SOCIOLOGICAL METHODOLOGY

2

DEALING WITH EXTREME RESPONSE STYLE IN CROSS-CULTURAL RESEARCH: A RESTRICTED LATENT CLASS FACTOR ANALYSIS APPROACH

*Meike Morren**
*John P.T.M. Gelissen**
*Jeroen K. Vermunt**

Cross-cultural comparison of attitudes using rating scales may be seriously biased by response styles. This paper deals with statistical methods for detection of and correction for extreme response style (ERS), which is one of the well-documented response styles. After providing an overview of available statistical methods for dealing with ERS, we argue that the latent class factor analysis (LCFA) approach proposed by Moors (2003) has several advantages compared to other methods. Moors' method involves defining a latent variable model which, in addition to the substantive factors of interest, contains an ERS factor. In LCFA the observed ratings can be treated as nominal responses, which is necessary for modeling ERS. We find strong evidence for the presence of ERS and, moreover, find that the groups differ not only in their attitudes but also in ERS. These findings underscore the importance of controlling for ERS when examining attitudes in cross-cultural research.

This research was supported by a grant from the Netherlands Organization for Scientific Research, MaGW/NWO project number (no. 400-05-164). Direct correspondence to Meike Morren, Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; e-mail: m.morren@uvt.nl.

*Tilburg University, The Netherlands

1. INTRODUCTION

Public, political, and social scientific awareness of a rapidly globalizing world has provided an enormous impetus for the cross-cultural study of empirical value and attitude patterns in recent decades. More and more surveys are conducted in culturally diverse populations, either within one country or between two or more countries. A well-known single-country study with a cross-cultural focus is the General Social Survey in the United States. Well-known examples of cross-national studies are the International Social Survey, the European Social Survey, the European Values Study, and the World Values Study. The growing number of cross-cultural surveys and the wealth of publications that is forthcoming from these data are a testament to the heightened interest in cross-cultural differences in attitudes and values.

Cross-cultural analyses yield crucial insights into the substantive attitude and value structures of culturally diverse populations. It is likely that people who come from different sociocultural backgrounds will interpret the world differently. Their frame of reference forms a tool to make sense of the world and is influenced by cultural values and norms that are transmitted in their upbringing, neighborhood, and school. These experiences culminate in a certain pattern of values, attitudes, and behavior (Wallace and Wolf 1998). The goal of most cross-cultural studies is to reveal the differences in these reference frames in order to explain cross-cultural differences in behavior.

However, the validity of such studies can be seriously reduced by biases distorting the measurement of attitudes and possibly affecting the outcome of cross-cultural comparisons (Van de Vijver and Leung 1997). For example, it is not always evident whether a set of items measures the same attitudinal construct in each cultural group. A specific type of bias that distorts attitude measurement in general and therefore plays an important role in the literature on survey methodology is response style behaviors, which have been described as “the systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content” (Paulhus 1991:17). In this paper, we focus in particular on methods for detection of and correction for extreme response style (ERS) behavior because its presence may invalidate group differences in attitudes measured by rating questions (for instance, see Bachman and O’Malley 1984; Clarke III 2001; De Jong et al. 2008; Greenleaf 1992a). An extreme response style results in a response pattern where a respondent predominantly selects the outer response categories of rating questions irrespective of his or her opinion. This response behavior confounds attitude measurement because the nonrandom response error blends with the content of the items that is intended to reflect an underlying attitude. It also has a biasing effect on the average value of the responses and on their correlations with covariates of interest. Of particular relevance for cross-cultural research is that it has repeatedly been shown that the presence of extreme response style differs across cultures (for instance, see Clarke III 2001; Gibbons, Zellner, and Rudek 1999; Harzing 2006; Hui and Triandis 1989; Johnson et al. 2005). Since both attitudes as well as the extreme response style can differ cross-culturally, comparison of these attitudes between ethnic groups can reflect cultural differences in attitudes or response style (Eid, Langeheine, and Diener 2003), a type of problem that is sometimes also referred to as the duality between genuine and stylistic variance (Poortinga and Van de Vijver 1987).

Although applied researchers are usually aware of these complicating issues, they often silently assume that their measurements can be compared across groups and that response style behavior does not seriously affect their measurements. Needless to say, such assumptions should be empirically investigated. Moors (2003, 2004; see also Green and Citrin 1994) not only strongly advocated this basic principle but also observed that there is no single accepted methodological approach for dealing simultaneously with construct inequivalence and response style behavior, although it is generally accepted that both distort the comparison of attitudes across groups. Moors (2003, 2004) showed how to use the latent class factor analysis (LCFA) model proposed by Magidson and Vermunt (2001) to define a statistical model for detecting attitudinal differences in culturally diverse groups that are corrected for group differences in extreme response style behavior. Moors' method involves defining a latent variable model that, in addition to the substantive factors of interest, contains an ERS factor. This LCFA approach bears close resemblance to the confirmatory factor models proposed for dealing with an acquiescent response style (Billiet and McClendon 2000; Cheung and Rensvold 2000). Differences are that in LCFA the latent variables are treated as ordinal and, moreover, that the ratings can be treated as nominal items, which is necessary for modeling ERS, as will be shown in the remainder of this papaer. Recent advances in statistical software for latent structure analysis make this model readily available to applied researchers.

This paper contributes to the existing literature in several ways. We provide the reader with an overview of approaches for detecting extreme response styles in survey data. In addition, we give a step-by-step exposition of the modeling approach proposed by Moors (2003, 2004) for detecting and adjusting for response style behavior in culturally diverse groups, and we discuss how it relates to other approaches. Furthermore, we propose an important extension of Moors' original model by making more strict (ordinal) assumptions about the items' relationships with the content-related factors. This not only leads to more parsimonious models but also makes a more clear distinction between the content-related factors and the response-style factor. Moors' approach as well as our extended LCFA approach are illustrated using data from the Dutch survey "The Social Position of Ethnic Minorities and Their Use of Services" (SPVA),[1] which allows the investigation of—and correction for—differences in extreme response style behavior between four culturally diverse groups. Thus, we heed the call of many authors, including Van de Vijver and Leung (1997, 2000), Cheung and Rensvold (2000), Krosnick (1999), Moors (2003, 2004), and Green and Citrin (1994), and empirically investigate the degree to which response style behavior confounds the measurement of attitudes.

2. METHODS FOR DETECTING EXTREME RESPONSE STYLE: AN OVERVIEW

Extreme response style is commonly defined as the tendency of respondents to express themselves extremely by choosing the end points on a rating scale, independent of the extremity of their opinions. This tendency is typically assumed to exist irrespective of the substantive item content but to show up in consistency with the positive or negative

[1] In Dutch, the abbreviation SPVA stands for *Sociale Positie en Voorzieningengebruik van Allochtonen* (Social Position and use of Facilities by Immigrants). We thank Data Archiving and Networked Services (DANS) for providing the data files.

formulation of an item[2] (De Jong et al. 2008; Greenleaf 1992b; Moors 2004; Paulhus 2002). Whereas several studies have found ERS to be a consistent individual difference (e.g., Hamilton 1968; Peabody 1962), others find that the influence of ERS changes as the survey progresses (Hui and Triandis 1985; Krosnick 1991), as the questions are formulated in another language (Gibbons, Zellner, and Rudek 1999), or as different survey methods are used (Bachman and O'Malley 1984; Van Herk, Poortinga, and Verhallen 2004). Following Hui and Triandis (1989) and Podsakoff and colleagues (2003), we argue that the occurrence of ERS is the result of an interaction of characteristics of the respondent and of the item concerned. More specifically, ERS is a characteristic of the respondent (a trait) indicating whether he or she tends to answer more extremely than other respondents in the investigated population. The degree to which this tendency actually appears in a particular rating scale depends on item characteristics such as response format, item content, location in the questionnaire, and so forth. Thus, some questions are more likely to elicit extreme response style than others.

Whether an extreme answer reflects a truly extreme attitude or rather ERS is impossible to determine from a single response. However, with multiple ratings it is sometimes possible to determine whether a respondent tends to provide more extreme answers than others in the sample. Several methods—ranging from straightforward descriptive methods to rather advanced statistical models—have been developed to measure ERS using multiple ratings. Whereas some researchers are mainly interested in methods for detecting ERS (De Jong et al. 2008; Greenleaf 1992b; Hui and Triandis 1989; Johnson et al. 2005; Sudman and Bradburn 1974), others focus on methods for correcting the biasing influence of ERS on the measurement of attitudes (Greenleaf 1992a; Marin, Gamba, and Marin 1992; Saris 1998), or comparing the influence of ERS on attitudes across different survey methods (King et al. 2004; Saris and Aalberts 2003; Weijters, Schillewaert, and Geuens 2008).

