The survey response

Morren, M.H.

The Survey Response.

A Mixed Method Study of Cross-Cultural Differences in Responding to Attitude Statements

ISBN: 978-90-8891-280-1

Cover design:	Paul Ivan
Printed & Lay Out by:	Proefschriftmaken.nl || Printyourthesis.com
Published by:	Uitgeverij BOXPress, Oisterwijk

This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO), grant number 400-05-164.

The Survey Response.

A Mixed Method Study of Cross-Cultural Differences in Responding to Attitude Statements

Proefschrift
ter verkrijging
van de graad van doctor
aan de Universiteit van Tilburg
op gezag van de rector magnificus, prof. dr. Ph. Eijlander,
in het openbaar te verdedigen ten overstaan van
een door het college voor promoties
aangewezen commissie
in de aula van de
Universiteit

op vrijdag 1 juli 2011 om 14.15 uur

door

Maria Hendrica Morren,

geboren op 10 september 1978 te 's-Hertogenbosch.

Promotiecommissie:

Promotor:	Prof dr. J. K. Vermunt
Copromotor:	Dr. J. P. T. M. Gelissen
	Prof. dr. J. A. P. Hagenaars
	Dr. ir. J. P. Fox
	Prof. dr. E. D. de Leeuw
	Dr. G. B. D. Moors
	Dr. J. Dagevos

Acknowledgements

The last four years of my life revolved around extreme response style, latent class factor analysis, measurement equivalence, cognitive interviews, and mixed methods. Precisely this combination of qualitative and quantitative methods initiated my decision to pursue a PhD at Tilburg University. I was attracted by the statistical knowledge of my supervisors, the change of academic environment and the challenging contents of the proposal. Even though I was not aware at the time, the department of Methods and Statistics has an excellent track record of finishing PhD projects within four years. Sometimes it felt as if my personal circumstances challenged this competence: I seriously developed an expertise in moving within and between cities in the Netherlands (10x). Nevertheless, thanks to the often-unappreciated campus that offers little distraction, the enlightening conversations with colleagues, and not the least, the enduring support of my supervisors, this thesis finally lies before you. During these four years, I greatly appreciated the professional environment at Tilburg and learned a lot about statistical methods as well as the process of transforming a good idea into a publishable article.

First of all, I would like to thank my supervisors John en Jeroen who always supported me with valuable advises throughout the years. The discussions we had about the direction of the project, the specific advantages of the methods used, the precise contents of the chapters and especially the publishability of the separate chapters have sharpened my understanding of the separate stages of the scientific process. Second, I also thank the respondents, without their willingness to participate and their effort to answer also the more difficult questions during the interviews I could have never written the last two chapters. Throughout the years, I've had several 'roomies' that I always (and gratefully) could rely on small talk, statistical advises, much needed coffee breaks, and long walks in the forest. I thank Carmen and Marieke for giving me the welcoming feeling in Tilburg, Hendrik for always having an answer to questions concerning statistics, and Margot for being an inspiring student-assistant during the first phases of my field work. I hope that they and the new roommates, Stéfanie and Ingrid will continue the much appreciated traditions of celebrating birthdays, going out for walks, and having dinner parties. I also treasure the small and long chats in the canteen, during fitness, playing tennis, conferences, dinners and other numerous occasions I've had with Judith, Natalia, Gabriela, Wobbe, Peter, Johan, Rudy, Gonul, Wilco, Marie-Anne, Marcel and many others. To all of you, the PhD time became so much more rememberable you being there!

Maar, ook met de beste professionele voorwaarden was dit boekje nooit zo goed terecht gekomen zonder het plezier en de steun van goede vrienden en familie. Janine, het was heerlijk af en toe er tussenuit te gaan met jou in die vier jaar. Die eerste keer Sziget was een ervaring om nooit meer te vergeten, des te meer omdat het waarschijnlijk het laatste festival is waarbij ik de tijd van mijn leven had ondanks het slapen in een tent vlakbij het podium. Janske, in de laatste jaren is er veel veranderd voor jou en voor mij, en ook deze veranderingen waren zoals altijd heel fijn om te delen met jou. Gesche, Simone en Marijke, bedankt voor het luisteren naar alle beslommeringen en ik hoop dat we nog vele lange avonden samen tegemoet gaan! Lotta, YOU are the best!! You truly believed in me all along. Paul, lief schatje, alles is zoveel fijner met jou erbij. Zoals met vele belangrijke maar ook de minder belangrijke momenten van de afgelopen vier jaar, heb je ook dit proefschrift en de weg er naar toe mooier gemaakt met je creatieve en positieve inzicht.

Contents

Introduction

The Comparability of Attitude Measurements in Cross-Cultural Surveys

People, governments and other institutions have always been interested in how citizens think and what they believe about important social issues. This has been the impetus for social scientists to extensively study people's attitudes towards many different social issues, such as the family, politics, gender equality, and religion. To collect data on people's predispositions towards such issues the survey or questionnaire is usually the preferred tool of the social scientist. Depending on their field of research, what is being measured in these surveys is referred to by researchers as opinions, beliefs, attitudes, values or subjective indicators, and these concepts are often used interchangeably by many researchers[1].

Irrespective of which view one holds about what label should be given to the complex theoretical concepts that are sought to be measured in surveys, a procedure has been developed that is commonly accepted and broadly applied in survey research to measure these complex constructs. In particular, since the early 1930s, social scientists agree that a complex construct can be measured by multiple closed-end questions (or statements) treating several topics that relate to separate aspects of the concerning attitude. These sets of questions are commonly referred to as measurement scales (Likert, 1932; Thurstone, 1928). In this study, we focus on Likert scales as they are predominantly used in social surveys to measure attitudes. A typical Likert scale consists of three or more statements to which the respondent can answer by selecting a response category on a bipolar response scale that runs from 'totally disagree' to 'totally agree', usually containing 5 and sometimes 7 ordered categories.

As modern Western societies have become more culturally heterogeneous in composition in recent decades, comparative public opinion research among culturally diverse populations has become enormously important for social scientific analysis and policy development. Consequently, the Likert scale method has been extensively applied in survey questions measuring attitudes, beliefs and behaviors of non-western immigrants, such as beliefs about other ethnic groups and contacts with members of other groups, with one of the important goals to assess the degree of social and cultural integration of the immigrant groups. Obviously, the quality of measurement instruments to make such assessments and valid comparisons between the cultural groups is crucial. However, the comparability of attitude measurements across culturally diverse groups can be threatened by various sources of bias (Belson, 1986; Heine, Lehmann, Peng, & Greenholtz, 2002; Van Hemert, Baerveldt,

[1] In this study, we mostly refer to 'attitudes' – which we define, following Oskamp and Schultz (2005, p. 9) as 'a predisposition to respond in a favorable or unfavorable manner with respect to a given attitude object'

& Vermande, 2001; Moors, 2004). Indeed, the approach of comparing responses to multilingual survey questions by members of different non-western immigrant groups within a multicultural society inevitably raises several important methodological issues: *To what extent are responses to survey questions from culturally diverse populations really comparable? And if they are not, what may cause this incomparability? In other words, how can we study the issue of incomparability of measurements – that is, detect and explain their incomparability in methodologically adequate ways?* The research that is reported in this thesis seeks to provide an answer to these questions.

