Advances in Methodology, Data Analysis, and Statistics
Anuška Ferligoj (Editor), Metodološki zvezki, 14
Ljubljana: FDV, 1998

On the Identification of Acquiescence in Balanced Sets of Items Using Structural Models

Jacques B. Billiet and McKee J. McClendon*

Abstract

Attitude surveys often use sets of items with identical response scales in order to construct scales consisting of multiple indicators of attitude constructs. These response scales are often written in a Likert format in which respondents are asked how strongly they agree or disagree with each attitude statement. Among sociologists, there is considerable evidence that such a response format can be susceptible to an agreeing-response bias called acquiescence. Several methods have been proposed for controlling for acquiescence, including forced choice items, split ballots, and balanced scales. In an important paper, Mirowsky and Ross (1991) showed how to specify structural models for the measurement and control of acquiescence in latent variable models. In this paper, a number of conditions are specified that must be met before concluding about the existence of the agreement component of the response style called "acquiescence". These conditions deal with hypotheses concerning the identification of a style factor within the context of structural models. The focus here is on the specification of a common style factor behind at least two independent theoretical concepts, each measured by a (quasi-) balanced set of item. The model is explored using a small random sample from the general population (N = 188) at the occasion of pilot interviews. It is argued that a model with only two content factors and a style factor represents the data better than a model with two content factors. The proposed model is then confirmed in two subsamples of a large scale survey (N = 2,100) in the same population.

---

* Department of Sociology, University of Leuven, Edward Van Evenstraat 2B, B-300 Leuven, Belgium.

This research is part of a research project on the reliability and validity of scales in surveys, and is supported by the Fund for Scientific Research-Flanders and by the Research Fund of KU Leuven (OT/93/4). The authors wish to thank Marc Swyngedouw, An Carton and Roeland Beerten of the Inter-University Center for Political Opinion Research (ISPO) for their support in the data collection. We are also grateful to Willem Saris, Dagmar Krebs, the other members of the International Research Group on Methodology and Comparative Survey Research (IRMCS), and two anonymous referees for their suggestions.

1 Introduction

Attitude surveys often use sets of items with identical response scales in order to construct scales consisting of multiple indicators of attitude constructs. These response scales are often written in a Likert format in which respondents are asked how strongly they agree or disagree with each attitude statement. There is considerable evidence that such a response format can be susceptible to an agreeing-response bias called acquiescence (McClendon 1991a). The arguments of Rorer (1965) and many of his colleagues in psychometrics (e.g. Nunnally 1967) who deny the relevance of the agreeing response style in both personality research and measurement theory, have been countered by many others (see: Bentler et al. 1991: 187). Among researchers which are involved in research on method effects in attitude measurement, the agreeing response style continued to receive considerable attention (e.g. Schuman and Presser 1981; Ray 1983, 1984a, 1984b, 1990; McClendon 1991a, 1991b, 1992; Mirowsky and Ross 1991; Toner 1987). Some scholars proposed the use of forced-choice items to overcome the acquiescent response bias (e.g. Schuman and Presser 1981: 207; Toner 1987). Others argued that the advice of using forced-choice items overlooked several serious problems (Ray 1990) and proposed balanced sets of items in Likert format ("balanced scales") in order to correct for acquiescence (Martin 1964; Nunnally 1967; Cloud and Vaughn 1970; Ray 1979; Paulhus 1991; Spector 1992).

In a balanced scale half of the items are worded in a positive direction and half in a negative direction toward the attitude object. The responses to the negative items are in reverse order coded and then added to the responses of the positive items to give a single score as a measure of the attitude construct. Acquiescence to the negative items will offset acquiescence to the positive items and consequently the mean of the response distribution will not be biased. Moreover, the covariance between the balanced scale and another variable will not be biased by acquiescence to both the positive and the negative items. However, the relatively simple notion justifying balanced sets of items is that acquiescence to the positively worded items will be canceled out by acquiescence to the negative items (McClendon 1992). In other words, it is assumed that the negative and positive worded items are equally susceptible to acquiescence. McClendon (1992) showed that this assumption does not necessarily hold.

In an important paper on the elimination of defense and agreement bias from measures of the Sense of Control, Mirowsky and Ross (1991) showed how an agreeing-response bias will inflate the variance and reliability estimates for an unbalanced scale and lead to either over-estimates or under-estimates of the relationship between the construct, being measured by the unbalanced scale, and other constructs. Moreover, they show how to specify structural models for the measurement and control of acquiescence in latent variable models. Watson (1992) applied the structural equation modeling technique in order to extract the covariance that is due to acquiescence in an unbalanced set of items. In the absence of a balanced set of items, she used seven control items in order to construct an acquiescence scale. Marsh (1996) also used confirmatory factor analysis to evaluate whether factors associated with positively and negatively worded items are substantively meaninfgul or artifactors.

In this study, we will try to validate the measurement of acquiescence in balanced scales using a structural modeling approach. Following Bentler et al. (1971) several hypotheses about the agreeing style factor are investigated in order to answer questions regarding the measurement of acquiescence. Others who tried to model acquiescence with the structural equations approach were confronted with the phenomenon that models with a positive and a negative factor were just as likely to describe the data as models with a bipolar factor and a style factor (Mirowsky and Ross 1991; McClendon 1992). Here we met a well known and often debated phenomenon in psychometrics and sociometrics: indicators that were meant to operationalize one (bipolar) dimension are often found to represent two independent factors (Marsh 1996: 810).

If opposite worded statements are used as indicators for one concept, it often appears that two dimensions are found underlying the theoretical construct, one for the positive set of items and one for the negative set of items (e.g. McClendon 1992). This happened for example in studies of global self-esteem (Marsh 1992), of mood state (Lorr and Shea 1979), self-descriptive adjectives like "happy" and "sad" (Diener and Emmons 1985), and political preferences (Weisberg 1970). Some researchers support a substantial explanation of this phenomenon saying that two independent factors is a correct finding while others promote a methodological explanation (see Marsh 1996).

According to van Schuur and Kiers (1994: 97), the identification of two factors, when one factor is expected, is an artifact caused by using factor analysis on data that would be more appropriately analyzed with a unidimensional unfolding model. According to these authors, the inappropriate specification of a linear model behind factor analysis is responsible for two factors. They suggest a number of diagnostics that specify necessary but not sufficient conditions for the unfolding model: double-centering of the data matrix; inspecting the form of the matrix, evaluating the signs of the partial correlations; and principal component analysis of the centered factor loading matrix (van Schuur and Kiers 1994: 101-103).