The easiest and most intuitive method for detecting ERS is to construct an ERS sum-score index (Gibbons et al. 1999; Harzing 2006; Johnson et al. 2005). Such an index is obtained by dichotomizing the

[2]This separates extreme response style from acquiescence, where respondents tend to agree or disagree with all items of a set *regardless* of their positive or negative content (Moors 2004: 304).

original items, where 1 refers to an extreme answer and 0 to one of the other item categories, and subsequently counting the number of extreme answers (number of 1s). The validity of such an ERS measure is improved by using a set of items that are unrelated in content. Greenleaf (1992a) developed a specifically designed measurement instrument for ERS consisting of unrelated 16-items, which was included in a survey by Arce-Ferrer (2006).

Greenleaf's ERS scale or a sum-score using related items in content can be used not only to detect respondents with ERS but also to assess differences in ERS between sociocultural groups as well as to control for ERS in subsequent statistical analysis (Bachman and O'Malley 1984; Clarke III 2001; Hui and Triandis 1989; Marin et al.1992). The measurement of ERS by means of a separate ERS scale has found very limited application, for various reasons including the additional costs it involves during the data collection process (De Jong et al. 2008).

Despite its simplicity, the use of the sum-score method with survey items developed for the measurement of one or more substantive dimensions has several drawbacks as well. The first drawback is that the recoded items no longer reflect the attitude dimensions of interest. It is clear that by collapsing the responses into two new categories (extreme versus remaining answer categories), which is needed for the calculation of the sum-score, most information about the underlying attitudes (reflected in the original response scale) is lost. Another drawback is that the ERS dimension may be confounded with substantive dimensions when items used to measure ERS are related in terms of attitudes (De Jong et al. 2008; Greenleaf 1992b). Typically, a large number of items on different topics are included to ensure that no single dominant attitude dimension has a substantial effect on the item responses. However, in the sum-score method it is not possible to control the ERS measurement, the reason being that pairs of items may be associated because they concern the same attitude or correlated attitudes. A final problem is that all items get the same weight when constructing the ERS scale, which is incorrect when proneness to ERS differs across items.

An alternative approach that overcomes the problems associated with the sum-score method involves the use of a latent variable model, such as an item response theory (IRT) model, confirmatory factor analysis (CFA) or structural equation modeling (SEM), or latent class analysis (LCA). First, in a latent variable model the items can be

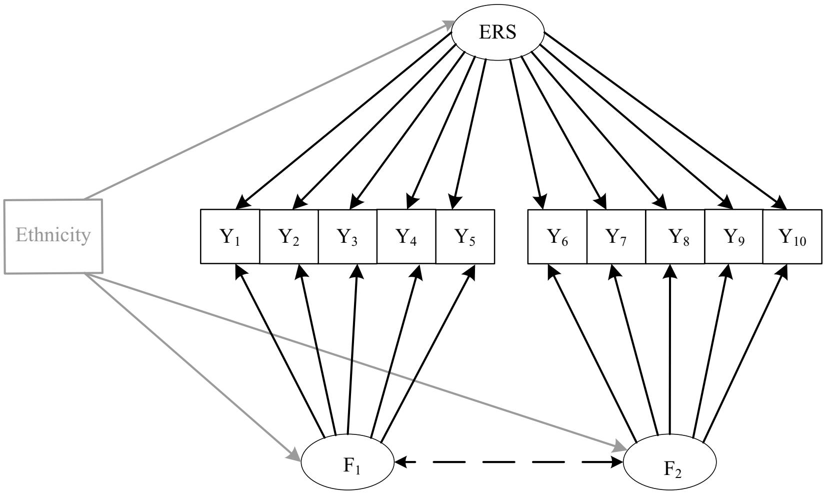

FIGURE 1. The latent variable model for the detection of a response style.

used in their original scales rather than in their dichotomized extreme response forms. This makes it possible to account for the substantive correlations among items measuring the same construct by including a separate latent variable for each construct. Second, in a latent variable model we can also include a latent variable representing the response style. This makes it possible to measure ERS controlling for substantive factors and vice versa. The latent style factor may have different effects across items, which is a way to take into account that items may be differently affected by ERS or—related to this—that some items may simply be inappropriate for detecting ERS. Lastly, and most importantly in the context of cross-cultural research, such a latent variable model may yield estimates for the group differences in attitudes while controlling for group differences in ERS.

An example of such a latent variable model is depicted in Figure 1. Here, Y1–Y10 represent item responses, F_1 and F_2 are two substantive factors, and ERS is the extreme response style factor. Ethnicity is a covariate affecting the substantive factors as well as the ERS factor. Note that when a separate measurement instrument for ERS is available, it could be used as an observed control variable or as a latent control variable with its own indicators in the latent variable model for the substantive factors of interest.

De Jong et al. (2008) proposed an IRT model for measuring ERS, which assumes that a continuous, stable, latent ERS trait underlies an individual's observed extreme response pattern in a multiple item set. An important feature of their model is that they use the items in dichotomized form (extreme versus remaining categories). Since IRT models typically assume unidimensionality—in other words, only one latent variable is included in the model (for instance, see Sijtsma and Molenaar 2002)—a multidimensional extension was needed to be able to control for correlations caused by content factors. As De Jong et al. indicated, their method improves on existing procedures by allowing items to be differentially useful for measuring ERS and by relaxing the requirement that the items in an ERS measure should be (marginally) uncorrelated, which allows the construction of an ERS measure based on substantively correlated items and eliminates the need for a dedicated ERS scale. A disadvantage of this approach is that it uses the items in their dichotomized form, which means that most of the information about the attitudinal constructs is lost. Another disadvantage is that estimation of the parameters of the model by De Jong et al. (2008) requires the use of rather complex Bayesian Markov chain Monte Carlo (MCMC) procedures, which makes the approach less accessible to applied researchers.

A model for dealing with response styles using the original ordinal items was proposed by Rossi, Gilula, and Allenby (2001). It is a hierarchical multivariate probit model with a location and a scale parameter that varies across individuals. Though it can capture various types of scale usage heterogeneity (this is what they call response style), it cannot deal with ERS as defined in the current paper—namely, the tendency to select the more extreme (or more moderate) response irrespective of whether the true option is negative or positive. Johnson (2003) proposed an extension of the Rossi et al. (2001) model that overcomes this limitation; that is, he defined a heterogeneous threshold model that can be seen as a model in which the person-specific scale factors differ across item categories. Two simplifying assumptions made by Johnson are that thresholds are symmetric across negative and positive responses and equal across items. It should be noted that neither the model by Rossi et al. (2001) or the one by Johnson (2003) is an IRT or factor analytic model. However, Johnson (2003) showed how his model can be used to define a latent variable model with substantive factors in addition to response style factors. Both the Rossi et al. (2001)

and Johnson (2003) models require tailor–made MCMC procedures for parameter estimation.

Two types of methods for investigating response styles have been proposed within the CFA or SEM framework, which is more accessible to applied researchers than IRT modeling. The first approach uses multiple-group CFA techniques (Byrne 1989; Byrne, Shavelson, and Muthen 1989), sometimes referred to as multiple-group LISREL modeling (Joreskog 2005). Rather than specifying a latent variable model with a response style factor as displayed in Figure 1, we use a model with content factors only. The aim is not to measure ERS but to check whether differential response styles distort the comparison of attitudes across groups. When item intercepts and factor loadings are invariant across groups, it is argued that the group comparison is not biased by differential response style effects (Van de Vijver 1998; Van de Vijver and Tanzer 1997). As Cheung and Rensvold (2000) show in a simulation study, differential ERS across groups results in noninvariant factor loadings (larger loadings for the more extreme group), and it may also affect item intercepts. This multiple-group SEM approach is useful if we wish to check whether group comparisons are invalidated by a differential response style. One limitation of this approach is, however, that it is a rather indirect way to deal with response styles: noninvariant intercepts and loadings may also be caused by factors other than a differential response style. Another limitation is that it cannot be used to measure or correct for differential response styles.

A second, very different use of CFA for the investigation of response styles involves the inclusion of a response style as a separate latent variable (factor) that directly affects the observed variables (items), in addition to the content-related latent factors (Billiet and McClendon 2000). The basic idea is that controlling for response style in attitude measurement requires the simultaneous specification of a response style factor and at least one substantive factor, the latter being measured by a balanced set of items. Our model depicted in Figure 1 is in agreement with the approach of Billiet and McClendon (2000), in which two related attitudes and one style factor measuring acquiescence are distinguished. We included two weakly related attitudes because the validity of the measurement of the response style increases when it occurs across items that are weakly related, or unrelated: The association between the items measuring unrelated substantive dimensions can be explained only by the response style factor. At the individual level, this

TABLE 1(a)
Pairwise Response Combinations that Are More Likely for Two Items Measuring the Same Attitude (with the Value of the Attitude Enclosed in Parentheses)

	Totally Disagree	Disagree	Neither Agree nor Disagree	Agree	Totally Agree
Totally disagree	X (−)				
Disagree		X (−)			
Neither agree nor disagree			X (0)		
Agree				X (+)	
Totally agree					X (+)

means that respondents who are subject to ERS are more likely to select the extreme response categories in both item subsets, controlling for his or her true scores on the two substantive dimensions.