In this study, we focus on two particular sources of bias: Members of groups may have a different understanding of, or reaction to questions and members of groups may use a culturally specific response style. With respect to the first issue, consider the following example: In a survey, two women – one belonging to a Moroccan and the other to an Antillean minority – both agree to the question “A man and a woman are allowed to live together without being married”. However, they motivate their response to the question differently: The ethnic Moroccan woman argues that she would never do so but it should be allowed, whereas the ethnic Antillean woman argues that marriage is not important and therefore it should be allowed. Thus, although minorities may have similar attitudes, they can react differently to a survey question because they think of different situations due to cross-cultural differences in language use, the meaning attached to the question wording, the topics that are considered taboo, their habits related to the attitudes or the familiarity to surveys.

Another factor that may influence the comparability of attitudes across cultural groups is a response style: The tendency of people to use certain categories when answering to a survey question, irrespective of their attitude towards an attitude object. As the presence of a response style can seriously bias the measurement of attitudes, researchers have been searching for ways to deal with response styles (Berg & Collier, 1953; Hippler & Schwarz, 1987; Sudman & Bradburn, 1974; Van Herk, Poortinga, & Verhallen, 2004). The detection of and correction for response styles usually revolves around statistical methods to disentangle the influence that the attitude and the response style have on the survey response. Either advanced statistical modeling is used – which is how we proceed in this study –, or a separate measurement scale is developed to measure the response tendency (Arce-Ferrer, 2006; Ross & Mirowsky, 1984). Major disadvantages of these separate measurement scales are that they might also be affected by a response style and they have to be included in the original survey which increases the costs of data collection.

In this study, we focus on a particular response style that survey respondents may use when responding to survey questions and to which much attention has been paid in the survey-methodological literature: The systematic preference or avoidance of choosing extreme answer categories. Because survey-methodological research on response styles is rather broad, we choose to limit the scope of the investigation to this particular response style. This means that other important response styles, such as agreeing to questions that are formulated negatively and positively (acquiescence), the preference for the middle category (midpoint response style), or social desirability, though equally important are not considered in this study. Some people are more likely than others to select – irrespective of the content of the question – the extreme categories such as 'totally agree' and 'totally disagree' (Clarke III, 2001). Survey methodologists commonly refer to this tendency of selecting the extreme categories as extreme response style (ERS) (Groves, 1989). However, some researchers are particularly interested in the extremity of judgment measured by the tendency to use the entire range of the response scale, i.e. the dispersion of answers around the midpoint category (Greenleaf, 1992a), whereas others distinguish between positive and negative Extreme Response Style which is the tendency to select positive or negative extreme categories (Harzing, 2006). For clarity, when we address ERS in this study, we refer to the preference for the extreme categories of a survey item as an indicator of a particular attitude measurement.

Since the 1950s researchers have been aware that the presence of ERS biases the measurement of attitudes in surveys (Cronbach, 1950; Greenleaf, 1992b; Peabody, 1962), and distorts the comparison of attitudes across individuals (for an overview: Hamilton, 1968) as well as across culturally diverse groups (for instance: Bachman & O'Malley, 1984; Clarke III, 2001; Hui & Triandis, 1989). With respect to the latter, early studies approached the relation between ERS and culture being mediated by personal attributes that might differ across ethnic groups or countries (Chun, Campbell, & Yoo, 1974; Peabody, 1962). More recently, cross-cultural differences in ERS were addressed by offering explanations such as the acculturation process (Hui & Triandis, 1989; Marin, Gamba, & Marin, 1992), the language of the survey (Gibbons, Zellner, & Rudek, 1999; Ralston, Cunniff, & Gustafson, 1995) and/or the survey method (Bachman & O'Malley, 1984). The influence of acculturation – defined as the process of immigrants arriving and accommodating to a new culture (Berry, 1990) – on a response style is investigated by comparing response patterns of minority and majority populations, or less and more acculturated minorities (Van Hemert, et al., 2001; Van de Vijver & Phalet, 2004).

If such a preference for choosing particular response categories differs systematically across culturally diverse groups, the presence of a response style may influence the comparability. In particular, a cross-cultural difference in the preference for extreme versus adjacent categories leads one group to appear as if they are more (less) extreme in their attitudes than other groups. This issue of the cross-cultural comparability in preferring extreme or adjacent categories is one of the issues that are being addressed in this study.

1.1 An Integrated Mixed-Method Approach

Because we want to learn more about the occurrence of response differences across groups, and, in its wake, also understand why these group differences in responding occur, we have chosen to implement a mixed method approach for the current research. This is deemed appropriate for several reasons. A mixed method design combines the strengths of quantitative and qualitative research by allowing for an expansion in knowledge in the breadth and scope of the research project (Greene & Caracelli, 2003; O'Cathain, Murphy, & Nicholl, 2007). The quantitative methods allow for answering what-questions while the qualitative methods are more useful in investigating how and why-questions (Morgan & Krueger, 1993; Wooley, 2009). This quality makes a mixed method approach especially fruitful in answering our research questions.

We integrate quantitative research methods with qualitative research methods by using the same set of survey-questions as the object of analysis in both methods. In particular, in the first phase of the study, we conduct quantitative analyses on a selection of questions from a large scale, nationally representative survey among the four largest ethnic minorities in The Netherlands, namely Turkish, Moroccan, Antillean and Surinamese people. In the second phase, we present the same survey questions to members of these four largest minorities during cognitive interviews. Whereas the quantitative analyses show us to what extent the four minorities differ in responding to the survey questions and whether members of the minorities use a particular response style, the qualitative analyses allow us to explore the reasons for these differences in responding among the four minorities in the Netherlands, by investigating how individuals motivate their answers during the response process and perhaps in culturally specific ways. Lastly, we integrate the findings from the quantitative and the qualitative analyses by assigning a response style to the interviewees' answers in the cognitive interviews based on the model estimates obtained from the large-scale survey in the quantitative analyses. This allows us to examine how people with or without a particular response style motivate their answer to the specific survey questions. Below we elaborate on some of the key aspects of our approach.

1.2 Latent Variable Modeling to Assess Measurement Equivalence and Response Style Usage

An approach to Latent Class Modeling that exists within the Latent Variable framework is used in this study to assess the degree of comparability of attitude measurements between the cultural groups and to detect whether a response style leads to measurement inequivalence between the different cultural groups. Based on one survey-question or one scale, it is difficult to assess whether an answer reflects the attitude that is intended to be measured, a response style, or another attitude. However, simultaneously considering the answers to several questions that relate to diverse, weakly related measurement scales can illuminate the influence of a response style or another attitude on a response pattern. To detect bias in the response patterns, we distinguish between the answers that are directly observed in the response pattern and the attitudes that are indirectly measured by the answers. It is assumed that the attitude – also referred to as a latent variable – forms an underlying dimension connecting the questions in the measurement scale. Each respondent is assigned a score on this latent underlying dimension measuring the attitude. Respondents with similar latent scores should have similar response patterns. However, if a respondent's answer diverts from the response pattern we expect based on his or her attitude score, it is likely that he or she had something else in mind than the attitude when answering this question.