Other methodological explanations were suggested by Green (1988), Hamilton (1968) and Bentler (1969). They explain the phenomenon as response bias or as a style effect (acquiescence). Saris (1988) suggested an interpretation in terms of variation in individual response functions. The present study supports this methodological interpretation in terms of a response style effect (acquiescence) that is observed when identical response scales in a Likert format are used for several sets of indicators.

2 A response style called "acquiescence"

Cronbach (1942, 1946, 1950) called the tendency to agree with statements or questions, independent of their content a "response set" which he defined as the disposition of a respondent to respond systematically to questions in another way than he would if the questions were offered in another form. A large number of empirical studies supported the existence of these response biases. Jackson and Messick (1958) emphasized the importance of these biases by highlighting a distinction between *content* and *style*. They ascribed the major common factors in personality inventories of the true-false or agree-disagree type to style rather than specific item content (Jackson and Messick 1958: 247). According to these scholars, a response style is a potentially measurable personality variable or trait. It refers to a behavioral consistency operating across measures of several conceptually distinct content traits (Jackson and Messick 1962: 134).

Rorer (1965), who examined a large number of studies dealing with acquiescence, arrived at a radically different conclusion. He concluded that there has never been any confirmatory evidence for the hypothesis that subjects will not respond consistently to the content when that content is presented in more than one form" (Rorer 1965: 151-152). It is important to state that Rorer used the term "style" in a restrictive way as a tendency to select some response category a disproportionate amount of time independently of the item content, excluding several other types of response sets such as social desirability (Rorer 1965: 134). He argues that "styles, like sets, have been inconsistently defined; designations include such terms as yeasaying, naysaying, and extreme position response bias" (Rorer 1965: 151). According to Rorer, "acquiescence has been conceptualized as a generalized tendency to be agreeable (a set), but has been operationally defined in terms of a disproportionate tendency to select a certain response category (a style)" (Rorer 1965: 151). Because measures of response styles fail to correlate across tests, Rorer hypothesized that these styles are test specific (Rorer 1965: 151).

Rorer's position was followed by some authors of leading books in psychological measurement (Block 1965; Nunnally 1967). This led Schuman and Presser (1981: 204-206) to conclude that a dominant stream in psychometrics declared acquiescence to be in the domain of fantasy. As was already mentioned, the survey researchers (and social psychologists) continued to pay attention to the agreement response bias. One of the reasons for this discrepancy in the views of psychologists and sociologists may be the kind of samples they use in their research. In psychology, empirical research often makes use of homogeneous students populations and written questionnaires. One can argue that it must be very hard to detect acquiescence in such a context since it may vary with respondent characteristics as age and level of education. In sociology, more heterogeneous samples and personal face-to-face interviews are used (Schuman and Presser 1981: 204-206).

One of Rorer's arguments for denying acquiescence was the absence of appreciable correlations between various scales which were built to measure acquiescence. Others have argued that the low correlations suggest that acquiescence to questionnaire and test items is not caused by an unitary personality trait. Martin (1964) and Ray (1983) distinguished at least two types of acquiescence, acquiescence to social aphorisms (e.g. items in the F-Scale or in scales about ethnocentrism) and acquiescence to personality inventories. In a well known study, Bentler et al. (1971) admitted the existence of style effects, but they formulated a number of empirical hypotheses (conditions) that must be tested before concluding that a particular style effect has been detected. Furthermore, Bentler et al. offered empirical evidence from psychological tests that acquiescence was best understood as composed of at least two different processes, rather than one: *agreement* and *acceptance* acquiescence (Bentler et al. 1971: 187, 190). We will not elaborate on this argument, since the component of acceptance applies to statements about characteristics that are conceived as self-descriptive (true/false personality inventories) and not to social aphorisms (agreement acquiescence) which are used in survey research.

Insofar a style refers to behavior consistency across measures of distinct traits, it can be identified by the existence of a latent variable which is usually, but not necessarily, a common factor in linear analysis. A style may be manifest in responses to stimuli were the content is presented in various formats, or by various methods as far as these formats or methods have some sensitivity to a particular response style in common (e.g. in case of agreement acquiescence: yes/no questions or agree/disagree statements as opposed to forced-choice questions). The degree of generality of any particular style is an empirical question, but conceptually we may expect that a common style effect can be detected in measures of, for example, abstract issues such as ethnic prejudice or distrust in politics, when measured by certain kinds of methods. In so far as a style is a personality variable, it will ordinarily exhibit correlations with other behavioral consistencies or content traits and it will attenuate the discriminant validity of personality scales. Further, style and content are not mutually exclusive. This means that the identification of stylistic variance does not necessarily contradict the presence of content (Bentler et al. 1971: 188-189).

Following this way of reasoning, the following hypotheses apply to the agreement component of acquiescence:

1. acquiescence can be identified as a common factor behind a set of agree/disagree items that are semantically balanced (half of the items are worded in the opposite direction of the other items with respect to a general construct);
2. the common style factor will have a non-zero variance which is anyhow smaller than the variance of the content factor;
3. the common style factor can be found in two or more balanced sets of indicators measuring two or more independent constructs;

4. within a sample of respondents, a style factor which is interpreted as agreement acquiescence must be apparent in a number of respondents agreeing with both, negative and positive statements that are conceived as (quasi) opposite in meaning with respect to the general construct;
5. finally, but very hard to assess, as a personality trait agreement acquiescence should be consistent over time when the same measurements of the same traits are used among the same respondents.

Those who do not consider acquiescence a personality trait or style will certainly not accept the last statement regarding the stability over respondents (e.g. Ray 1983). They presumably also will not accept the propositions about the common style factor behind the measurements of two or more different constructs suggesting that acquiescence is content specific and not entirely general. Ray (1983) for example found that acquiescence scales (the sum of agreements to the items in a balanced scale) for different constructs were not so highly correlated. It is argued that acquiescence depends of the ambiguity of the concepts. However, we know of only one study that supports ambiguity as the cause of acquiescence (McBride and Moran 1967). Ambiguity implies that the respondents that are likely to agree with (nearly) all items of one balanced scale may not be identical as those who agree with the items of another balanced scale. Content that is ambiguous to some respondents may be relative clear to others since ambiguity may have something to do with how much one knows about the topic (McClendon 1991; Krosnick 1991). According to Krosnick (1991), acquiescence and other response effects are produced by satisfying, which in turn is caused by high task difficulty, low motivation, and low ability (related with the amount of knowledge about the topic).