Billiet and McClendon (2000) and Welkenhuysen-Gybels, Billiet, and Cambré (2006) used this SEM-based model for measuring and correcting for acquiescence. Although a conceptually similar approach could be used for detecting ERS, there is one fundamental reason why the structural equation approach has not been applied for this purpose: ERS has a nonmonotone effect on item responses. Whereas factor analysis assumes a linear (and thus monotonic) relation between latent variables and item responses—a higher factor score induces a higher response[3]—the influence of ERS is nonmonotonic in the sense that a higher ERS score increases the response probabilities for both the lowest and the highest category. The two-way tables that follow clarify the difference between a monotonic and a nonmonotonic pattern by showing how these patterns impact the association between two items.

Tables 1(a) and 1(b) show the dominant association pattern between two items arising from a shared attitude factor and ERS, respectively. The Xs indicate that combinations of responses can be expected to be more likely than responses that are independent, and the symbol enclosed in parentheses indicates whether these responses are given by persons with a low (−), middle (0), or high (+) attitude/ERS score. Table 1(a) shows that for items measuring the same underlying construct, cell frequencies on the diagonal of the table can be expected

[3]A higher score on the latent factor will induce a lower item score when the loading is negative.

TABLE 1(b)
Pairwise Response Combinations that Are More Likely When Both Items Are Affected by ERS (with the Value of the ERS Factor Enclosed in Parentheses)

	Totally Disagree	Disagree	Neither Agree nor Disagree	Agree	Totally Agree
Totally disagree	X (+)				X (+)
Disagree		X (−)		X (−)	
Neither agree nor disagree			X(0)		
Agree		X (−)		X (−)	
Totally agree	X (+)				X (+)

to be larger, with respondents having a negative value on the attitude dimension scoring low (*disagree* or *totally disagree*) on both items and respondents with high values scoring high (*agree* or *totally agree*). Table 1(b) illustrates the very distinctly different pattern arising from the nonmonotonic effect of an ERS factor: Cell frequencies for combinations of two extreme responses (irrespective of their direction) are larger because these are selected by individuals with positive scores on the ERS factor, and cell frequencies for two nonextreme responses are larger because these are selected by individuals with negative scores on the ERS factor. This means that when responses are affected by ERS, the association pattern of two items measuring the same attitude will be a mixture of these patterns shown in Tables 1(a) and 1(b). The association between two items measuring different dimensions will be of the form of Table 1(b), though in the case of correlated dimensions it may also be a mixture between 1(a) and 1(b), but the importance of 1(a) will be much less than for items measuring the same dimension.

The nonmonotone association implies that the relationship between the latent variable representing the response style and the item responses will be U-shaped (or even more complex) in the item. Specification of such a relationship requires using either complex nonlinear terms or treating items as nominal rather than ordinal/interval measurements. It will be clear that this is not possible within a standard SEM-framework that relies on linear relations and interval (or ordinal) level measurements (Joreskog 1994, 2005). Therefore, the structural equations approach where the response style is included in the SEM model as a separate latent variable cannot be applied to the case of ERS.

3. DETECTION OF ERS BY LATENT CLASS FACTOR ANALYSIS

Moors (2003) developed an SEM-like model for dealing with ERS using the latent class factor analysis (LCFA) approach proposed by Magidson and Vermunt (2001). The key contribution of Moors' study is that it resolves the problem of the standard SEM-approach discussed above; that is, it allows defining a U-shaped relationship between the latent ERS factor and the item responses. Using an empirical example, Moors showed that ignoring ERS may yield latent attitudinal factors that are seriously confounded with ERS. This emphasizes the usefulness of latent class factor analysis (LCFA) and the importance of correction for ERS.

The main differences between latent class analysis (LCA), IRT, and CFA/SEM concern the assumptions about the measurement levels of the item responses and the latent variable(s). In LCA and IRT the observed responses can be assumed to be measured at a nominal instead of an interval or ordinal level, as in CFA (Heinen 1996; Skrondal and Rabe-Hesketh 2004). Rather than analyzing a data set summarized in the form of a covariance matrix and a mean vector, LCA and IRT use the original response patterns that are typically summarized in a multidimensional frequency table. As was already indicated above, being able to treat the items as nominal makes it possible to detect that some respondents are more likely to choose the extreme categories in both directions, controlling for their true opinions.

Whereas in SEM (as in IRT) the latent variables are assumed to be normally distributed continuous variables, they are either specified as nominal in standard LCA or as ordinal in LCFA. The LCFA model proposed by Magidson and Vermunt (2001) is actually a variant of latent class analysis with multiple ordinal latent variables. Similar to factor analysis, it can be used in a more exploratory way or, as we do here, in a confirmatory way. It should be noted that the distinction between discrete latent variables with ordered categories (LCFA) and continuous latent variables (SEM or IRT) is not fundamental for the detection of ERS. In fact, the model we propose can be tested within the IRT as well as the LCA framework, the only difference being the assumed measurement level for the latent variables. A model similar as to the one proposed by Moors (2003) could also be defined using continuous latent variables—that is, as a multidimensional variant of an IRT model called the nominal response model (Bock 1972). Such a model could even be estimated with the same software used by Moors and in this paper—that is, by defining the latent variables to be continuous instead of ordinal (see also the appendix).

The LCFA model is graphically presented in Figure 1. We denote the scores of person i on the substantive factors by F_{1i} and F_{2i} and on the ERS factor as E_i. The response of individual i to rating item j is denoted by Y_{ij}, a particular response by c, and the number of response categories by C. Whereas standard factor analysis involves defining linear regression models for the items with the latent factors as predictors, Moors' LCFA model for ERS involves defining multinomial logistic regression models for the item responses with F_{1i}, F_{2i}, and E_i as predictors. Since the assumed distribution for the latent variables does not alter the model part for the item responses, we define it without explicitly specifying whether the latent variables are continuous or discrete. We show below how the latent variables can be modeled as discrete interval variables, as suggested by Moors (2003). This is the relevant regression equation for Y_{ij}:

$$P(Y_{ij} = c|F_{1i}, F_{2i}, E_i) = \frac{\exp(\beta_{0jc} + \beta_{1jc}F_{1i} + \beta_{2jc}F_{2i} + \beta_{3jc}E_i)}{\sum_{d=1}^{C} \exp(\beta_{0jd} + \beta_{1jd}F_{1i} + \beta_{2jd}F_{2i} + \beta_{3jd}E_i)}. \quad (1)$$

The β parameters are the item parameters to be estimated: β_{0jc} is an intercept term, β_{1jc} and β_{2jc} are slope parameters corresponding to the substantive factors, and β_{3jc} is a slope parameter for the ERS factor. The index j expresses that parameters may differ across items. As is typical in multinomial logistic regression models, each category of the item concerned has its own set of parameters, which is expressed by the index c (Agresti 2002). For identification purposes, the parameters should be fixed to 0 for one category or be restricted to sum to 0 across response categories. We used the latter constraint, which is often referred to as effect coding. Note that the ERS model for the ten items depicted in Figure 1 assumes that the first five items are not related to F_{2i}, which means that their β_{2jc} parameters are assumed to be equal to 0. Likewise, the last five items are assumed to be unrelated to the first substantive factor.

The desired interpretation of the latent substantive factors is that the higher a respondent's position on the latent dimension concerned, the more likely it is that he or she gives a high response (or a low response for a reversed formulated item). Such an interpretation is valid in the model defined in equation (1) if the β parameters for the substantive factor increase (decrease) monotonically across response categories. The ERS dimension measures the extent to which a respondent prefers the extreme answers relative to the other respondents in the sample. Thus, a higher score on the ERS dimension means that a person is more likely to give an extreme response than another person with the same value on the content factor. We stress that a low score on the ERS dimension does not necessarily imply an absence of ERS but instead indicates the opposite tendency—that is, a larger preference of nonextreme answers compared to other respondents. The interpretation of the ERS factor is valid if the extreme answer categories (for example, categories 1 and 5 on a five-point scale) have positive β_{3jc} values, but the nonextreme categories (for example, categories 2 and 4) and possibly also the middle categories have negative values. The larger the β_{3jc} values, positive and negative, the stronger the items concerned are affected by ERS. This illustrates clearly that the interpretation of the style factor is always post hoc; that is, it is based on the pattern of estimated values of the item- and category-specific parameters for the style factor. Since these parameters are not restricted, it is possible that we find a response style other than ERS—for instance, acquiescent response style (ARS). Similar to the attitude, ARS would correspond with positive values for the agree categories and negative values for the disagree categories. To distinguish between ARS and the attitudes, a balanced set of items is required because both ARS and the attitude affect the item responses linearly. To distinguish between a positive attitude and ERS, such a balanced set is not required as—in contrast to the attitude—the category item-parameters are affected by ERS in a nonmonotone manner. Nevertheless, a balanced set of items could increase the validity of ERS measurement as it allows differentiating between positive ERS (totally agree) and negative ERS (totally disagree) (see Harzing 2006).