Although many different forms of bias can be distinguished, they are subsumed under three general types of bias which may lead to measurement inequivalence (Van de Vijver & Tanzer, 1997). *Construct bias* occurs when different attitudes are measured across cultures using the same set of questions. To avoid this type of bias, one should ascertain that the concepts under study have a similar meaning across the culturally diverse groups. *Method bias* can be related to the familiarity with the survey process as a whole: Respondents from diverse cultural backgrounds may differ in the problems they have with the measurement scales or their understanding of the researcher's intention of the researcher. Furthermore, the presence of the interviewer could influence the answers: People may be more willing to open up to people from a similar background (Dohrenwend, Colombotos, & Dohrenwend, 1968; Hurtado, 1994) or assume that certain topics are common knowledge and not to be discussed (Davis, 1997). Lastly, *item bias* occurs at the level of the question, for example when the question treats topics that are considered taboo in one culture or too painful to discuss. Another source of item bias is the translation of the questions when the survey is held in the mother tongue of the participant (Ellis & Mead, 2000). Such sources of bias are examined in the qualitative part of the research.

1.3 Cognitive Interviewing

To further investigate how cultural differences that may have a biasing effect on survey responses appear in the data, we present the same survey questions to members of these four minorities and asked them about their understanding of the questions and how they justified their answers. This type of interviewing is called cognitive interviewing, because the main interest is in the respondent's cognitive process. The difference between a regular survey interview and a cognitive interview is that the interviewer, next to presenting questions to the respondent, prompts the respondent to reflect aloud on his thoughts on the question. We mainly focused on the understanding of the questions, the interpretation of the wording and the use of the response scale.

Two forms exist of cognitive interviewing: A think aloud (TA) interview and a verbal report (Willis, 2005). While think aloud techniques are used to obtain insight in the cognitive process without directing the thoughts, verbal reports are produced by using probes to elicit more specific information about the response process. Probes are additional questions used to reveal the interpretation of the question, for example: 'Can you repeat the question in your own words?' or 'What does that word mean to you?'. These verbal reports are very useful in testing the validity of survey questions (Willis, Royston, & Bercini, 1991). As we were interested in comparing respondents with respect of specific parts of the response process, we used probes to learn more about their thoughts on the formulation of the question, the intention of the question, and the use of response scale.

We regard the answers of the respondents as justifications of their survey responses and not as immediate (unbiased) reflections of the response process. To interpret the answers correctly, we take into account that the verbalizations might have alter their thoughts on the questions (Russo, Johnson, & Stephens, 1989), and that for some respondents these verbalizations merely reflect justifications of their answers, fabricated on the spot (Nisbett & Wilson, 1977). While we certainly acknowledge such limitations, we argue that cognitive interviews are a valid and important procedure to get an impression of the problems respondents encounter while answering the questions, and to learn more about the way in which respondents justify their answers.

1.4 Background of the Dutch Survey Data

Since the 1980s, a tradition exists in the Netherlands to compare the attitudes about social-cultural topics across minorities using large-scale surveys. Dutch social scientific interest in the degree of social, cultural and economic integration of minorities is particularly reflected in the large-scale survey project *Sociale Positie en Voorzieningen gebruik van Allochtonen* [Social Position and Utility Use of Ethnic Minorities] (SPVA). For conducting the statistical analyses in this study we have used the data from this project that was gathered in 2002. The survey consists of face-to-face interviews with all members of the household. However, in our research we only use the heads of the household to avoid an overlap in attitudes among the members of the household, or in other words, to guarantee independence of observations.

In this study we focus on particular attitudes of the four largest minorities in the Netherlands: People with a Surinamese, Antillean, Turkish, or Moroccan ethnic background. Even though the analyses and conclusions are drawn with respect to these four cultural Dutch minorities, the methods developed in this study are equally well applicable to cross-national comparative survey research. In fact, cross-national differences in languages, customs, norms and values, and behavior are presumably greater, not smaller than the cross-cultural differences. However, we focus on the specific problems in comparing answers of minorities, that is people who – or descend from others who – immigrated to the host country. Many immigrants of the four Dutch minorities can be described as economic immigrants who search for a better life for them and their children in the prosperous Netherlands. Even though their motives may be similar, the social and political context in which the immigration took place differs across the four minorities. Note that we only include immigrants of the first or second generation in our study: People who were born abroad themselves or at least one of their parents. To give the reader some background information on the populations under study, we introduce the four minorities below.

Surinamese. Surinam, a country at the north-east coast of South America, was one of the colonies of the Netherlands since 1674 and became independent after 1975. The immigration from Surinam to the Netherlands was primarily stimulated by the need for labor force in the Netherlands during the 1960s. Later, immigrants were also driven by political motives because of the political chaos that ensued after the independence declaration in the 1980s (Lucassen & Penninx, 1994). Immigration continues until today initiated by family reunion, educational purposes or economic reasons. From the perspective of Surinam, more than a third of the population emigrated, and in 2010 342,000 people with a Surinamese ethnicity live in the Netherlands (CBS, 2010). Although Dutch is the official language in Surinam and is spoken in the majority of the households, many of these immigrants also speak Surinaams (Sranan Tongo) or other languages (ABS, 2005).

Antilleans. The Dutch Antilles (six islands in the Caribbean Sea) were a colony until 1954, the year in which they were assigned the legal status of a country within the Kingdom of the Netherlands. In October 2010, the Dutch Antilles subsumed to exist. Three islands, Curaçao, Sint Maarten and Aruba, continue as separate countries within the Netherlands, while the other islands obtained a legal status comparable to the municipalities in the Netherlands. A consequence of this political situation is that Antilleans have the Dutch nationality and are acquainted to the Dutch language which makes it easier for them to immigrate to the Netherlands. The steady migration flow likely resulted from the poor economic situation at the Antilles (Lucassen & Penninx, 1994). In 2010, about 138,000 Antilleans live in the Netherlands (CBS, 2010).

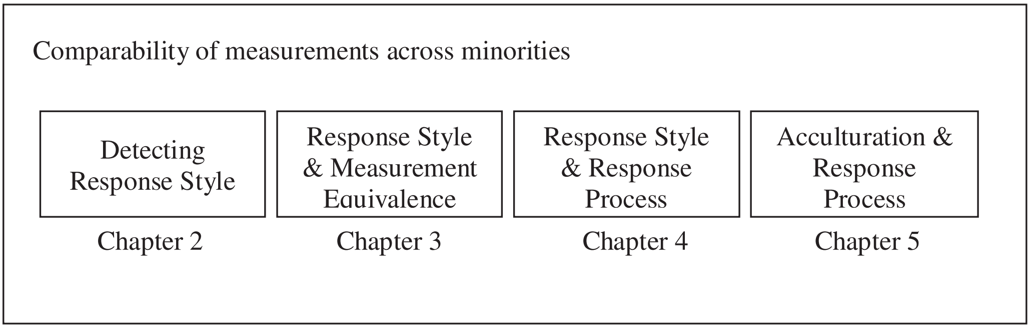
Turks. Initially, Turkish immigrants came to work in the Netherlands. After the economic crisis in the beginning of the 1970s, it became clear that they would stay in the Netherlands and bring their families from Turkey. An increasing number of Turkish migrants came from poorly developed regions in Turkey. In the 1990s, another stream of Turkish migration originated: New families were formed through marriage between Turkish migrants in the Netherlands and Turkish citizens. These family reunions and family formations resulted in being the largest immigrant group consisting of 384,000 people (CBS, 2010).