We are not aware of panel studies in which sound empirical arguments can be found for sustaining the fifth proposition. Schuman and Presser (1981: 207-212) reported a repeated split ballot study but it was not possible to evaluate the stability of acquiescent respondents since balanced scales were not used. There are some indications of acquiescence in split ballot studies with both yes/no items and agree/disagree statements (Schuman and Presser 1991; McClendon 1991a). In the present study we do not have panel data nor split ballots but we will try to identify the style factor of acquiescence by investigating the propositions 1 to 4 with two sets of balanced items measuring two independent constructs.

3 Method and data

In a previous study, several of the mentioned propositions were explored with sets of positively and negatively worded items about ethnic prejudice. The data were collected by face-to-face interviews in 1989 in random samples of 664 Flemish respondents, 518 Walloon respondents and 418 respondents from Brussels, all between 18 and 75 years of age. Four negatively worded and three positively worded statements about feelings of threat against immigrants were selected (Billiet et al. 1990). It was not possible to select couples of pure reversals but some of the statements were clearly contradictory in meaning with respect to the concept 'feeling threatened by immigrants' (Moroccans or Turks). About 10% of the respondents agreed with pairs of quasi-contradictory items in each of the three samples. Eight percent of the respondents agreed with at least three negatively and two positively worded items (hypothesis 4). It was shown that the Mirowsky and Ross (1991) approach could be applied to the balanced set of threat items. A model with a content factor and a style factor was fitted and compared with alternative models (Billiet 1995b). However, it was not possible to test the proposition about two or more balanced sets of items because the 1989 surveys did not contain balanced sets of items for other concepts than ethnic prejudice.

3.1 Data

In the present study, we will use two balanced sets of items that were constructed in view of the identification of acquiescence. The scales were tested in a pilot study with a random sample of 188 Flemish and French speaking Belgians in June 1995, and after evaluation used in the 1995 general survey on political attitudes with a random sample of 2,100 Flemish respondents (October 1995-February 1996). Both studies used face-to-face interviews[1].

The questionnaires contained two quasi balanced sets of 13 Likert items about ethnic prejudice and political efficacy. The response scales were all 5-point (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree). In the analysis, the scores were reversed (strongly agree = 5; strongly disagree = 1). It was possible to select several subsets of balanced scales (equal numbers of positively and negatively worded items with respect to the attitude object)[2]. In this study, the following subset is used since in our view it contains the clearest (quasi) contradictions: 'not to be trusted' - 'welcome'; 'endanger for jobs'- 'contribute to prosperity'; 'cultural threat' -'cultural enrichment'; 'no ability to listen' - 'take our views into account'; 'feel too good' - 'able people'. The positively worded items are not reversals of the negatively worded items, and it is possible to agree with some of

[1] The survey was conducted by the Inter-University Center for Political Opinion Research (ISPO). Most of the respondents (80%) are second-wave panel respondents. The first wave and the new samples are two-stage samples with equal probabilities. In the first stage, the municipalities were selected at random. About 120 Flemish villages out of 316 were included in the sample. In the second stage, a random sample of respondents was selected from the national population registers. The response rates (non contacts included) of the panel respondents and those in the original sample were 70% and 65%.

[2] The proposed models apply to different subsets of items, even when more negatively than positively worded items were used, as long as they belong to one dimension of the theoretical construct.

the positively and some of the negatively worded items. However it is assumed that we are able to identify a part of the covariance which can be ascribed to a style factor (acquiescence) with these two sets of items.

In the general survey on political attitudes about 7.5% of the respondents agreed with seven or more items; 3% agreed with eight or more (hypothesis 4). The two subsets of items were worded in exactly the same way in both the pilot study and the general survey, but with a slight difference in order. The correspondence of the threat and distrust items in the two questionnaires is shown in the Table 1.

Table 1. Correspondence between the threat indicators and the distrust indicators of the questionnaires from the pilot study and the General Election Survey

Pilot study	General survey	Balanced sets of items
v32_2	v108_2	In general, immigrants are not to be trusted.
v32_4	v108_4	Guest workers endanger the employment of the Belgians.
v32_7	v108_7	Muslims are a threat for our culture and customs.
v32_6	v108_6	The immigrants contribute to the prosperity of our country.
v32_8	v108_8	The presence of different cultures enriches our society.
v32_12	v108_10	We should kindly welcome the foreigners who come to live here.
v26_9	v97_7	The politicians have lost the ability to listen to ordinary people like me.
v26_12	v97_9	Once they are elected, most politicians feel themselves too good for people like me.
v26_4	v97_3	If people like me make their views know, they generally take them into account.
v26_6	v97_4	Most of our politicians are able people who know what they are doing.

3.2 Model specifications

In the case of both concepts, four hypothetical measurement models are distinguished. The *first* model is the most simple one. It specifies two content factors η^t and η^d, each with equal positive and negative loadings according to the direction of the item wordings.[3] The factor loadings are fixed +1 or -1, indicating that they are equal in absolute value within each construct, and all error covariances of the indicators are fixed zero ($\varepsilon_{i,j} = 0$ for $i \neq j$). The covariance between the two latent variables ($\psi_{1,2}$) is freed because we may expect a substantial positive correlation between the two concepts (see Billiet 1995a).

The *second* model is the model in which we are interested. It is the model that is specified in the hypotheses 1 and 2 concerning acquiescence in one balanced set of indicators and in hypothesis 3 about two balanced sets of indicators. It specifies two

[3] Actually, since the item scores were reversed, we expect positive loadings for the negatively worded items and negative loadings for the positively worded items. Thus high values on the latent variables mean 'feelings of threat' or 'distrust'.

content factors (η^t and η^d), and one style factor (η^{st}). All the loadings λ^{st}_i on the style factor are fixed +1 which means that we expect that the response to all items is equally affected by the style effect. The covariances $\psi_{1,3}$ and $\psi_{2,3}$ between the two content factors and the style factor are fixed to zero..[4].

The *third* model specifies two unipolar content factors for each concept, one pair for the negatively worded items (η^{t+} and η^{d+}) one pair for the positively worded items (η^{t-} and η^{d-}). From a theoretical point of view, this is not a nice situation since a set of indicators that was expected to measure one concept is split into two parts. When the correlations between the positive and negative factors are substantially high, then it seems better to use only one latent variable since this indicates that one is measuring the same concept. When the correlations between the factors are low, then one may seriously doubt the unidimensionality of the set of indicators or one may have doubts about the appropriateness of the linear model behind factor analysis (van Schuur and Kief 1994).

The *fourth* model specifies two content factors (η^t, η^d) and two style factors (η^{st1}, η^{st2}) indicating that there is a common factor behind the negatively and positively worded indicators of each of the two concept. This would indicate that we had identified two different response styles among different groups of respondents, or one style (acquiescence) which is content specific and not general. However, a substantially high correlation between the two styles suggests that they are measuring the same latent variable and permits us to switch to Model 2.