As explained above, the modeling of the effect of ERS requires that items are treated as nominal response variables, as is done in equation (1). However, this requirement does not apply to substantive factors. Note again that a valid interpretation of these factors requires that their β parameters are monotonically increasing (decreasing) across response categories. So, in fact, it would be more natural to treat responses as ordinal in their relationship with the content factors—that is, to impose restrictions which guarantee a monotone relationship between F and Y. This can be achieved by means of an adjacent-category ordinal logit specification, which is also used in IRT models for rating items, as in the partial credit model (Masters 1982).

The specification of such a restricted model for ERS is possible because an adjacent-category logit model is a restricted multinomial logit model (Agresti 2002). More specifically, we assume that

$$P(Y_{ij}=c|F_{1i}, F_{2i}, E_i) = \frac{\exp(\beta_{0jc}+\beta_{1j}c\, F_{1i}+\beta_{2j}c\, F_{2i}+\beta_{3jc}\, E_i)}{\sum_{d=1}^{C}\exp(\beta_{0jd}+\beta_{1j}\, d\, F_{1i}+\beta_{2j}\, dF_{2i}+\beta_{3jd}\, E_i)}. \quad (2)$$

The imposed constraints are $\beta_{1jc} = \beta_{1j}c$ and $\beta_{2jc} = \beta_{2j}c$, which automatically guarantee that the implied β_{1jc} and β_{2jc} are monotone in c. The parameters for the ERS factor remain unchanged compared to equation (1). This hybrid ordinal-nominal regression model can also be written as a linear model for the logit of responding in category c+1 instead of c; that is,

$$\log \frac{P(Y_{ij} = c + 1|F_{1i}, F_{2i}, E_i)}{P(Y_{ij} = c|F_{1i}, F_{2i}, E_i)} = (\beta_{0jc+1} - \beta_{0jc}) + \beta_{1j}F_{1i} + \beta_{2j}F_{2i} + (\beta_{3jc+1} - \beta_{3jc})E_i. \quad (3)$$

This equation shows how the various model parameters are related to the adjacent-category logits. The β_{1j} and β_{2j} parameters are thus effects on the adjacent-category logits. Note that the effect of the ERS factor on the adjacent category logit ($\beta_{3jc+1} - \beta_{3jc}$) should be negative when comparing categories 2 and 1 and positive when comparing categories 5 and 4, assuming we have a 5-point scale. The same model, but now formulated in term of odds instead of logits, is as follows:

$$\frac{P(Y_{ij} = c + 1|F_{1i}, F_{2i}, E_i)}{P(Y_{ij} = c|F_{1i}, F_{2i}, E_i)} = \exp(\beta_{0jc+1} - \beta_{0jc})\exp(\beta_{1j})^{F_{1i}}\exp(\beta_{2j})^{F_{2i}}\exp(\beta_{3jc+1} - \beta_{3jc})^{E_i} \quad (4)$$

These exponentiated parameters are the ones that typically will be interpreted.

One advantage of this more restricted specification compared to the one proposed by Moors (2003) is that it is more parsimonious. Rather than $C - 1$ parameters for each item, only one parameter has to be estimated to capture the influence of the attitude on an item response. This single parameter is similar to a factor loading in standard factor analysis. A second advantage is that the relationship between content factors and the responses are forced to be monotone, which gives the model structure a clearer distinction between the ERS factor on the one hand and the content factors on the other hand. The restriction imposed in equations (2) can be tested by comparing the fit of this model with the fit of the unrestricted model of equation (1).

We will show below that the ERS factor specification can also be similarly be restricted using scores for the response categories—for example, scores with a W-shaped or U-shaped pattern. A U-shaped pattern can be obtained, for example, using scores 1.5, −1, −1,−1, and 1.5 or equivalently 1, 0, 0, 0, and 1.[4] We will specify such models to investigate the robustness of the results obtained with an unrestricted ERS factor. Until now, we did not provide any details about the specification of the latent variables in the proposed ERS model. One option is to assume that these are continuous normally distributed variables, in which case the model estimation by maximum likelihood involves the numerical approximation of a three-dimensional integral. Another option, also used by Moors (2003), is to treat the latent variables as discrete variables with a few (e.g., three) ordered categories. Such a discrete specification with ordered classes can be perceived in two different ways: We can firmly believe that there are three classes or we can see it as a way to approximate a continuous distribution with an unknown (possibly non-normal) form. Some authors refer to the latter as a semiparametric or nonparametric specification of the distribution of a latent variable (Heinen 1996; Skrondal and Rabe-Hesketh 2004). In fact, an ordinal specification is more flexible than a continuous specification because no unverifiable distributional assumptions are made. The latent classes are merely assumed to represent three points with equal distances on an

[4]The model remains unchanged when applying this same linear transformation of each of the scores; adding 1 and dividing by 2.5 does not change the model.

underlying—possibly continuous—dimension, which is achieved by assigning scores to the classes of -1, 0 and 1 (in the case of three classes). A larger number of classes could be used to position the respondents more accurately on the ERS dimension; however, in our analysis this does not alter the conclusions with regard to ERS. Another specification issue related to the latent variables is that they can be regressed on covariates—for example, on a set of dummies for the cultural group one belongs to (see Ethnicity in Figure 1). The regression model used for the ordinal latent variables is also an adjacent-category ordinal logit model.

Figure 1 shows that the substantive factors are allowed to be correlated with one another but that the ERS factor is assumed to be uncorrelated with the substantive factors. We make the latter assumption because it seems to be logical in most applications; that is, usually there is no reason to assume that a person's response style is correlated with the substantive dimensions to be measured. It should, however, be noted that it is not a problem to relax this assumption. We can simply include the associations between the substantive factors and the style factor in the model, which will have little impact on the measurement part of the model.

4. APPLICATION

The data used to illustrate the ERS model described in the previous section come from the Dutch survey named SPVA (see footnote 1), which was repeated every four years from 1988 until 2002. For our study, we use the data collected in 2002 among the four largest ethnic minorities in the Netherlands—namely, Turks, Moroccans, Surinamese, and Antilleans. Note that only the answers of the heads of the households are included in the analyses to secure independent observations. Response rates lie between 44% for Surinamese and Antilleans and 52% for Turks (see Table 2). Among the topics treated in the survey are family values, work values, religion, women's emancipation, work, and education (Dagevos, Gijsberts, and Van Praag 2003). In this application, we use two sets of five questions, each subset referring to a cultural dimension; that is, *attitude toward the Dutch society* and beliefs about the *autonomy of the children within the family*. The respondents were asked to report on a fully labeled 5-point Likert scale, ranging from *totally agree* (1) to *totally disagree* (5), with *neither agree nor disagree* as a neutral midpoint. For the statistical analyses, the category order was reversed in order to facilitate the interpretation of scale that now runs from a negative (1) toward a positive (5) response to the item.

Table 2 reports the means of the ten items for each of the four ethnic groups. While these groups are fairly similar when it comes to their attitude toward the Dutch society, Turks and Moroccans are slightly more positive—on average—about the autonomy of the children compared to Surinamese and Antilleans. Note that a high score on the attitude toward the autonomy of the children actually means that they have little tolerance toward children making their own decisions. However, to reach more valid conclusions about the differences between these groups, confounding factors, such as differential ERS, should be controlled for since they may have biased the measurement of attitudes.

We estimated various LCFA models using the SPVA data. For this purpose we used the syntax module of the Latent GOLD 4.5 program[5] (Vermunt and Magidson 2008), a program for the maximum likelihood estimation of latent class models and other types of latent variable models. Table 3 reports the log-likelihood and BIC values for the most relevant models. BIC can be used to compare models with one another: The lower the BIC values the better the model is in terms of fit and parsimony. It should be noted that several of the estimated models are nested; for example, models with and without ERS factor are nested when the remaining part is the same. The former can be obtained from the latter either by fixing the β parameters for the ERS factor to 0 or by reducing the number of categories of the ERS factor to 1. However, as is known from model selection in latent class and mixture modeling, models with different numbers of classes cannot be compared using an asymptotic likelihood-ratio test because certain regularity conditions have not been met (McLachlan and Peel 2000). A possible way out would be to use likelihood-ratio tests with bootstrap p values, which are, however, rather computationally intensive procedures. We thus decided to use only BIC for model selection, which is the most common procedure in latent class analysis.

[5]See the appendix for details on model specification using the syntax module of the Latent GOLD 4.5 program.