Moroccans. Similar to the Turks, the Moroccan immigrants came for economic reasons at first, afterwards family reunion and newly formed marriages between Moroccan people in the Netherlands and in Morocco became more important. Also similar to the Turkish immigrants, the Moroccan immigrants more often came from poor regions in Morocco than from the cities or more developed regions. As a result of their social and cultural background, immigrant groups from Morocco as well as Turkey are less familiar with the educational system and the language than the other minorities (CBS, 2005). In 2010, about 349,000 people of a Moroccan ethnic background live in the Netherlands (CBS, 2010).

1.5 Overview of the Thesis

The Chapters in this thesis all relate to methods for investigating the comparability of survey measures across culturally diverse populations. Figure 1.1 provides a schematic overview of the chapters in the thesis:

Figure 1.1 Schematic Overview of the Chapters in the Thesis



Chapter 2. This Chapter treats the question whether a response style distorts the comparison of attitudes across ethnic groups. Using 10 questions that belong to the SPVA dataset treating attitudes toward the Dutch society and the autonomy of children, we illustrate 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. This modeling approach allows us to show that the cross-cultural differences in attitudes are affected by the influence of ERS and that controlling for ERS is vital to a valid cross-cultural comparison of attitudes.

Chapter 3. Whereas in the previous Chapter we assessed that ERS affects the comparability of attitudes, in Chapter 3 we are interested in assessing to what extent the cross-cultural differences in extreme responding may lead to incomparable survey responses when not controlled for. To examine how extreme responding affects the cross-cultural comparability of responses to 10 SPVA questions treating the attitude toward the Dutch society and family values, Chapter 3 shows how groups are compared on basis of the factor loadings, intercepts and factor means in a Latent Class Factor Model. Similar to Chapter 2, a latent factor measuring the response style is explicitly included as an explanation for group differences found in the data. We find that ERS can account for the observed cross-cultural differences in factor loadings and partly for the cross-cultural differences in intercepts.

Chapter 4. This Chapter presents a mixed methods approach that integrates quantitative and qualitative methods to analyze why the four largest minorities in the Netherlands respond differently to 15 items treating the attitude toward the Dutch society, family values and autonomy of the children. In the quantitative study, we investigate whether group differences between minorities can partially be explained by response styles, and whether it is possible to detect particular kinds of response style (see Chapter 2 and 3). In the qualitative study we subsequently investigate how people with a specific response style arrive at their answer during the response process. Using the quantitative results, we infer response styles to the interviewees and assess whether the explanations of their responses to the probes in the interviews can illuminate the employment of a response style. We conclude that two different response styles can be distinguished and that these response styles are related to how people deal with difficulties in answering survey questions.

Chapter 5. This Chapter addresses a similar research question to the previous Chapter: Do respondents from different cultural backgrounds use a particular response style and response strategies when responding to attitude statements, and are these characteristics of the response process related to a respondent's ethnicity and generation of immigration? Similar to Chapter 4, we combine qualitative and quantitative methods: Differences in response style are related to a selection of characteristics of respondents in the large representative sample of minorities in the Netherlands as well as in a purposively selected qualitative sample of persons belonging to the same cultural groups. Furthermore, analysis of cognitive interviews of the latter is used to examine how respondents may overcome the difficulties of responding to survey items in a cross-cultural survey by means of particular response strategies. In this chapter, we particularly investigate whether (and how) the group differences in responding are related to the level of acculturation and whether respondents with a response style use culturally specific response strategies. Acculturation refers to the extent to which the respondents are familiar with the culture of the host society and the extent to which they accept the norms and values of the host society (as well as their ethnic identity).

Finally, in **Chapter 6**, we summarize and discuss the methodological insights that can be derived from this study. We also reflect on the limitations of the research that is reported in this thesis, and propose some ideas for future research which may contribute to improving the quality of attitude measurements in cross-cultural surveys. Note that Chapters 2, 3, 4, and 5 are adapted versions of articles that were submitted for publication in international peer-reviewed journals. We made it possible to read each chapter separately which creates some overlap in text among the chapters.

Chapter 2*

Dealing with Extreme Response Style in Cross-Cultural Research: A Restricted Latent Class Factor Analysis Approach

---

* This Chapter has been accepted for publication as: Morren, M., Gelissen, J.P.T.M., and Vermunt, J.K. (in press a). Dealing with extreme response style in cross-cultural research: A restricted latent class factor analysis approach. To appear in *Sociological Methodology*.

2.1 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 held across 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 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 is 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 socio-cultural 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 & 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 & 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 which therefore plays an important role in the literature on survey methodology is response style behavior; that is, 'the systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content' (Paulhus, 1991, p.17). In this chapter, we particularly focus 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 (see for instance Bachman & 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 non-random 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 (see for instance, Clarke III, 2001; Gibbons, et al., 1999; Harzing, 2006; Hui & Triandis, 1989; Johnson, Kulesa, Cho & Shavitt, 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 & Diener, 2003), a type of problem that is sometimes also referred to as the duality between genuine and stylistic variance (Poortinga & 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 & 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 which are corrected for group differences in extreme response style behavior. Moors' method involves defining a latent variable model which, 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 & McClendon, 2000; Cheung & 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 contribution. Recent advances in statistical software for latent structure analysis make this model readily available to applied researchers.

This chapter 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)[2], 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, among which 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.

2.2 Methods for Detecting Extreme Response Style: An Overview

Extreme response style is commonly defined as the tendency of a respondent to express him- or herself extremely by choosing the end-points on a rating scale, independent of the extremity of his or her opinion. This tendency is typically assumed to exist irrespective of the substantive item content but to show up in consistency with the positive or negative formulation of an item[3] (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 & Triandis, 1985; Krosnick, 1991), as the questions are formulated in another language (Gibbons, et al., 1999), or as different survey forms are used (Bachman & O'Malley, 1984; Van Herk et al., 2004). Following Hui and Triandis (1989) and Podsakoff, MacKenzie, Lee, and Podsakoff (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 (s)he tends to answer more extreme than other respondents in the investigated population. To which extent this tendency actually appears in a particular rating scale depends on item characteristics such as response format, item content, location in the questionnaire, etcetera. Thus, some questions are more likely to elicit extreme response style than other questions.

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

[2] In Dutch, the abbreviation SPVA stands for *Sociale Positie en Voorzieningengebruik van Allochtonen*. We thank Data Archiving and Networked Services (DANS, 2002) for providing the data files.

[3] 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, p. 304).