In the first step, the data of the pilot study are used for model specification, identification, estimation, and testing (Bollen and Long 1992) LISREL® 8 (Jöreskog and Sörbom 1993). In a second step, the selected model is tested and respecified in a randon part of the sample of the General Election Survey (N1 = 986), and then after final specification replicated in the second random part (N2 = 992). The models are estimated using different kinds of input matrices and estimation procedures. Most procedures lead to the same conclusions. In the pilot study the maximum likelihood estimations are reported in the tables. In the much larger samples of the General Election Survey, polychoric correlations and weighted least squares are used. This procedure is suggested by Jöreskog (1990).

[4] It is possible that acquiescence is correlated with specific attitudes but we have no a priori arguments for assuming that this is the case here.

Table 2. Four hypothetical measurement models for two balanced sets of items measuring two concepts

Items	(1) Two bipolar content factors($\psi_{1,2}\neq 0$)		(2) Two bipolar content factors and one style factor ($\psi_{1,2}\neq 0$)		
	η^t	η^d	η^t	η^d	η^s
T1	+1	0	+1	0	+1
T2	+1	0	+1	0	+1
T3	+1	0	+1	0	+1
T4*	-1	0	-1	0	+1
T5*	-1	0	-1	0	+1
T6*	-1	0	-1	0	+1
D1	0	+1	0	+1	+1
D2	0	+1	0	+1	+1
D3*	0	-1	0	-1	+1
D4*	0	-1	0	-1	+1

	(3) Four unipolar content factors (all $\psi_{i,j}\neq 0$)				(4) Two bipolar content factors and two style factors (4a: $\psi_{1,2}\neq 0$; 4b: $\psi_{1,2}$ and $\psi_{3,4}\neq 0$			
Items	η^{t+}	η^{t-}	η^{d+}	η^{d-}	η^t	η^d	η^{st1}	η^{st2}
T1	+1	0	0	0	+1	0	+1	0
T2	+1	0	0	0	+1	0	+1	0
T3	+1	0	0	0	+1	0	+1	0
T4*	0	+1	0	0	-1	0	+1	0
T5*	0	+1	0	0	-1	0	+1	0
T6*	0	+1	0	0	-1	0	+1	0
D1	0	0	+1	0	0	+1	0	+1
D2	0	0	+1	0	0	+1	0	+1
D3*	0	0	0	+1	0	-1	0	+1
D4*	0	0	0	+1	0	-1	0	+1

* Worded in reversed direction (opposite in content).

4 Results of the analysis

4.1 Exploration

The hypotheses 1 and 2 about a style factor (acquiescence) behind one balanced set of indicators are included in the models with two balanced sets of indicators. Let us therefore consider directly the two sets of indicators simultaneously. This exploration starts with the results of an initial exploratory factor analysis using the ML method in order to test an hypothesis about the number of factors. This means that the factors are extracted to only account for the common variance and not for the remaining unique variance (Hatcher 1993: 71). Table 3 contains the results from separate analyses for the 10 items which are intended to measure the two concepts THREAT and DISTRUST.

Several criteria are used for deciding about the number of factors (Hatcher 1994: 85-86). After the initial extraction of the factors, a promax rotation was performed since it is expected that the two factors are not orthogonal. The three highest eigenvalues are 6.99, 2.54 and 0.43. This suggests already a two factor solution as specified in Model 1. The signs and the values for the loadings of the standardized regression coefficients after promax rotation are also in line with that model. However, a Chi-square test indicates that the hypothesis stating that two factors are sufficient is rejected ($p = .005$). According to this criterion, it is reasonable to look also for a three factor solution.

The four hypothetical models are shown in Table 4. All λ's belonging to a construct are constrained to be equal as and all error covariances are fixed to zero. The models in Table 4 are identical to those in Table 2. All models are evaluated by several fit indices. The drop in Chi-square value, in combination with the number of degrees of freedom (df) and the p-value of the Chi-square statistic[5] provide information about the improvement of a model against the other models. On the basis of simulations, the Root Mean Square Error of Approximation (RMSEA) in combination with the p-value of close fit (Ho: RMSEA < .05) is preferred for evaluation because this measure discriminates best between "bad" and "good" models (Shevlin and Miles 1997). Models with RMSEA < .05 and with a high p-value of close fit (close to 1) are generally considered acceptable.

In this small sample, none of the models met the RMSEA criterion (< .05) for selection of an adequate model. This is not a problem because it is still possible to relax the equality constraints on the factor loadings λ^t_i and λ^d_i of the two content factors in order to improve the fit. The drop of the Chi-square value, in relation to the loss of degrees of freedom (at least 3 Chi-square units for 1 df), provides information about the improvement of one model over another one.

Model 2 with two content and one style factor is preferred over Model 1 with only two content factors, because of a drop in Chi-square of about 22 units of a loss of 1 df, and because the smaller RMSEA value with a larger p-value of close fit. For the same reasons, Model 2 is also superior to Model 4a with two content factors and two uncorrelated style factors. Model 2 is preferable over Model 4b because there is no substantial drop in the Chi-square value for a loss of 2 df and, more important, the two style factors are so strongly correlated ($r_{3,4} = .89$) that they must be quasi identical. At first glance, Model 3 with two pairs of unipolar content factors seems better than Model 2. However, we can seriously doubt this premature conclusion because of two reasons. Firstly, the drop in Chi-square of about 12 units is rather small for a loss of 6 df. Secondly, the correlations between the positive and negative

---

[5] The p-value criterion is only used in the case for small samples. In structural equation modeling with large samples, the p-value of the Chi-square statistic is not recommended as a strict criterion for model fit (Saris, Satorra and Sörbom 1987; Saris and Satorra 1993).

factors are so large ($r_{1,2}$ = -.89 and $r_{3,4}$ = -.83) that we may conclude that the same concepts are measured twice. Therefore, it is more parsimonious to specify a single factor for each concept. Summarizing, for both theoretical and statistical reasons, Model 2 is preferable over the other models.