TABLE 2
Mean Observed Item Response Per Ethnic Group (N = 3574)

	Turks	Moroccans	Surinamese	Antilleans
Factor 1: Attitude Toward Dutch Society				
Item 1: In the Netherlands immigrants get many opportunities	3.53 (1.058)	3.42 (1.075)	3.26 (1.106)	3.24 (1.148)
Item 2: The Netherlands is hostile to immigrants[a]	2.80 (1.015)	2.47 (0.879)	2.40 (0.880)	2.52 (0.906)
Item 3: In the Netherlands your civil rights as an immigrant are respected	3.40 (0.905)	3.55 (0.857)	3.52 (0.862)	3.44 (0.843)
Item 4: The Netherlands is a hospitable country for immigrants	3.03 (0.971)	3.47 (0.915)	3.69 (0.885)	3.60 (0.908)
Item 5: The Netherlands is tolerant toward foreign cultures	3.83 (0.909)	3.57 (0.872)	3.84 (0.815)	3.69 (0.830)
Factor 2: Autonomy of the Children				
Item 6: Children should live at home until marriage	3.69 (1.049)	3.76 (1.120)	2.94 (1.272)	2.59 (1.233)
Item 7: Elderly should be able to move in with their children	3.13 (1.129)	3.79 (0.965)	3.10 (1.153)	3.01 (1.175)
Item 8: Adult children should be able to move in with their parents	3.88 (0.884)	3.94 (0.852)	3.32 (1.081)	3.14 (1.110)
Item 9: Parents should always be respected, even if they do not deserve it based on their behavior or attitude	3.11 (1.147)	3.36 (1.127)	2.86 (1.115)	2.88 (1.092)
Item 10: Older family members should have more influence in important decisions (for instance, about moving) than younger ones	4.11 (0.830)	4.21 (0.890)	3.61 (1.111)	3.70 (1.094)
N	914	862	1016	782
Response Rate (%)[b]	0.52	0.52	0.44	0.51

Note: Standard errors are shown in parentheses.
[a] This item has a reversed formulation.
[b] The response rate excludes those who were not at home, refused, or otherwise were unavailable (see DANS 2005: 44).

TABLE 3
Goodness of Fit Statistics for Latent Class Factor Models (N = 3574)

Model	Log-Likelihood	BIC	Number of Parameters
(A) Null model	−47616.6	95560.4	40
(B) One-factor model	−44513.0	89696.8	82
(B_1) Model B + style factor	−43032.7	87079.9	124
(B_2) Model B + ordinal specification	−46196.9	92835.6	54
(B_3) Model B + ordinal specification + style factor	−43196.7	87162.5	94
(C) Two-factor model	−43910.2	88515.8	85
(D) Model C + style factor	−42233.8	85506.6	127
(E) Model C + ordinal specification	−45008.3	90466.6	55
(F) Model C + ordinal specification + style factor	−42338.0	85469.6	97

Models A, B, and C are models with 0, 1, and 2 substantive factors but without a style factor. Note that the null model (Model A) assumes that item responses are independent of one another. Based on the BIC values, it can be seen that a two-factor model outperforms a one-factor model, which is, of course, in agreement with what could be expected given the content of the items. In Model D the style factor is included, and finally in Models E and F the items are treated as ordinal in relation to the substantive factors, with Model F also including a style factor. The analyses in Models D, E, and F are repeated in Models B_1, B_2, and B_3, which contain only one substantive factor.

Inspection of the β_{1jc} and β_{2jc} parameters (loadings) obtained with Models C—not shown here—pointed out that the relationships between the items and the two factors are not monotonic, as is required for a valid interpretation of the substantive factors. In fact, the loadings were more in agreement with the type of U-shaped pattern corresponding to an ERS factor: positive values for the lowest and highest categories and negative values for the other three categories. Such a pattern is more likely to be associated with a response style such as ERS than an attitude and led us to conclude that the factors that were supposed to measure substantive content are confounded with ERS. Not surprisingly, the inclusion of an additional factor measuring ERS improves the model fit considerably, as can be seen by comparing the BIC values of Models C and D. Moreover, the β_{1jc} and β_{2jc} coefficients of Model D show a monotone pattern: They increase or decrease—depending on the item formulation—along the response scale. These results show that controlling for ERS ensures an interpretation of the two content factors as could be expected.

As a last step, we specified the more restricted variant of Moors' ERS model described in equation (2); that is, the items were treated as ordinal instead of nominal in their relationship with the substantive latent variables. Whether such ordinal restriction is appropriate when controlling for ERS is confirmed by the monotone pattern in the multinominal logit coefficients in Model D. Lastly, we could check the appropriateness of the ordinality assumption in Model F by comparing the BIC value of Model F with Model D, which shows that the model with the linearity restriction on the category-specific loadings is the one that should be preferred. Note that the ordinal restriction deteriorates the model without a correction for ERS (compare the BIC of Model C and Model E) due to the presence of the nonmonotone pattern that is caused by ERS.

To check whether the style factor is not just absorbing misspecifications of the substantive dimensions (for example, that the cross loadings are wrongly assumed to be equal to 0), we estimated a series of models similar to Models D, E, and F but with only one substantive factor. These three variants of Model B are called Model B_1, B_2, and B_3, respectively. As can be seen, according to the BIC criterion, Models B_1, B_2, and B_3 fit much worse than Models D, E, and F, which shows that we really need two substantive factors in addition to a style factor. This is confirmed by the parameter estimates for the ERS factor in Models B_1 and B_3, which show an ERS pattern and not a pattern corresponding to a substantive dimension.

Table 4 reports the β_{1j}, β_{2j}, and β_{3jc} parameters obtained with Model F. As can be seen, for the two substantive factors we have one parameter per item and for the response style factor we have five parameters (which sum to 0). For the interpretation of these β parameters it is important to note that the latent variables are specified to have three ordinal categories scored as -1, 0, and 1. Since the logit parameters are effects of a one-point change in the latent variable, these parameters correspond to a shift from class 1 to class 2 or from class 2 to class 3. For the substantive factor the classes correspond with a negative, neutral, and positive attitude respectively. The three ERS classes can be labeled low, middle, and high (see also the discussion below).

TABLE 4
Parameter Estimates Obtained with Model F, Containing the Two Content Factors and One Style Factor (N = 3574)

	Factor 1: Attitude Toward Dutch Society	Factor 2: Autonomy of Children	Factor 3: ERS Factor				
			TD	D	N	A	TA
Item 1	1.03		1.70	−1.32	−0.39	−1.13	1.13
	(0.06)		(0.12)	(0.09)	(0.08)	(0.09)	(0.16)
Item 2	−1.03		1.47	−0.98	−0.56	−1.19	1.26
	(0.06)		(0.18)	(0.09)	(0.07)	(0.08)	(0.12)
Item 3	2.68		2.19	−1.81	−1.26	−1.39	2.26
	(0.17)		(0.22)	(0.13)	(0.12)	(0.14)	(0.34)
Item 4	2.11		1.38	−1.46	−0.67	−1.18	1.93
	(0.13)		(0.16)	(0.10)	(0.09)	(0.11)	(0.24)
Item 5	1.29		1.30	−1.46	−0.57	−0.73	1.47
	(0.08)		(0.12)	(0.10)	(0.09)	(0.11)	(0.24)
Item 6		1.50	1.87	−1.46	−0.25	−1.26	1.10
		(0.12)	(0.13)	(0.10)	(0.10)	(0.10)	(0.12)
Item 7		1.25	1.89	−1.30	−0.45	−1.36	1.22
		(0.09)	(0.12)	(0.09)	(0.08)	(0.09)	(0.14)
Item 8		1.51	1.91	−1.40	−0.45	−1.42	1.36
		(0.12)	(0.13)	(0.09)	(0.09)	(0.11)	(0.19)
Item 9		0.98	1.13	−1.50	−0.27	−0.89	1.54
		(0.08)	(0.10)	(0.09)	(0.12)	(0.11)	(0.20)
Item 10		0.74	1.63	−1.21	−0.37	−1.31	1.27
		(0.05)	(0.12)	(0.08)	(0.07)	(0.08)	(0.12)

Note: Standard errors are shown in parentheses. All parameters shown here are significantly different from 0 at $p < .05$. Item category labels are denoted by TD (totally disagree), D (disagree), N (neither agree nor disagree), A (agree), and TA (totally agree).

When ordinally restricted as in Table 4, the β coefficients are most easily interpreted in terms of effects on the adjacent category odds ratios (see equation 4). For example, a one-point change in the latent factor measuring the attitude toward the Dutch society increases the odds of choosing category $c + 1$ rather than category c by a factor 3, exp (1.03), for the first item. It can be seen that there are large differences across items in the strength of their association with the substantive factors.

The category-specific β parameters belonging to the ERS factor show the expected nonmonotone pattern: The higher the respondents' ERS scores, the more likely they are to select the outer categories and the less likely they are to select the other categories. The style factor has a large effect on the item responses that can be seen by computing its effects on the odds of choosing *totally agree* over *agree* or choosing *totally disagree* over *disagree.* For the first item, these odds increase by a factor 20 and 10 [exp(1.70 + 1.32) and exp(1.13 + 1.13)], respectively, when we change from one class to the next. Thus, the higher respondents stand on the ERS dimension, the (much) more likely they are to choose *totally agree* (*totally disagree*) instead of *agree* (*disagree*). We emphasize that this result is *given the substantive factors*, meaning that this respondent selects these categories more often than would be expected on the basis of his or her attitude.