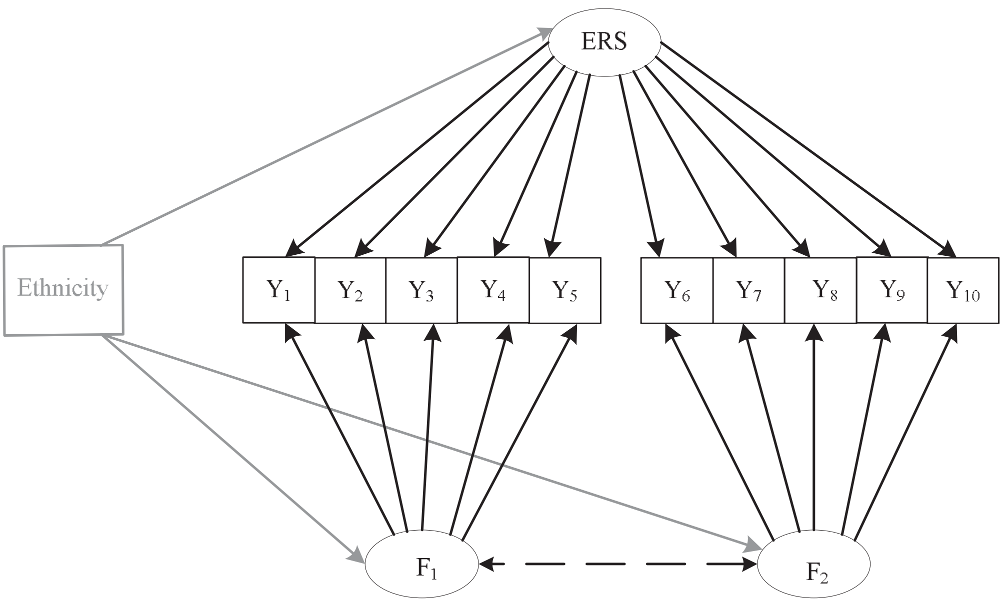
sometimes possible to determine whether a person tends to answer more extreme than other persons in the sample. Several methods – ranging from straightforward descriptive methods to rather advanced statistical models – have been developed to measure ERS using multiple ratings. Whereas some researchers are mainly interested in methods for detecting ERS (De Jong, et al., 2008; Greenleaf, 1992b; Hui & Triandis, 1989; Johnson, et al., 2005; Sudman & Bradburn, 1974), others focus on methods for correcting the biasing influence of ERS on the measurement of attitudes (Greenleaf 1992a; Marin, et al., 1992; Saris, 1998), or compare the influence of ERS on attitudes across different survey methods (King, Murray, Salomon, & Tandon 2004; Saris & Aalberts, 2003; Weijters, Schillewaert & 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 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 not only be used to detect respondents with ERS but also to assess differences in ERS between socio-cultural groups as well as to control for ERS in subsequent statistical analysis (Bachman & O'Malley, 1984; Clarke III, 2001; Hui & Triandis, 1989; Marin, et al., 1992). The measurement of ERS by means of a separate ERS scale has found very limited application, among other things because of 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. One 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 problem 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 for the fact 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 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 one 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.

Figure 2.1 The latent variable model for the detection of a response style



An example of such a latent variable model is depicted in Figure 2.1. 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 (see for instance, Sijtsma and Molenaar, 2002) – a multidimensional extension was needed to be able to control for correlations caused by content factors. As they 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 constructing 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 how they call response style), it cannot deal with ERS as defined in the current chapter, 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 which 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) and 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 by Johnson (2003) model 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, & Muthen, 1989), sometimes referred to as multiple-group LISREL modeling (Jöreskog, 2005). Rather than specifying a latent variable model with a response style factor as displayed in Figure 2.1, one uses a model with content factors only. The aim is not measuring ERS, but checking 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 & Tanzer, 1997). As Cheung and Rensvold (2000) show in a simulation study, differential ERS across groups results in non-invariant factor loadings (larger loadings for the more extreme group) and it may also affect item intercepts. This multiple-group SEM approach is useful if one wishes 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: non-invariant intercepts and loadings may also be caused by other factors 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 & 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 2.1 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 or unrelated: the association between the items measuring unrelated substantive dimensions can only be explained by the response style factor. At the individual level, this 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 equations 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[4] – 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 following two-way tables clarify the difference between a monotonic and a non-monotonic pattern by showing how these patterns impact the association between two items.

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

Table 2.1a

Pairwise response combinations which are more likely for two items measuring the same attitude (between braces the value of the attitude)

	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 (+)

Table 2.1b

Pairwise response combinations which are more likely when both are affected by ERS (between braces the value of the ERS factor)

	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 (+)

Tables 2.1a and 2.1b show the dominant association pattern between two items arising from a shared attitude factor and ERS, respectively. The Xs indicate which combinations of responses can be expected to be more likely than if responses were independent, and the symbol between braces indicates whether these responses are given by persons with a low (-), middle (0), or high (+) attitude/ERS score. Table 2.1a shows that for items measuring the same underlying construct cell frequencies on the diagonal of the table can be expected 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 2.1b illustrates the very different pattern arising from the non-

monotonic effect of a 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 non-extreme 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 2.1a and 2.1b. The association between two items measuring different dimensions will be of the form of Table 2.1b, though in the case of correlated dimensions it may also be a mixture between Tables 2.1a and 2.1b, but the importance of Table 2.1a will be much less than for items measuring the same dimension.

The non-monotone 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 either using 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 which relies on linear relations and interval (or ordinal) level measurements (Jöreskog, 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.

2.3 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 is that it resolves the problem of 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 (2003) showed that ignoring ERS may yield latent attitudinal factors which 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 & 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 which 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. Similarly 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 similar model as 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 as Moors and we used; that is, by defining the latent variables to be continuous instead of ordinal (see also Appendix A).

The LCFA model is graphically presented in Figure 2.1. 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. Below we show 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}:

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

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 ten items depicted in Figure 2.1 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 non-extreme answers compared to other respondents. The interpretation of the ERS factor is valid if the extreme answer categories (for example, categories 1 and 5 of a five point scale) have positive β_{3jc} values but the non-extreme 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 one finds another response style than ERS, for instance, acquiescent response style (ARS). Similarly 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 non-monotone 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, such 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

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

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,

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

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 (β_{3jc+1} - β_{3jc}) should be negative when comparing category 2 and 1 and positive when comparing categories 5 and 4, assuming we have a 5-point scale. The same model but now in term of odds instead of logits is formulated as follows:

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

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).

Below, we will show that also the ERS factor specification can similarly be restricted using scores for the response categories; for example, scores with a W-shape or U-shape pattern. A U-shape pattern can be obtained, for example, using scores 1.5, -1, -1,-1, and 1.5 or equivalently 1, 0, 0, 0, and 1[5]. 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: one can really believe that there are three classes or one 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 semi-parametric or non-parametric specification of the distribution of a latent variable (Heinen, 1996; Skrondal & 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 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

[5] 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.

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, such as, for example, a set of dummies for the cultural group one belongs to (see Ethnicity in Figure 2.1). The regression model used for the ordinal latent variables is also an adjacent-category ordinal logit model.

Figure 2.1 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 one wishes to measure. It should, however, be noted that it is not a problem to relax this assumption, One 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.