Table 3. Result from an exploratory factor analysis (ML method): principal factors and promax rotation

Items	Principal factors		Promax rotation (Std. Reg. Coeff.)*	
	Factor 1	Factor 2	Factor 1 THREAT	Factor 2 DISTRUST
v32_2	.689	-.231	.686	.087
v32_4	.734	-.121	.626	.218
v32_7	.592	-.199	.590	.075
v32_6	-.640	.302	-.711	.006
V32_8	-.671	.276	-.711	-.034
v32_12	-.506	.385	-.684	.152
v26_9	.493	.378	.037	.605
v26_9	.594	.613	-.088	.887
v26_4	-.386	.-253.	-.069	-.431
v26_6	-.304	-.173	-.073	-.314
Variance explained	73.3%	26.6%	57.1%	42.9%

Inter-factor correlation = .42

The fit of the models is substantially better when the equality constraint on the λ^t_i and λ^d_i parameters is no longer maintained. Actually, freeing the loadings of the content factors (except one in each set for scaling reasons), is still in harmony with our hypotheses about acquiescence as long as the signs of λ's correspond with the positive and negative wordings of the items. An evaluation of the four models without equality constraints on the loadings of the content factors leads to the same conclusions as before. Model 2 (Chi-square = 49.221; df = 33; p =.034) has the smallest RMSEA (.051) with the highest p-value of close fit (.443). Model 3 is again second best (Chi-square = 46.382; df = 29; p = .02; RMSEA = .057; p-value of close fit = .334), but the correlations within each pair of content factors are quasi unchanged ($r_{1,2}$ = -.87 and $r_{3,4}$ = -.83). Therefore, Model 2 is still preferable over the other models. We observed that the modifications of the models did not substantially affect the correlation between the THREAT and DISTRUST that has a stable value of about .49.

Table 4. Pilot study: test information for models with two balanced sets of indicators (All λ's are fixed +1 and -1 and all $\varepsilon_{ij} = 0$ for $i \neq j$) (N = 188)

	Model	Goodness of fit indices					variances of the latent variables			
		Chi-square	df	prob.	RMSEA	p-value close fit	ψ_1 (t_1)	ψ_2 (t_2)	ψ_3 (t_3)	ψ_4 (t_4)
1	η^t, η^d,	88.788	42	< .0001	.077	.025	.611	.399	-	-
	$\psi_{1,2} \neq 0$						(8.114)	(6.341)		
							$r_{1,2}$ = .50			
2	η^t, η^d, η^{st}	66.539	41	.007	.058	.290	.622	.416	.043	-
	$\psi_{1,2} \neq 0$;						(8.231)	(6.611)	(3.672)	
	$\psi_{1,3} = \psi_{2,3} = 0$						$r_{1,2}$ = .49			
3	η^{t1}, η^{t2}, η^{d1}, η^{d2}	54.751	35	.018	.055	.361	.716	.605	.646	.285
	all $\psi_{i,j} \neq 0$						(7.584)	(7.100)	(6.258)	(3.341)
							$r_{1,2}$ = .-.89		$r_{3,4}$ = -.83	
							$r_{1,3}$ = .56		$r_{2,3}$ = -.34	
							$r_{1,3}$ = -.44		$r_{2,4}$ = .48	
4a	η^t, η^t, η^{st1}, η^{st2}	75.950	40	.0005	.069	.090	.622	.421	.040	.061
	$\psi_{1,2} \neq 0$;						(8.227)	(6.654)	(2.449)	(2.165)
	all other $\psi_{i,j} = 0$						$r_{1,2}$ = .49		-	
4b	η^t, η^t, η^{st1}, η^{st2}	66.018	39	.004	.061	.227	.621	.421	.043	.061
	$\psi_{1,2} \neq 0$; $\psi_{3,4} \neq 0$;						(8.221)	(6.634)	(3.086)	(2.268)
	all other $\psi_{i,j} = 0$						$r_{1,2}$ = .49		$r_{3,4}$ = .89	

The evaluation of Model 2 is reported in Table 5. This is the model in which the equality constraint on the λ^t_i and λ^d_i parameters (see Table 1) is dropped. All the error covariances $\varepsilon_{i,j}$ between the observed variables are all fixed to zero.

The signs of the unconstrained λ parameters of the two content factors are exactly as was expected in Model 2 (see Table 2). They are in general large. Only the parameters of the two positively worded indicators of DISTRUST have rather low values, indicating that a large amount of unique variance is still unexplained in this model. This is also apparent in the squared multiple correlations (R^2). Less than 1/3 of the variance in these two indicators is explained by the latent variables (content and style) in the model. The variance of the STYLE factor (ψ_3) is substantive (η^{st} = .039; t = 3.400), but is, as was expected, lower than the variances of the two content factors ψ_1 (.709; t = 4.116) and ψ_2 (.481; t = 3.982). The λ parameters belonging to the STYLE factor all differ from zero though they are considerably smaller than the λ's for the two content factors. This is reasonable: the indicators are mainly affected by each of the content factors and not by a style factor. There is no correlation between the two content factors and the style factor, and the correlation between THREAT and DISTRUST is strong ($r_{1,2}$ = .49) as was expected on theoretical grounds (Billiet 1995a).

Table 5. Pilot study: evaluation of measurement Model 2 with two content factors and one style factor (all error covariances $\varepsilon_{i,j}$ fixed to zero)

Items	Unstandardizedλ parameters THREAT (t-value)	DISTRUST (t-value)	STYLE (t-value)	Standardized λ parameters THREAT	DISTRUST	STYLE	R^2
v108_2	1 (fixed)	0	1	.842	0	.197	.561
v108_4	1.049 (9.433)	0	1	.884	0	.197	.563
v108_7	.871 (8.030)	0	1	.733	0	.197	.413
v108_6	-.952 (-8.670)	0	1	-.802	0	.197	.527
v108_8	-.944 (-8.975)	0	1	-.795	0	.197	.572
v108_10	-.759 (-7.225)	0	1	-.639	0	.197	.371
v97_7	0	1 (fixed)	1	0	.693	.197	.422
v97_9	0	1.218 (6.374)	1	0	.845	.197	.607
v97_3	0	-.778 (-5.302)	1	0	-.539	.197	.308
v97_4	0	-.697 (-4.616)	1	0	-.482	.197	.220
$\psi_{i,j}$	THREAT	DISTRUST	STYLE	THREAT	DISTRUST	STYLE	
THREAT	.709 (5.437)			1			
DISTRUST	.283 (4.116)	.481 (3.982)		.49	1		
STYLE	0	0	.039 (3.400)	0	0	1	

Chi-square = 49.221; df = 33, p = .034; RMSEA = .051, p-value for test of close fit = .443 (N = 188) ML estimation

Overall, the existence of a style or response-set factor is not rejected. In keeping with the hypotheses, Model 2 identified a common factor (hypothesis 1) with non-zero variance behind two sets of indicators (hypothesis 3), and with method variance which is lower than the variance of the two content factors (hypotheses 2). Is this STYLE acquiescence? All the λ's on the style factor corresponding with positively or negatively worded items, have positive loadings. This is what we expect in the presence of acquiescence. However this is not decisive proof and additional support for acquiescence is required. In order to provide additional support for the hypothesis of acquiescence, a confirmatory analysis will be conducted using data of a much larger sample from the same population. Moreover, we will to identify the style factor on a more substantial way by observing its relationship with other constructs like education and "scoring for acquiescence" (Ray 1997).