The parameter estimates of Model F confirm that the style factor is indeed an ERS factor. However, we did not indicate *a priori* that the parameters should have the specific structure corresponding to ERS. To investigate the robustness and validity of the encountered ERS factor, we will compare Model F with models using more restricted specifications for the ERS factor. Moreover, we will check the validity of our ERS factor by comparing it with ERS scores obtained using two other methods described in our overview—that is, with an ERS index and an IRT-based ERS score using all 55 rating items from the SVPA survey.

Restricted variants of Model F in which the β parameters for the relationship between the ERS factor and the responses are specified to have W-shaped or U-shaped patterns can be obtained in a similar way as the ordinal models for the content factors—that is, by using prespecified scores for the categories of response variables. A W-shaped pattern (Model F_1) is obtained using category scores 1, -1.5, 1, -1.5, and 1, and a U-shaped pattern (Model F_2) using scores 1.5, -1, -1, -1, and 1.5. These two specifications differ in the treatment of the middle category, which is either assumed to be similar to the outer or the inner categories as far as the relationship with the style factor is concerned. As can be seen from the fit measures reported in Table 5, both Model F_1 and Model F_2 are fit worse fits than the unrestricted Model F, showing that the restriction of the midpoint category parameter to exactly equal the outer or inner category parameters is too strong. However, based on the fact that Model F_2 fits better than Model F_1, it can be concluded that the style factor is better approximated by a U-shaped pattern of category parameters than a W-shaped pattern. We also estimated a model using category scores 1.25, -1, -0.5, -1, 1.25 (Model F_3) in which the middle category is assumed to be similar to inner categories

TABLE 5
Fit Measures for Variants of Model F Using a Restricted Specification for the Style Factor (N = 3574)

Model	Log-Likelihood	BIC	Number of Parameters
(F) Model F	−42338.0	85469.6	97
(F_1) Model F + W-shaped pattern	−43126.0	86800.2	67
(F_2) Model F + U-shaped pattern	−42535.9	85619.9	67
(F_3) Model F + W-U-shaped pattern	−42434.3	85416.9	67

Note: The W-shaped pattern is obtained using category scores 1, −1.5, 1, −1.5, and 1; the U-shaped pattern with scores 1.5, −1, −1, −1 and 1.5, and the W-U shaped with scores 1.25, −1, −0.5, −1, and 1.25.

but not identical. According to BIC, this very parsimonious model should be preferred over the unrestricted Model F.

Using the results of our LCFA model, it is possible to compute an ERS score for each individual in the sample (these are posterior mean estimates). As indicated in our overview, there are also other methods to compute ERS scores, two of which are the ERS index and the IRT-based ERS score. We recoded all rating items of the SPVA survey (55 in total) as 0 (nonextreme response) and 1 (extreme response). The ERS index is simply the proportion of items with an extreme response.[6] Moreover, we estimated a unidimensional IRT model using these 55 dichotomous items, and computed IRT-based ERS scores.[7] The LCFA-based ERS score (using Model F) correlates strongly to the ERS index (.81) and IRT-based ERS score(.76). The fact that these scores based on 55 items correlate highly with our ERS score demonstrates the validity of our procedure. The ERS score based on Model F also correlates highly with the scores based on the restricted models F_1, F_2, and F_3—that is, .88, .95, and .99, respectively. This shows that the

[6]This ERS index is similar to the index discussed in the overview in Section 2. The proportion is based on the items without missing values.

[7]This is a slightly simplified version of the IRT modeling approach proposed by De Jong et al. (2008), as we do not account for the fact that despite recoding into 0 and 1, items measuring the same substantive dimension may still be more strongly related to one another. However, the style factor turned out to capture 93.3% of the interitem associations, showing that the remaining associations are not very large. The IRT model was estimated using ML with the missing values.

proposed procedure is robust toward the specification used for the ERS factor.

In the literature, different meanings are attached to the dimension underlying an extreme response style factor (Greenleaf 1992b). Some characterize the dimension as representing the tendency to select extreme responses (for instance, see De Jong et al. 2008); others start from the point of view that the dimension describes the dispersion of responses around the center of the response scale (for instance, see Baumgartner and Steenkamp 2001). Both argue that one end point corresponds to a response pattern containing many extreme responses and signifies "*strongly affected by ERS*". In our view, the conceptualization of the other end point depends on the operationalization of ERS. In the sum-score method, where one simply counts the number of extreme responses, the opposite end point of the dimension represents response patterns with few extreme responses—that is, with the tendency to prefer the nonextreme categories *agree*, *disagree*, or *neither agree nor disagree*.

Table 6 reports the probabilities of belonging to each of the three ERS classes (based on Model F) given the number of responses in the extreme, adjacent, and middle categories, respectively. As expected, the class membership probabilities conditional on the number of extreme responses show that the smaller this number, the more likely a respondent belongs to class 1; the larger this number, the more likely a respondent belongs to class 3. For the number of responses in the adjacent categories, the opposite pattern occurs: Many such responses make it more likely to belong to the first class while few of them make it more likely to belong to the third class. As far as the number of responses in the middle categories is concerned, it can be observed that the larger this number, the more likely that a respondent belongs to the second class of the ERS factor. These findings seem to indicate that the ERS dimension picks up both the tendency to select as well as to avoid extreme responses, irrespective of the respondent's attitude. However, more research is needed to confirm whether this interpretation of the ERS factor is useful and valid in other situations.

One purpose of our research was to investigate the attitude differences between the four ethnic groups as well as how these differences are confounded by differential response styles. In Table 7, every model mentioned in Table 2 includes ethnicity dummies as predictors in the regression equations for the latent variables. The fact that the likelihood

TABLE 6
Membership Probabilities for the Three ERS Classes Given the Number of Extreme, Middle, and Adjacent Category Responses Obtained with Model F (N = 3574)

	Extreme Responses				Midpoint Responses				Adjacent Responses			
	ERS Class				ERS Class				ERS Class			
Number	1	2	3	N	1	2	3	N	1	2	3	N
0	0.82	0.18	0.00	1605	0.54	0.32	0.13	1113	0.00	0.07	0.93	76
1	0.44	0.56	0.00	587	0.51	0.38	0.10	810	0.00	0.17	0.82	111
2	0.09	0.90	0.01	424	0.46	0.43	0.11	610	0.00	0.37	0.63	153
3	0.01	0.96	0.03	283	0.39	0.50	0.11	449	0.02	0.62	0.36	208
4	0.00	0.85	0.15	192	0.33	0.59	0.07	264	0.04	0.83	0.13	258
5	0.00	0.64	0.36	164	0.27	0.68	0.05	173	0.14	0.82	0.04	388
6	0.00	0.28	0.72	98	0.10	0.83	0.06	83	0.26	0.74	0.01	417
7	0.00	0.07	0.93	89	0.07	0.90	0.02	48	0.42	0.58	0.00	494
8	0.00	0.01	0.99	71	0.03	0.94	0.02	19	0.66	0.34	0.00	495
9	0.00	0.00	1.00	35	0.02	0.98	0.01	3	0.89	0.11	0.00	490
10	0.00	0.00	1.00	26	0.01	0.98	0.02	2	0.98	0.02	0.00	484
Overall	0.45	0.44	0.11	3574	0.45	0.44	0.11	3574	0.45	0.44	0.11	3574

TABLE 7
Goodness of Fit Statistics for Latent Class Factor Models with Ethnicity Included as a Covariate in Every Model (N = 3574)

Model	Log-Likelihood	BIC	Number of of Parameters
(A_g) Null model	−47616.6	95560.4	40
(B_g) One-factor model	−44482.9	89661.3	85
(C_g) Two-factor model	−43815.4	88399.8	94
(D_g) Model C_g + style factor	−41850.4	84838.0	139
(E_g) Model C_g + ordinal specification	−44699.1	89921.9	64
(F_g) Model C_g + ordinal specification + style factor	−41950.6	84793.0	109

values of all models in Table 7 show a significant improvement of the fit of the models in Table 2 indicates that ethnicity is indeed associated with the (supposed) substantive dimensions.

Table 8 reports the logit coefficients for the ethnicity dummies in the regression models for the latent factors as obtained with Model C_g, Model D_g, Model E_g, and Model F_g (the subscript *g* stands for group). Note that the parameters for Turks are fixed to 0, which means that this category serves as the reference category. A positive parameter value means that the group concerned is more likely to belong to a higher class than the Turkish respondents.

First, the encountered group differences in ERS show that Moroccans are somewhat more likely to use the extreme categories and Surinamese somewhat less likely than Turks. This differential ERS can only partially explain the encountered differences between the models with and without ERS. These are mainly the result of large, individual differences in response style existing within groups. Second, Table 8 illustrates that ERS suppresses the group differences somewhat and that the standard errors are smaller in Models C_g and E_g. Although not further investigated, this finding indicates that the ordinal specification used not only in our LCFA analyses but also in our multigroup SEM analyses removes the contamination of the items parameters by ERS. Nevertheless, a correction for ERS is preferable to avoid misspecifications and type II errors.

