore **(-)**
c. La Chiesa cattolica è troppo lontana dal paese reale, fraintendendo così i suoi problemi più urgenti **(-)**
d. Nella Chiesa cattolica ci sono troppi preti di dubbia moralità **(-)**
e. La struttura della Chiesa cattolica garantisce che l'operato dei suoi membri segua sempre i principi di solidarietà **(+)**
f. La Chiesa cattolica dà il giusto sostegno a tutti coloro che ne hanno bisogno **(+)**
g. La Chiesa cattolica ha una struttura così rigida da non riuscire ad aiutare sufficientemente il

Paese **(-)**

h. La Chiesa cattolica rende il paese più civile **(+)**
i. La Chiesa cattolica è troppo legata ad aspetti materiali della vita **(-)**
l. I preti sono delle figure fondamentali nel dare forza alla gente **(+)**

Banks

a. Le Banche andrebbero eliminate, perché non tutelano in alcun modo il valore del denaro **(-)**
b. Le Banche hanno regolamenti che riescono a garantire la tutela dei clienti **(+)**
c. Il personale di banca cerca spesso di comprendere le esigenze del cliente e di venirgli incontro in qualche modo **(+)**
d. Le Banche investono il denaro dei risparmiatori in fondi e azioni troppo rischiosi **(-)**
e. Le Banche cercano di essere abbastanza chiare e trasparenti nel comunicare condizioni e costi dei loro prodotti **(+)**
f. Le Banche hanno regole troppo rigide che impediscono di assecondare le richieste dei clienti **(-)**
g. Le Banche danno priorità agli interessi dei gruppi forti di potere, perché costrette da interferenze esterne **(+)**
h. Le Banche svolgono bene il loro compito di far circolare la ricchezza **(+)**
i. Le Banche concedono pochi prestiti e così facendo soffocano la crescita economica del Paese **(-)**
l. Sono troppe le persone che lavorano in banca che si relazionano al cliente in modo impersonale e burocratico **(-)**

References

Alwin D.F. & Krosnick J.A. (1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods Research, 20 (1),* 139-181.

Alwin D.F. (1992). Information transmission in the survey interview: Number of response categories and the reliability of attitude measurement. *Sociological Methodology, 22,* 83-118.

Armstrong S.J., Denniston W.B., & Gordon M.M. (1975). The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance, 14,* 257-263.

Bernheim J.L., Theuns P., Mazaheri M., Hofmans J., Fliege H., & Rose M. (2006). The potential of Anamnestic Comparative Self-Assessment (ACSA) to reduce bias in the measurement of subjective well-being. *Journal of Happiness Studies, 7,* 227-250.

Biorcio R., & Diani M. (1993). *Primo rapporto sull'associazionismo in Lombardia*. Milano, Italy: Regione Lombardia.

Birkett N.J. (1986). Selecting the number of response categories for a Likert-type scale. *Proceedings of the American Statistical Association 1987 Annual Meeting,* Section on Survey Research Methods, 488-492.

Cantril H., & Free L.A. (1962). Hopes and fears for self and country: The self-anchoring striving scale in cross-cultural research. *American Behavioral Scientist, 6 (1),* supplement.

Fishbein M.E., & Ajzen I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Groves R.M. & Kahn R.L. (1979). *Surveys by telephone*. New York: Wiley.

Guttman L. (1950). The basis for scalogram analysis. In S.A. Stouffer et al. *Measurement and Prediction*. New York: Wiley.

Kenny D.A. (1979). *Correlation and causality*. New York: Wiley.

Kilpatrick F.P. & Cantril H. (1960). Self anchoring scaling: A measure of individuals' unique reality worlds. *Journal of Individual Psychology, 16 (2),* 158-173.

Krosnick J.A. & Berent M.K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science, 37 (3),* 941-964.

Krosnick J.A. & Fabrigar L.R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg et al. (Eds.), *Survey measurement and process quality* (pp. 141-164). New York: Wiley-Interscience.

Krosnick J.A. (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236.

Malhotra N., Krosnick J.A., & Thomas R.K. (2009). Optimal design of branching questions to measure bipolar constructs. *Public Opinion Quarterly, 73 (2),* 304-324.

Marsh H.W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*, 335–361.

Masters J.R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement, 11*, 49-53.

Matell M.S. & Jacoby J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement, 31*, 657-674.

Miller G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review 63 (2),* 343-355.

Miller P.V. (1984). Alternative question forms for attitude scale questions in telephone interviews. *Public Opinion Quarterly, 48*, 766-778.

Mohler P.P., Smith T.W., & Harkness J.A. (1998). Respondent's ratings of expressions from response scales: A two-country, two-language investigation on equivalence and translation. In J.A. Harkness (Ed.), *Nachrichten Spezial, Cross-Cultural Survey Equivalence 3* (pp. 159-184). Mannheim: ZUMA.

O'Muircheartaigh C., Gaskell G., & Wright D.B. (1995). Weighing anchors: Verbal and numeric labels for response scales. *Journal of Official Statistics, 11 (3),* 295-307.

Rasch G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research (Expanded edition, 1980. Chicago: University of Chicago Press).

Saris W. & Gallhofer I. (2007a). *Design, evaluation and analysis of questionnaires or survey research*. New York: Wiley.

Saris W. & Gallhofer I. (2007b). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods, 1 (1),* 29-43.

Schaeffer N.C., & Presser S. (2003). The science of asking questions. *Annual Review of Sociology, 29*, 65-88.

Schwarz N. & Hippler H.J. (1991). Response alternatives: The impact of their choice and presentation order. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (41-56). New York: John Wiley.

Schwarz N., Knauper B., Hippler H.J., Noelle-Neumann E., & Clark L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55*, 570-582.

Thurstone L.L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology, 26 (3),* 249-269.

Tourangeau R., Rips L.J., & Rasinski K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Wedell D.H. & Parducci A. (1988). The category effect in social judgment: Experimental ratings of happiness. *Journal of Personality and Social Psychology, 55*, 341-356.