2.4 Application

The data used to illustrate the ERS model described in the previous section comes 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 2.2). Topics treated in the survey are among others: family values, work values, religion, women's emancipation, work and education (Dagevos, Gijsberts, & 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 Netherlands* 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 which now runs from a negative (1) toward a positive (5) response to the item.

Table 2.2

Mean observed item response per ethnic group (N=3574)

Factors and items	Turks	Moroccans	Surinamese	Antilleans
Factor 1: Attitude toward NL				
Item 1: In the Netherlands immigrants get many opportunities	3.53 (1.06)	3.42 (1.08)	3.26 (1.11)	3.24 (1.15)
Item 2: The Netherlands is hostile to immigrants	2.80 (1.02)	2.47 (0.88)	2.40 (0.88)	2.52 (0.91)
Item 3: In the Netherlands your civil rights as an immigrant are respected	3.40 (0.91)	3.55 (0.86)	3.52 (0.86)	3.44 (0.84)
Item 4: The Netherlands is a hospitable country for immigrants	3.03 (0.97)	3.47 (0.92)	3.69 (0.89)	3.60 (0.91)
Item 5: The Netherlands is tolerant towards foreign cultures	3.83 (0.91)	3.57 (0.87)	3.84 (0.82)	3.69 (0.83)
Factor 2: Autonomy of children				
Item 6: Children should live at home until marriage	3.69 (1.05)	3.76 (1.12)	2.94 (1.27)	2.59 (1.23)
Item 7: Elderly should be able to move in with their children	3.13 (1.13)	3.79 (0.97)	3.10 (1.15)	3.01 (1.18)
Item 8: Adult children should be able to move in with their parents	3.88 (0.88)	3.94 (0.85)	3.32 (1.08)	3.14 (1.11)
Item 9: Parents should always be respected, even if they do not deserve it based on their behavior or attitude	3.11 (1.15)	3.36 (1.13)	2.86 (1.12)	2.88 (1.09)
Item 10: Older family members should have more influence in important decisions (for instance about moving) than younger ones	4.11 (0.83)	4.21 (0.89)	3.61 (1.11)	3.70 (1.09)
N	914	862	1016	782
Response Rate (%)[a]	52	52	44	51

Note. Standard errors are shown in parentheses. Item 2 is formulated in reversed manner where a positive answer indicates a conservative attitude. [a] The response rate excludes those who were not at home, refused, or otherwise were unavailable (see DANS, 2002, p. 44).

Table 2.2 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.

Table 2.3.

Goodness of fit statistics for Latent Class Factor Models (N=3574)

Model	Log-Likelihood	BIC (based on LL)	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

We estimated various LCFA models using the SPVA data. For this purpose we used the syntax module of the Latent GOLD 4.5 program[6] (Vermunt and Magidson, 2008), a program for the maximum likelihood estimation of latent class models and other types of latent variable models. Table 2.3 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

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

estimated models are nested: for example, models without and with 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 asymptotic likelihood-ratio test because certain regularity conditions are not met (McLachlan & Peel, 2000). A possible way out would be to use likelihood-ratio tests with a bootstrap p values, which are, however, rather computationally intensive procedures. We, therefore, decided to use only BIC for model selection which is the most common procedure in latent class analysis.

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, Model F also including a style factor. The analyses in Model D, E, and F are repeated in Models B_1, B_2 and B_3 containing 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 which led us to conclude that the factors that were supposed to measure substantive content are confounded with ERS. Not surprisingly, including 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, one 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 2.4

Parameter estimates obtained with Model F containing the content factors autonomy of children, attitude toward the Netherlands and a style factor (N=3574)

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

Note. Standard errors are shown in parentheses. All parameters shown in Table 2.4 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).

Table 2.4 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 discussion below).

When ordinally restricted as in Table 2.4, 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 Netherlands 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 a respondent's ERS score, the more likely he or she selects the outer categories and the less likely he or she will select the other categories. The style factor has a large effect on the item responses which 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 one changes from one class to the next. Thus, the higher a person stands on the ERS dimension, the (much) more he or she is likely to choose *totally agree* (*totally disagree*) instead of *agree* (*disagree*). We emphasize that this result is *given the substantive factors*, meaning that this person selects these categories more often than would be expected on 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.

Table 2.5.

Fit measures for variants of Model F using a restricted specification for the style factor (N=3574)

Model	Log-Likelihood	BIC (based on LL)	Number of parameters
F) Model F	-42338.0	85469.6	97
F_1) Model F + W-shape pattern	-43126.0	86800.2	67
F_2) Model F + U-shape pattern	-42535.9	85619.9	67
F_3) Model F + W-U-shape pattern	-42434.3	85416.9	67

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

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-shape or U-shape patterns can be obtained in a similar way as the ordinal models for the content factors; that is, by using pre-specified scores for the categories of response variables. A W-shape pattern (Model F_1) is obtained using category scores 1, -1.5, 1, -1.5, and 1, and a U-shape 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 2.5, both Model F_1 and Model F_2 fit worse 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-shape pattern of category parameters than a W-shape 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 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 SPVA survey (55 in total) as 0 (non extreme response) and 1 (extreme response). The ERS index is simply the proportion of items with an extreme response.[7] Moreover, we estimated a uni-dimensional IRT model using these 55 dichotomous items, and computed IRT-based ERS scores.[8] The correlations between LCFA-based ERS score (using Model F) with the ERS index and IRT-based ERS score 0.81 and 0.76, respectively. 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 0.88, 0.95, and 0.99, respectively. This shows that the proposed procedure is robust towards 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 (see for instance 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 (see for instance Baumgartner & Steenkamp, 2001). Both argue that one endpoint corresponds to a response pattern containing many extreme responses and signifies '*strongly affected by ERS*'. In our view, the conceptualization of the other endpoint depends on the operationalization of ERS. In the sum-score method, where one simply counts the number of extreme responses, the opposite endpoint of the dimension represents response patterns with few extreme responses; that is, with the tendency to prefer the non-extreme categories *agree*, *disagree* or *neither agree nor disagree*.

---

[7] This ERS index is similar to the index discussed in the Overview. The proportion is based on the items without missing values.

[8] 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 of the 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 inter-item associations, showing that the remaining associations are not very large. The IRT model was estimated using ML with the missing values.

Table 2.6.

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			
Number	ERS class			N	ERS class			N	ERS class			N
	1	2	3		1	2	3		1	2	3	
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 2.6 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 could be expected, the class membership probabilities conditional on the number of extreme responses show that the smaller this number, the more likely one belongs to class 1 and the larger this number, the more likely one belongs to class 3. For the number of responses in the adjacent categories this pattern is the other way around: Many of such responses makes it more likely to belong to the first class while few of them makes 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 one 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 ERS factor is useful and valid in other situations.

Table 2.7.

Goodness of fit statistics for Latent Class Factor Models with ethnicity included as a covariate in every model (N=3574)

Model	Log-Likelihood	BIC (based on LL)	Number 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

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 2.7, every model mentioned in Table 2.2 includes ethnicity dummies as predictors in the regression equations for the latent variables. The fact that the likelihood values of all models in Table 2.7 show a significant improvement of the fit of the models in Table 2.2 indicates that ethnicity is indeed associated with the (supposed) substantive dimensions.