4.2 Confirmation

Does measurement Model 2, with the indicators selected in the pilot study, apply to the general population? Is it stable in a different question context? In the questionnaire of the 1995 General Election Survey in Flanders, the two sets of indicators are in the same order but the previous questions differ from those in the pilot study.

Table 6. Estimation of Model 2 with two content factors and a style factor in the second random sample of the 1995 General Election Survey Flanders (t-values for freed parameters within brackets).

	Unstandardized λ parameters			Standardized λ parameters			
Items	THREAT (t-value)	DISTRUST (t-value)	STYLE (t-value)	THREAT	DISTRUST	STYLE	R^2
v108_2	1 (fixed)	0	1	.786	0	.169	.645
v108_4	.967 (27.324)	0	1	.759	0	.169	.605
v108_7	.823 (20.336)	0	1	.647	0	.169	.446
v108_6	-.913 (-24.572)	0	1	-.717	0	.169	.543
v108_8	-.919 (-21.077)	0	1	-.722	0	.169	.549
v108_10	-.889 (-22.867)	0	1	-.699	0	.169	.516
v97_7	0	1 (fixed)	1	0	.676	.169	.485
v97_9	0	1.271 (12.703)	1	0	.859	.169	.766
v97_3*	0	-.745 (-10.907)	1	0	-.504	.169	.282
v97_4*	0	-.582 (-9.355)	1	0	-.393	.169	.183
$\psi_{i,j}$	THREAT	DISTRUST	STYLE	THREAT	DISTRUST	STYLE	
THREAT	.671 (20.422)			1			
DISTRUST	.252 (9.280)	.457 (8.855)		.48	1		
STYLE	0	0	.028 (3.938)	0	0	1	

Chi-square = 55.068; df = 32, p = .007; RMSEA = .027, p-value for test of close fit = 1.00 (N = 992)

WLS estimation based on polychoric correlations and the asymptotic variance-covariance matrix

* Error covariance $\varepsilon_{9,10} \neq .147$ (z = 3.823)

We decided to split the sample into two random parts. The models reported in the pilot study are tested in one part of the sample (N = 986). Freeing the error covariance between the two positively worded DISTRUST items ($\varepsilon_{9,10}$) led to a substantial improvement of the model fit. We had already observed that a large amount of variance in these two indicators was not explained by the latent variables in the model. Accepting common error covariance in these two items indicates that they may be also affected by an unidentified source. There are indications that it is item specific acquiescence that was found, but we will explore this idea further in another study. In this analysis, Model 2 is again selected as the most adequate model on both theoretical and statistical grounds (Chi-square = 67.452, df = 32; RMSEA = .043, p-value of close fit = .849). Exactly the same models are then applied to the second part of the sample (N = 992) without any further modification. Model 2 with two content factors an a style factor, is clearly reproducible.

Table 6 shows the parameters of Model 2 estimated in the second random sample (N = 992) of the 1995 General Election Survey. It is exactly the same model that was selected in the other part of the sample and in the pilot study, with the exception of the non-zero error covariance $\varepsilon_{9,10}$ between the two positively DISTRUST indicators.

According to the fit indices, in the second sample the estimated model is somewhat closer to the data than it was in the first sample.

So far, we may conclude that it is theoretically meaningful and empirically possible to specify an additional response-style factor for two sets of balance items, measuring two concepts. Our empirical data sustain four of the five hypotheses. The values and signs of parameters in Model 2 are in accordance with what we might expect from 'acquiescence', however those who do not accept the existence of this response effect may still require more evidence. Therefore, we will provide two additional arguments, one about the relationship of our latent variable STYLE with the degree of education; the other about the relationship of this latent factor with the number of agreements on all of the items.

4.3 Identification of the response-style factor

It is sometimes claimed that acquiescence has a negative relationship with education (Schuman and Presser 1981; Mirowsky and Ross 1991; McClendon 1991b; Watson 1992; Narayan and Krosnick 1996). There may be several reasons for this negative correlation, for example the abstractness of the items. Lower educated respondents may have a less clear views on certain types of items. We estimated a model with a latent exogenous variable EDUCATION (measured by the level of the obtained certificate) as a predictor for the three latent variables.

The standardized regression parameters[6] of EDUCATION on THREAT ($\gamma_{1,1}$ = .-.45) and on DISTRUST ($\gamma_{2,1}$ = -.31) are in harmony with theoretical expectations and with previous studies (Billiet 1995a). The standardized regression coefficient expressing the effect of EDUCATION on STYLE ($\gamma_{3,1}$) is -.23. The latter effect indicates that a standard unit drop in education results in an increase of .23 units on the style variabele. This is a good sign, but not a proof, that the style factor identifies acquiescence, because such relationship is also expected for other response effects such as a middle-alternative effect (Narayan and Krosnick 1996). For balanced sets of items, the middle alternative effect might result in the same kind of structural relationships as we specified in Model 2.

[6] The relationships between the predictors and the dependent variables are nearly the same in Model 2 as in Model 1 in which only two content factors are specified.

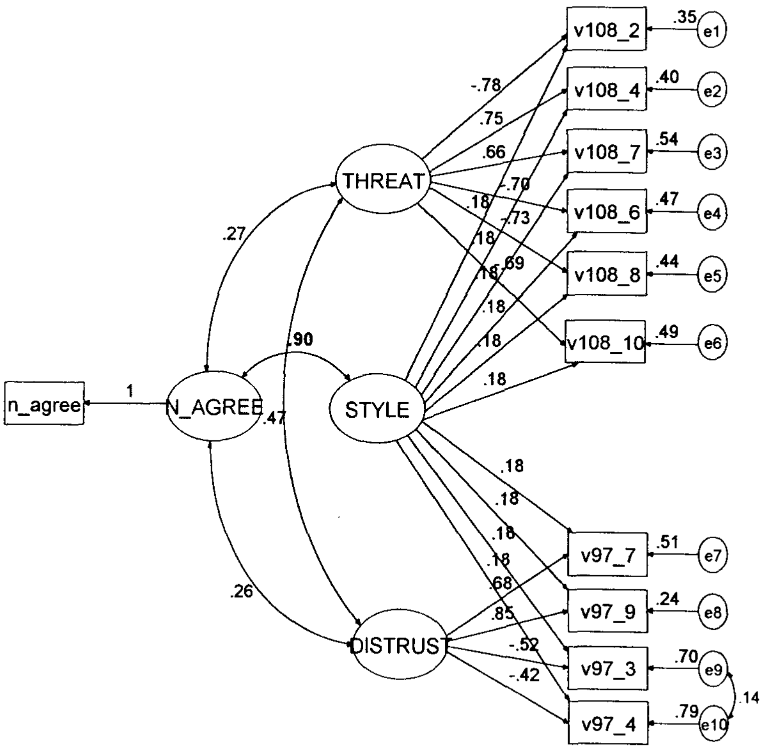

Figure 1: Model 2 with two content factors, a style factor and a score for agreement in the 1995 General Election Survey (standardized parameters)