TABLE 8
Effect of Ethnicity in Model C_g, Model D_g, Model E_g and Model F_g (N = 3574)

	Ethnicity	Factor 1: Attitude Toward Dutch Society	Factor 2: Autonomy of Children	Correlation	Factor 3: ERS
Model C_g[a]	Turks	0.00	0.00	0.93	
				(0.13)	
	Moroccans	0.18	−0.33	0.38	
		(0.11)	(0.11)	(0.13)	
	Surinamese	0.82	1.28	0.89	
		(0.11)	(0.12)	(0.13)	
	Antilleans	0.46	1.33	0.66	
		(0.12)	(0.13)	(0.16)	
Model D_g[b]	Turks	0.00	0.00	−0.09	
				(0.13)	
	Moroccans	0.38	−0.52	0.42	0.10
		(0.11)	(0.12)	(0.15)	(0.08)
	Surinamese	1.08	1.67	0.57	−0.02
		(0.12)	(0.15)	(0.15)	(0.08)
	Antilleans	0.60	1.85	0.06	0.09
		(0.14)	(0.15)	(0.17)	(0.09)
Model E_g[c]	Turks	0.00	0.00	0.30	
				(0.12)	
	Moroccans	0.42	−0.52	0.37	
		(0.10)	(0.12)	(0.11)	
	Surinamese	1.11	1.32	0.66	
		(0.14)	(0.12)	(0.13)	
	Antilleans	0.59	1.81	0.24	
		(0.18)	(0.15)	(0.18)	
Model F_g[d]	Turks	0.00	0.00	−0.04	0.00
				(0.12)	
	Moroccans	0.42	−0.45	0.57	0.14
		(0.10)	(0.11)	(0.14)	(0.08)
	Surinamese	1.15	1.60	0.57	−0.16
		(0.15)	(0.15)	(0.17)	(0.08)
	Antilleans	0.70	1.87	0.19	−0.06
		(0.17)	(0.17)	(0.20)	(0.08)

Note: Standard errors are shown in parentheses.
[a] No style factor and nominal specification of items
[b] With style factor and nominal specification of items
[c] No style factor and ordinal specification of items
[d] With style factor and ordinal specification of items

5. DISCUSSION

The findings of Moors (2003, 2004) have been confirmed in our study. First, the response style factor turns out to affect the responses to such an extent that it invalidates substantive findings when not controlled for. This is due to the fact that the presence of ERS causes the items to be related to the supposed substantive factors in a nonmonotonic rather than a monotonic way. Second, when not controlled for, response style affects the encountered differences between culturally diverse groups. The inclusion of the style factor yields not only more valid substantive factors but also more valid conclusions with respect to the group differences on these factors. Third, we proposed the items to be ordinally restricted in their relation to the substantive factors but to remain unrestricted (nominal) in their relation with the style factor. This more parsimonious model turned out to be the preferred model specification in our application. Finally, we showed that the ordinal specification suppresses the influence of ERS on the items.

The ERS models discussed in this paper can be expanded in several interesting ways. The unrestricted style factor is able to detect not only a nonmonotone pattern caused by ERS but also a monotone pattern caused by other response styles such as the acquiescent response style (ARS). This unrestricted modeling approach can always be used to detect a response style even though the type of response style that can be detected may not be known beforehand. In this sense the method is exploratory. Similar to the association model (Goodman 1981), the category scores are estimated in Model F without assuming equal distances or order. Any kind of survey would permit the unrestricted approach; however, we believe that the W pattern particular to the extreme response style is most likely to be found in Likert scales (Chun, Campbell, and Yoo 1974; Cronbach 1950; Peabody 1962). If the unrestricted Model F should be applied to other survey designs, other response styles such as acquiescence can appear. Estimating models in which the parameters of the response style factor are restricted to a particular pattern (e.g., the W-shaped pattern) may be applicable in survey designs where knowledge of a particular response style may become available during the course of the study, as might occur in research studies using panel designs. For example, the unrestricted model may be used to detect a particular response style in a first wave of data collection, with more restricted models being tested in subsequent waves, given the findings of the unrestricted model in the first wave.

More than a single style factor can be incorporated in the model, but then the *post hoc* interpretation of the category-specific item parameters can no longer be used to label the multiple response style factors. Multiple style factors require a more confirmatory approach by imposing *a priori* restrictions on the response style parameters so that they are in agreement with a particular response style. For example, in the case of a 5-point scale, category scores with a U-shaped pattern could be used for an ERS factor (as in our Model F2) and monotonic category scores (-2, -1, 0, 1, and 2) for an ARS factor, with the additional restriction that the effect of the ARS factor should be positive irrespective of the positive or negative wording of the item concerned. Note that the modeling of ARS requires balanced item sets in order to be able to differentiate between ARS and substantive factors. Although in Likert type data the unrestricted style factor can detect ERS and ARS in balanced item sets, this modeling approach can be used across survey designs to detect other response styles.

Another possible extension is to allow (some of) the parameters of the measurement model to be group specific. The relationship between the items and the substantive factors can be made group specific, as can their relationship with the ERS factor. Another possible extension is the inclusion of additional predictors for which we would like to control the encountered ethnic group differences in the latent factors. Examples of such predictors are individual characteristics such as educational attainment, language proficiency, and age. A third possible extension is the integration of the proposed ERS model into a more general structural equation modeling framework in which one latent variable is used as a predictor of another latent variable.

In this study, we have illustrated the effect of an extreme response style on a response pattern of a Likert scale in general and, more specifically, on the validity of cross-cultural comparisons. The proposed ordinal restriction yields simpler models that do not have a worse fit and facilitate the interpretations of the model parameters. We recommend that survey researchers include an unrestricted style factor in their models for measuring attitudes in a more valid manner. In summary, we emphasize here the need for detecting and correcting for extreme response style in cross-cultural research.

APPENDIX: LATENT GOLD 4.5 SYNTAX USED FOR THE MOST COMPLEX MODEL

We used the syntax module of Latent GOLD 4.5 to estimate models A to F in Table 2 and Model A_g to F_g, in Table 5. The variables and equations sections of the syntax file for the most complex model F_g are as follows:

```
.
 variables
   dependent
     Y1 nominal, Y2 nominal, Y3 nominal, Y4 nominal,
     Y5 nominal, Y6 nominal, Y7 nominal, Y8 nominal,
     Y9 nominal, Y10 nominal;
   independent ethnicity nominal coding = first;
   latent
     F1 ordinal 3 scores = (−1 0 1),
     F2 ordinal 3 scores = (−1 0 1),
     ERS ordinal 3 scores = (−1 0 1);
 equations
   F1 <- 1 + ethnicity;
   F2 <- 1 + ethnicity;
   ERS <- 1 + ethnicity;
   F1 <-> F2 | ethnicity;
   Y1 -- Y5 <- 1 + (~ord) F1 + ERS;
   Y6 -- Y10 <- 1 + (~ord) F2 + ERS;
.
```

In the variables section we provide the relevant information on the dependent, independent, and latent variables to be used in the analysis: The dependent variables are nominal, the independent variable is nominal with the first category as the reference category (which overrides the default effect coding), and the latent variables are ordinal with the specified category scores. The first three equations define the regression models for the latent variables—which contain an intercept (indicated with "1") and an effect of ethnicity—and the fourth defines the association between F1 and F2 (which is allowed to vary across ethnic groups). The last two equations define the multinomial regression models for items Y1 to Y5 and Y6 to Y10, respectively. The term "(∼ord)" before F1 and F2 indicates that the nominal dependent variable concerned should be treated as ordinal in this term. As an alternative, we could define the items to ordinal instead of nominal and put "(∼nom)" before ERS to indicate that the ordinal items should be treated as nominal for these terms.

The other estimated models can easily be derived from this syntax example. For example, removing "+ ethnicity" and "| ethnicity" for the first four equations yields a model without ethnic group difference in the latent variables, removing "(∼ord)" yields a model in which the term concerned remains a standard multinomial logit term, and removing ERS from the latent variable definition and the equations yields a model without ERS factors.

REFERENCES

Agresti, Alan. 2002. *Categorical Data Analysis*. New York: Wiley.

Arce-Ferrer, Alvaro J. 2006. "An Investigation into the Factors Influencing Extreme-Response Style: Improving Meaning of Translated and Culturally Adapted Rating Scales." *Educational and Psychological Measurement* 66:374–92.

Bachman, Jerald G., and Patrick M. O'Malley. 1984. "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles." *Public Opinion Quarterly* 48:491–509.

Baumgartner, H., and Jan-Benedict E. M. Steenkamp. 2001. "Response Styles in Marketing Research: A Cross-National Investigation." *Journal of Marketing Research* 38:143–56.

Billiet, Jaak B., and McKee J. McClendon. 2000. "Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items." *Structural Equation Modeling* 7:608–28.

Bock, Darrell R. 1972. "Estimating Item Parameters and Latent Ability When Responses Are Scored in Two or More Nominal Categories." *Psychometrika* 37:29–51.

Byrne, Barbara M. 1989. "Multigroup Comparisons and the Assumption of Equivalent Construct Validity Across Groups: Methodological and Substantive Issues." *Multivariate Behavioral Research* 24:503–23.