Table 2.8.

Effect of ethnicity in Model C_g, Model D_g, Model E_g and Model F_g (N=3574)

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

Note. Standard errors are shown in parentheses. Model C_g: No style factor and nominal specification of items; Model D_g: With style factor and nominal specification of items; Model E_g: No style factor and ordinal specification of items; Model F_g: With style factor and ordinal specification of items.

Table 2.8 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 Turkish people.

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 2.8 illustrates that ERS suppresses the group differences somewhat and that the standard errors are smaller in Model C_g and E_g. Although not further investigated, this finding indicates that the ordinal specification used in our LCFA analyses but also used in multi-group 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.

2.5 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 chapter can be expanded in several interesting ways. The unrestricted style factor not only is able to detect 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 could 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, et al., 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, such as 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-shape 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. Not only can the relationship between the items and the substantive factors be made group specific, but also their relation 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 contribution, 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 fit worse 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, this contribution emphasizes the need for detecting and correcting for extreme response style in cross-cultural research.

Appendix A. 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 from Table 2.3 and Model A_g to F_g, from Table 2.7. The variables and equations sections of the syntax file for the most complex model F_g is 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, one 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 equation 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 factor.

Chapter 3*

The Impact of Controlling for Extreme Responding on Measurement Equivalence in Cross-Cultural Research

* This Chapter has been conditionally accepted for publication as: Morren, M., Gelissen, J.P.T.M., and Vermunt, J.K. (in press b). The Impact of Controlling for Extreme Responding on Measurement Equivalence in Cross-Cultural Research. *Methodology*

3.1 Introduction

Cross-cultural comparisons in which people from different nations or ethnic backgrounds are asked how they feel about social issues or how they behave constitute an important part of research in the social and behavioral sciences. More and more attention is being paid to the validity of such comparisons (Berry, Poortinga, Segall, & Dasen, 2002; Johnson, et al., 2005; Van de Vijver, 1998; Van de Vijver & Leung, 1997). In particular, this field of research questions whether it is possible to compare people with different cultural backgrounds on their attitudes and values. It is likely that people with a different frame of reference – rooted in their experiences, their social interactions, and the norms and values shared by their group – understand the topics that are raised in a survey differently (Triandis, 1990; Wallace & Wolf, 1998). Consequently, because describing and explaining differences in attitudes is the aim of most cross-cultural studies, one should empirically establish that respondents from different groups have the same topic in mind while answering a survey-item (Krosnick, 1999; Tourangeau, 2003). If this is not the case, comparing attitudes between groups is similar to comparing apples and oranges. The methodological literature refers to this situation as measurement inequivalence.

In this contribution, we argue that such a lack of measurement equivalence (ME)[9] can be related to group differences in response styles which cause respondents from culturally diverse backgrounds to respond differently to the items than one would expect on basis of their attitudes (Hui & Triandis, 1989). We show that these group differences in responding lead to what appears to be measurement inequivalence and that correcting for the response style results in more equivalent measurements.

The investigation is narrowed down to extreme response style (ERS) because it has repeatedly been shown that this response style seriously distorts attitude measurement in social survey research (see for instance Chun, et al., 1974; De Jong, Steenkamp, Fox & Baumgartner, 2008). The response pattern of an item that is affected by ERS shows a higher frequency of extreme responses – the endpoints of the item scale – than one would expect based on the respondent's attitude. This impedes a correct estimation of model parameters when modeling group differences in attitudes. Moreover, an extreme response pattern may represent a truly extreme attitude as well as ERS; that is, ERS may confound genuine and

[9] In other traditions ME is referred to as measurement invariance (MI) (Cheung & Rensvold, 2000; Meredith, 1993; Millsap, 1995; Steenkamp & Baumgartner, 1998), or as Differential Item Functioning (DIF) (Sijtsma & Molenaar, 2002). Alternatively, Adcock and Collier (2001) address ME as the contextual specificity of measurement validity.

stylistic variance (Van de Vijver & Leung, 1997). Thus, ERS leads to biased attitude measurement if not controlled for. Additionally, research findings show that people with differing cultural backgrounds may be subject to extreme responding to a different degree (Bachman & O'Malley, 1984; Gibbons, et al., 1999; Hui & Triandis, 1989; Johnson, et al., 2005; Marin, et al., 1992). In this chapter we show how a difference in ERS between culturally diverse groups imports measurement inequivalence in the data and, if not controlled for, biases the attitude measurement.

To this end, we propose a latent variable model that simultaneously allows for examining measurement equivalence as well as the detection of and the correction for ERS. Importantly, this model enables us to assess the implications of the presence of ERS for measurement equivalence. We build on the contributions of Moors (2003, 2004) and apply logistic Latent Class Factor Analysis (LCFA) (Eid, et al., 2003; Heinen, 1996; Vermunt & Magidson, 2004) instead of linear Structural Equation Modeling (SEM) that is commonly used in multiple group analyses (Billiet & McClendon, 2000; Byrne & Stewart, 2006; Cheung & Rensvold, 2000). As we will show, SEM is an inappropriate method to deal with the non-monotone response pattern caused by ERS because of the assumption of linear relationships between latent and observed variables. In contrast, the less restrictive LCFA approach proposed here does not make such stringent assumptions, it allows for the detection of and the correction for ERS and measurement equivalence can be assessed.

In the remainder of this contribution, we illustrate how multiple-group analysis within the LCFA framework can be used to detect measurement inequivalence. We present a latent variable model that disentangles style and substance and we explain how this model can be adjusted to a LCFA model that can also detect and correct for ERS. We then show in an analysis of a generated data set in which we simultaneously detect measurement inequivalence and correct for ERS that specific forms of measurement inequivalence relate to the presence of extreme responding. Finally, we apply the multiple group LCFA approach to data obtained from four ethnic groups within the Netherlands using the Dutch survey The Social Position of Ethnic Minorities and Their Use of Services (SPVA)[10] and demonstrate the usefulness of the approach in an empirical application.

[10] In Dutch, the abbreviation SPVA stands for *Sociale Positie en Voorzieningengebruik van Allochtonen*. We thank Data Archiving and Networked Services (DANS) for providing the data files.

3.2 A Latent Class Factor Approach to Multiple-Group Analysis

Complex constructs such as people's attitudes cannot be observed directly. To obtain a valid and reliable measurement of such constructs, researchers usually ask respondents multiple questions which indicate several important aspects of the attitude (Bollen, 2002; Skondral & Rabe-Hesketh, 2004). These ideas about attitude measurement are applied by modeling the attitude as a latent – unobserved – variable (also called factor or trait) and the questions as observed variables (hereafter called items). Within this latent variable framework, an important goal of multiple-group analyses is to measure the extent to which the groups have different attitudes in terms of the group means of the latent variables. However, these group differences in latent means can only be compared validly and reliably when the same latent variable model can be applied within each group as well as across groups (Byrne & Watkins, 2003; Meade & Lautenschlager, 2004a, p. 60; Mullen, 1995; Van de Vijver & Leung, 1997).