The second argument deals with the relationship of the STYLE variable with the sum of agreements across all items. The latter is called "scoring for acquiescence" (Ray 1979). This construct is an additive scale ranging from 0 to 14, defined by the sum of the agreements (N_AGREE) in four balanced distrust items, eight balanced threat items, and two clearly opposite worded items on individualism and communalism. We expect a very large positive correlation between the latent STYLE variable and the sum of agreements if the first indeed identifies acquiescence. There are two ways to test this hypothesis: (1) adding the observed variable "sum of agreements" as an additional indicator for STYLE and looking at the corresponding parameter ($\lambda_{11,3}$), or (2) specifying a new latent variable N_AGREE with one indicator "sum of agreements", and looking at the correlation between N_AGREE and STYLE. Actually, the two procedures lead to the same conclusion, a strong correlation ($r_{3,4}$) of N_AGREE with STYLE, or a strong correlation of the indicator "number of agreements" ($\lambda_{11,3}$) with the latent factor STYLE. Figure 1 shows the second way of testing.

The correlation between the one-indicator variable N_AGREE and the latent variable STYLE is .90 ($z = 22.260$). This is a strong indication that we are measuring the same construct in two different ways. We may conclude that STYLE is indeed measuring acquiescence because of its strong relationship with a variable (N_AGREE) that simply counts the number of agreements with the positively and negatively worded statements in a balanced set of fourteen items. The moderate positive correlations of scoring for acquiescence with the two content variables are not so surprising because those who tend to agree with positively and with negatively worded items are located in the middle of the scales of the latent content factors.

5 Conclusions and discussion

An advantage of this study is that a pilot study was used for exploration and two samples of a larger general survey for confirmation. Four out of the five hypotheses concerning the measurement of acquiescence in balanced sets of items are confirmed in the 1995 General Election Survey in Flanders. Our findings support the idea that acquiescence exists in responses to agree-disagree items which are used in samples from a general population. Moreover, it is clearly demonstrated that it is possible to control for acquiescence in the context of structural equation models.

Several questions are still open. In all our tests, several competitive models were acceptable, even models with two style factors when no correlated error variance between some indicators of one concept was accepted. Is this an indication that partly subsets of respondents are differently sensitive for acquiescence, depending on the concept and the type of items? Further researh is needed in order to answer that question. A positieve answer would seriously challenge the plausibility of the fifth proposition about acquiescence as a stable personality trait. Accepting correlated uniqueness among two items leads to an acceptable model with one response-style factor behind the two sets of items as was specified in the hypotheses one to three.

The decision to model acquiescence as in the present study, and the selection of a model must depend largely on theoretical grounds. From the start, we were trying to model a response-style factor of a method effect that met the requirements of indirect measurement of acquiescence as an additional common factor behind two balanced sets of items. However, we could only identify it at the end of the study by relating the latent construct with a variable measured by the number of agreements on a balanced set of fourteen items (scoring for acquiescence). The strong correlation

(.90) between the two variables indicates that it are two measurements of the same concept, acquiescence.

Several improvements are possible in future research. Our indicators were selected from larger sets of balanced items. Model 2 applied to several combinations but not to all. For this study, we choose subsets of items that were somewhat contradictory in meaning. However, we were not able to work with strict reversals or contrary statements. Is it possible that moderate or ambivalent respondents are more likely to agree with both positively and negatively worded items? About 8% of the respondents agreed with seven or more items of the balanced scale. This finding is in harmony with the fourth hypothesis, however we cannot convincingly exclude that these are moderate or even ambivalent respondents, and not acquiescent ones. We know that our style factor was negatively correlated with education, and because of this it is expected that sophisticated or moderate respondents are less likely to agree with both positively and negatively worded items. Pure reversals are needed in order to exclude the alternative explanations. However, in surveys it is not adequate to put pure reversals together in one set of items because even the acquiescent respondents will see the contradictions, and subsequently try to avoid inconsistent answers. In further research it is intended to split the items into two balanced subsets and to have one of these parts placed in a separate part of the questionnaire. Pure reversals are possible with this procedure.

Another improvement and even cross-validation is possibly with a somewhat different construction of "scoring for acquiescence". It is recommended to use another balanced set of items for the sum of agreements (N_AGREE), independent of the sets that are used in the structural model. We expect that the moderate correlation of that variable with the two content variables will disappear in that case.

We expected substantial changes in the structural relations between the content variables and the predictor, claiming that the estimations within a model with a style factor are more valid. In our data, the changes in the structural relationships are very small. However, we can argue that some changes can occur in other combinations of variables. In any case, the specification of a style or method factor behind (quasi) balanced sets of items that are intended to measure one concept may prevent an inflation of strongly correlated content factors in factor analysis.

References

[1] Bentler, P. M. (1969): Semantic space (approximately) bipolar. *The Journal of Psychology*, **71**, 33-40.

[2] Bentler, P. M., Jackson, D.N. and Messick, S. (1971): Identification of context and style: two-dimensional interpretation of acquiescence. *Psychological Bulletin*, **76**, 186-204.

[3] Billiet, J., Carton, A. and Huys, R. (1990): *Onbekend of onbemind. Een sociologisch onderzoek naar de houding van de belgen tegenover migranten*. Leuven: SOI.

[4] Billiet, J. (1995a): Church involvement, individualism, and ethnic prejudice among Flemish Roman Catholics. *Journal of the Scientific Study of Religion*, **34**, 224-233.

[5] Billiet, J. (1995b): *Theoretical dimensions and measurement models of attitudes towards ethnic minorities*. Paper presented at the Eight Conference of the International Group on Methodology and Comparative Survey Research, Prague, May 4-7, 1995.

[6] Block, J. (1965): *The challenge of response sets*. New York: Appleton-Century Crafts.

[7] Bollen K.A. and Long, J. S. (1992): Tests for structural equation models. *Sociological Methods and Research*, **21**, 123-131.