Byrne, Barbara M., Richard J. Shavelson, and Bengt Muthen. 1989. "Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance." *Psychological Bulletin* 105:456–66.

Cheung, Gordon W., and Roger B. Rensvold. 2000. "Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research Using Structural Equation Modeling." *Journal of Cross-Cultural Psychology* 31:187–212.

Chun, Ki-Taek, John B. Campbell, and Jong Hae Yoo. 1974. "Extreme Response Style in Cross-Cultural Research. A Reminder." *Journal of Cross-Cultural Psychology* 5:465–80.

Clarke III, Irvine. 2001. "Extreme Response Style Cross-Cultural Research." *International Marketing Review* 18:301–24.

Cronbach, Lee J. 1950. "Further Evidence on Response Sets and Test Design." *Educational and Psychological Measurement* 10:3–31.

Dagevos, Jaco, Merove Gijsberts, and Carlo Van Praag. 2003. *Rapportage Minderheden 2003. Onderwijs, Arbeid en Sociaal-Culturele Integratie*. The Hague: The Netherlands Institute for Social Research.

DANS (Data Archiving and Networking Services). 2005. *Documentation SPVA-2002*. Retrieved November 30, 2010 (http://www.scp.nl/dsresource?&objectid=18720).

De Jong, Martijn G., Jan-Benedict E. M. Steenkamp, Jean-Paul Fox, and Hans Baumgartner. 2008. "Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation." *Journal of Marketing Research* 45:104–15.

Eid, Michael, Rolf Langeheine, and Ed Diener. 2003. "Comparing Typological Structures Across Cultures by Multigroup Latent Class Analysis. A Primer." *Journal of Cross-Cultural Psychology* 34:195–210.

Gibbons, Judith L., Jennifer A. Zellner, and David J. Rudek. 1999. "Effects of Language and Meaningfulness on the Use of Extreme Response Style by Spanish-English Bilinguals." *Cross-Cultural Research* 33:369–81.

Goodman, Leo A. 1981. "Association Models and the Bivariate Normal for Contingency Tables with Ordered Categories." *Biometrika* 68:347–55.

Green, Donald P., and Jack Citrin. 1994. "Measurement Error and the Structure of Attitudes: Are Positive and Negative Judgments Opposites?" *American Journal of Political Science* 38:256–81.

Greenleaf, Eric A. 1992a. "Improving Rating Scale Measures by Detecting and Correcting." *Journal of Marketing Research* 29:176–88.

———. 1992b. "Measuring Extreme Response Style." *Public Opinion Quarterly* 56:328–51.

Hamilton, David L. 1968. "Personality Attributes Associated with Extreme Response Style." *Psychological Bulletin* 69:192–203.

Harzing, Anne-Wil. 2006. "Response Styles in Cross-National Survey Research." *International Journal of Cross Cultural Management* 6:243–66.

Heinen, T. 1996. *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks: Sage Publication.

Hui, Harry C., and Harry C. Triandis. 1985. "The Instability of Response Sets." *Public Opinion Quarterly* 49:253–260.

———. 1989. "Effects of Culture and Response Format on Extreme Response Style." *Journal of Cross-Cultural Psychology* 20:296–309.

Johnson, Timothy P. 2003. "On the Use of Heterogeneous Thresholds Ordinal Regression Models to Account for Individual Differences in Response Style." *Psychometrika*, 68(4): 563–83.

Johnson, Timothy P., Patrick Kulesa, Young I. Cho, and Sharon Shavitt. 2005. "The Relation Between Culture and Response Styles: Evidence from 19 Countries." *Journal of Cross-Cultural Psychology* 36:264–77.

Joreskog, Karl G. 1994. "On the Estimation of Polychoric Correlations and Their Asymptotic Covariance Matrix." *Psychometrika* 59:381–89.

———. 2005. *Structural Equation Modeling with Ordinal Variables Using LISREL*. Chicago: Scientific Software.

King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98:191–207.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213–36.

———. 1999. "Survey Research." *Annual Review Psychology* 50:537–67.

Magidson, Jay, and Jeroen K. Vermunt. 2001. "Latent Class Factor and Cluster Models, Bi-Plots, and Related Graphical Displays." Pp. 223–64 in *Sociological Methodology*, vol. 31, edited by Michael E. Sobel and Mark P. Becker. Boston, MA: Blackwell Publishing.

Marin, Gerardo, Raymond J. Gamba, and Barbara V. Marin. 1992. "Extreme Response Style and Acquiescence Among Hispanics." *Journal of Cross-Cultural Psychology* 23:498–509.

Masters, Geoff N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47:149–74.

McLachlan, G. J., and D. Peel. 2000. *Finite Mixture Models*. New York: Wiley.

Moors, Guy. 2003. "Diagnosing Response Style Behavior by Means of a Latent-Class Factor Approach: Socio-Demographic Correlates of Gender Role Attitudes and Perceptions of Ethnic Discrimination Reexamined." *Quality and Quantity* 37:277–302.

———. 2004. "Facts and Artefacts in the Comparison of Attitudes Among Ethnic Minorities: A Multigroup Latent Class Structure Model with Adjustment for Response Style Behavior." *European Sociological Review* 20:303–20.

Paulhus, Del L. 1991. "Measurement and Control of Response Bias." Pp. 17–59 in *Measures of Personality and Social Psychological Attitudes*, vol. 1, edited by J. P. Robinson, P. R. Shaver, and L. S. Wrightman. San Diego, CA: Academic Press.

———. 2002. "Socially Desirable Responding: The Evolution of a Construct." Pp. 49–69 in *The Role of Constructs in Psychological and Educational Measurement*, edited by H. I. Braun, D. N. Jackson, and D. E. Wiley. Mahwah, NJ: Lawrence Erlbaum.

Peabody, Dean. 1962. "Two Components in Bipolar Scales: Direction and Extremeness." *Psychology Review*:65–73.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88:879–903.

Poortinga, Ype H., and Fons J. R. Van de Vijver. 1987. "Explaining Cross-Cultural Differences: Bias Analysis and Beyond." *Journal of Cross-Cultural Psychology* 18:259–82.

Rossi, Peter E., Zvi Gilula, and Greg M. Allenby. 2001. "Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach." *Journal of the American Statistical Association*, 96(453):20–31.

Saris, Willem E. 1998. "The Effects of Measurement Error in Cross Cultural Research." *ZUMA-Nachrichten Spezial*:67–84.

Saris, Willem E., and Chris Aalberts. 2003. "Different Explanations for Correlated Disturbance Terms in MTMM Studies." *Structural Equation Modeling* 10:193–213.

Sijtsma, Klaas, and Ivo W. Molenaar. 2002. *Introduction to Nonparametric Item response Theory*, edited by R. M. Jeager. Thousand Oaks, CA: Sage.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. London: Chapman and Hall/CRC.

Sudman, Seymour, and Norman M. Bradburn. 1974. *Response Effects in Surveys. A Review and Synthesis*. Chicago: Aldine Publishing Company.

Van de Vijver, Fons J. R. 1998. "Towards a Theory of Bias and Equivalence." *ZUMA-Nachrichten Spezial*:41–65.

———. 2000. "Methodological Issues in Psychological Research on Culture." *Journal of Cross-Cultural Psychology* 31:33–51.

Van de Vijver, Fons J. R., and Kwok Leung. 1997. *Methods and Data-Analysis for Cross-Cultural Research*, vol. 1, edited by W. J. Lonner and J. W. Berry. Thousand Oaks, CA: Sage.

———. 2000. "Methodological Issues in Psychological Research on Culture." *Journal of Cross-Cultural Psychology* 31:33–51.

Van de Vijver, Fons J. R., and N. K. Tanzer. 1997. "Bias and Equivalence in Cross-Cultural Assessment: An Overview." *European Review of Applied Psychology* 47:263–80.

Van Herk, Hester, Ype H. Poortinga, and Theo M. M. Verhallen. 2004. "Response Styles in Rating Scales: Evidence of Method Bias in Data from Six EU Countries." *Journal of Cross-Cultural Psychology* 35:346–60.

Vermunt, Jeroen K., and Jay Magidson. 2008. *LG-SyntaxTM User's Guide: Manual for Latent GOLD 4.5 Syntax Module*. Belmont, MA: Statistical Innovations.

Wallace, Ruth A., and Alison Wolf. 1998. *Contemporary Sociological Theory. Expanding the Classical Tradition*. Upper Saddle River, NJ: Prentice Hall.

Weijters, Bert, Niels Schillewaert, and Maggie Geuens. 2008. "Assessing Response Styles Across Modes of Data Collection." *Journal of the Academy Marketing Science* 36:409–22.

Welkenhuysen-Gybels, Jerry, Jaak Billiet, and Bart Cambré. 2006. "Adjustment for Acquiescence in the Assessment of the Construct Equivalence of Likert-Type Score Items." *Journal of Cross-Cultural Psychology* 34:702–22.