[8] Cloud, J. and Vaughn, G. M. (1979): Using balanced scales to control acquiescence. *Sociometry*, **33**, 193-206.

[9] Cronbach, L.J. (1942): Studies of acquiescence as a factor in the true-false test. *Journal of Educational Psychology*, **33**, 401-415.

[10] Cronbach, L.J. (1946): Response sets and test validity. *Educational and Psychological Measurements*, **6**, 475-494.

[11] Cronbach, L.J. (1950): Further evidence on response sets and test design. *Educational and Psychological Measurement*, **10**, 3-31.

[12] Diener, E. and Emmons, R.A. (985): The independence of positive and negative affect. *Journal of Personality and Social Psychology*, **47**,1105-1117.

[13] Green, D.P. (988): On the dimensionality of public sentiment toward partisan and ideological groups. *American Journal of Political Science*, **32**, 758-780.

[14] Hamilton, D. (1968): Personality attributes associated with extreme response style. *Psychological Bulletin*, **69**, 192-203.

[15] Hatcher, L. (1993): *A Step-by-Step Approach to Using the SAS® System for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute Inc.

[16] Krosnick, J.A. (1991): Response strategies for copingwith the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, **5**, 213-236..

[17] Jackson, D.N. and Messick, S. (1958): Content and style in personality assessment. *Psychological Bulletin*, **55**, 243-252.

[18] Jackson, D.N. and Messick, S. (1958): Content and style in personality assessment. *Psychological Bulletin*, **55**, 243-252.

[19] Jackson, D.N. and Messick, S. (1962): Response styles and the assessment of psychopathology. In S. Messick and J. Ross (Eds.), *Measurement in personality and cognition*. New York: John Wiley.

[20] Jöreskog, K., G. (1990): New developments in LISREL. Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, **24**, 387-404.

[21] Jöreskog, K. G., Sörbom, D. (1993): *LISREL® 8 User's Reference Guide.* Chicago: Scientific Software International.

[22] Lorr, M. and Shea, T.M. (1979): Are mood states bipolar? *Journal of Personality Assessment*, **43**, 468-472.

[23] Marsh, H.W. (1996): Positive and negative global self-esteem: A substantively meaningful distinction of artifactors?, *Journal of Personality and Social Psychology*, **70**, 810-819.

[24] Martin, J. (1964): Acquiescence: measurement and theory. *British Journal of Social and Clinical Psychology*, **3**, 316-225.

[25] McBride, L. and Moran, G. (1967): Double agreement as a function of ambiguity and susceptibility to demand implications of the psychological situation. *Journal of Personality and Social Psychology*, **6**, 115-118.

[26] McClendon, McKee J. (1991a): Acquiescence and recency response-order effects in interview surveys. *Sociological Methods and Research*, **20**, 60-103.

[27] McClendon, McKee J. (1991b.): Acquiescence: tests of the cognitive limitations and Question Ambiguity Hypotheses. *Journal of Official Statistics*, **7**, 153-156.

[28] McClendon, McKee J. (1992*): On the measurement and control of acquiescence in latent variable models*. Presented at the 1992 meeting of the American Sociological Associations, Pittsburg, August 20-24.

[29] McClendon, McKee J., Billiet, J. and Waege, H. (1996): Measurement and control of acquiescence in surveys: combining split ballots and structural models. *Proceedings of the International Conference on Survey Quality and Process Control*. Alexandria: American Statistical Association, Bristol, April 1-4, 1995, 233-238.

[30] Mirowsky, J. and Ross, C. E. (1991): Eliminating defense and agreement bias from measures of the sense of control: A 2 x 2 Index. *Social Psychology Quarterly*, **54**, 127-45.

[31] Narayan, S. and Krosnick, J. A.. (1996): Education moderates some response effects. *Public Opinion Quarterly*, **60**, 89-105.

[32] Nunnally, J. C. (1967): *Psychometric Theory*. New York: McGraw-Hill.

[33] Paulhus, D. L. (1991): Measurement and control of response bias. In J.P. Robinson, P.R. Shaver and L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes*. San Diego CA: Academic Press, 17-59.

[34] Ray, J. J. (1979): Is the acquiescent response style problem not so mythical after all: Some results from a successful balanced F-scale. *Journal of Personality Assessment*, **43**, 638-643.

[35] Ray, J .J. (1983): Reviving the problem of acquiescent response bias. *Journal of Social Psychology*, **121**, 81-96.

[36] Ray, J. J. (1984a): Reinventing the wheel: Winkler, Kanouse, and Ware on acquiescent response set. *Journal of Applied Psychology*, **69**, 353-355.

[37] Ray, J. J. (1984b): A further comment on the Winkler, Kanouse, and Ware method of controlling for acquiescent response bias. *Journal of Applied Psychology*, **69**, 359.

[38] Ray, J.J. (1990): Acquiescence and problems with forced-choices scales. *Journal of Social Psychology*, **130**, 397-399.

[39] Rorer, L.G. (1965): The great response-style myth. *Psychological Bulletin*, **63**, 129-156.

[40] Toner, B. (1987): The impact of agreement bias on the ranking of questionnaire response. *Journal of Social Psychology*, **127**, 221-222.

[41] Saris, W. (1988): Individual response functions. In W.E. Saris and I.N. Gallhofer (Eds.), *Sociometric research: Data collection and scaling* (Vol 1.), 200-214.

[42] Saris, W.E., Satorra, A. and Sörbom, D. (1987): The detection and correction of specification errors in structural models. In: C.C. Clogg (ed.), *Sociological Methodology* 1987. San Francisco: Jossey Bass, 105-129.

[43] Saris, W.E. and A. Satorra (1993): Power evaluations in structural models. In: K.A. Bollen and J.S. Long (eds.), *Testing structural equation models*. Newbury Park: Sage, 108-204.

[44] Schuman, H. and Presser, S. (1981): *Questions and answers in Attitude Surveys*. New York: Academic.

[45] Shevlin, M. and Miles, J. (1997): *An investigation of the performance of incremental fit indices in confirmatory factor analysis. A Monte Carlo simulation* Paper presented at the Annual Conference of the British Psychological Society. Northern Ireland Branchn Dunfanaghy, April 25-27.

[46] Spector, P.E. (1992): *Summated rating scale construction: An introduction*. Newsbury Park, CA: Sage.

[47] van Schuur, W.H. and Kiers, H. 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.

[48] Watson, Dorothy. (1992): Correcting for acquiescent response bias in the absence of a balanced scale: an application to class consciousness. *Sociological Methods and Research*, **21**: 52-88.

[49] Weisberg, H.F. (1980): A multidimensional conceptualization of party identification. *Political Behavior*, **2**, 33-60.