In this chapter, we investigate whether the items and the response scale of attitude measurements are used homogeneously – which indicates measurement equivalence – or rather heterogeneously – which indicates measurement inequivalence – by people who come from culturally diverse backgrounds but who actually have similar attitudes (Hui & Triandis, 1985; Van de Vijver & Leung, 1997). If measurement equivalence is absent, which is shown by the fact that the associations between the items and the attitudes differ across groups in strength and significance, then we hypothesize that this absence of measurement equivalence can be partly or even completely be explained by a confounding effect due to a group-specific presence of ERS. In Figure 1, we graphically illustrate various Latent Class Factor Models which allow the investigation of such issues in a multiple-group analysis on a pooled sample.

Figure 3.1. The pooled approach to multiple-group analyses of metric and scalar equivalence

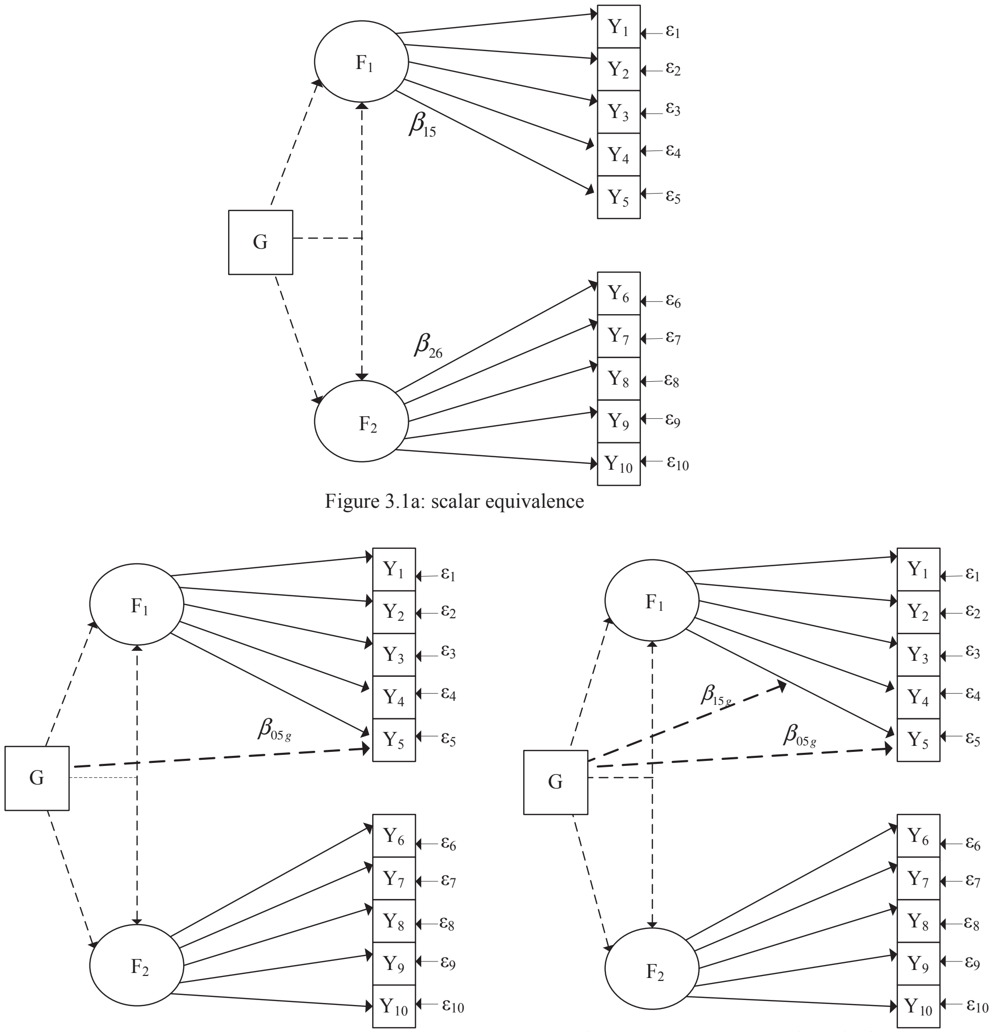


Figure 3.1a: scalar equivalence

Figure 3.1b: metric equivalence

Figure 3.1c: measurement inequivalence

Note. The dashed arrows indicate that the groups differ with respect to these relationships. The parameters are not denoted as category-specific item parameters to simplify the graphical display. In Figure 1a, the item parameters β_{1j} and β_{2j} are described by β_{15} and β_{26} for item 5 and 6. In Figure 1b, the inequivalent intercepts of item 5 are indicated by β_{05g}, and in Figure 1c the inequivalent factor loadings of item 5 are indicated by β_{15g}.

Here, Y_1 *to* Y_{10} represent the item responses which are directly related to the latent variables measuring the attitudes F_1 and F_2. Furthermore, the item responses may either be related directly, or indirectly, or in interaction with the effect of the latent variable measuring the attitudes to the observed group variable *G*. As we will explain below, which particular effects of the grouping variable G are included in the model depends on the type of measurement equivalence that the researcher seeks to investigate. The systematic variance among the item responses is captured by the factor loadings, i.e. the relations between the latent variables and the item responses; the random variation is represented by the error terms ε_j. In the models depicted in Figure 3.1, it is assumed that the five items in the first item subset do not relate directly to the second attitude which is modeled by fixing the item parameters β_{2j} to zero. In the same way, the parameters β_{1j} are fixed to zero for the five items in the second item subset.

An important advantage of the LCFA approach to multiple-group analysis in comparison to other well-known approaches that are based on the linear regression model is that the equivalence of item intercepts, factor loadings, factor means and (co)-variances – which is necessary for the evaluation of various forms of measurement equivalence – can be tested simultaneously without using restrictions. In particular, contrary to the linear regression model used in CFA analyses LCFA uses an ordinal logit regression model to measure the latent variables. The consequence of this difference in modeling is that whereas in the CFA approach the researcher needs to include certain restrictions in the model to fix the location and scale of the latent variables, this is unnecessary in the LCFA approach. The ordinal logit model for the latent variables in Figure 3.1 is described in Appendix B. In such multiple-group analyses, the group differences are introduced in the model by an explanatory covariate which measures group membership; in Figure 3.1 this is indicated by G. A direct effect of G on the latent variables denotes a group difference in the latent means and/or co-variances. These group differences in the attitudes can only be measured reliably and validly when the attitudes are measured equivalently across groups. Here, we are interested in two forms of measurement equivalence: scalar and metric equivalence[11].

Scalar equivalence is the most restrictive type of measurement equivalence and occurs when respondents from different backgrounds react similarly to the items given their attitudes. Establishing scalar equivalence is necessary to validly compare means of latent variables across groups. The situation of scalar equivalence is depicted in Figure 3.1a. Here, the group differences in the latent means are indicated by the dashed arrows between the observed variable G and the latent variables F_1 and F_2. The model in Figure 3.1a for the observed score of respondent i on item j is formally represented by:

[11] Prior to assessing scalar and metric equivalence, one also needs to establish configural equivalence, which holds that items measuring the attitudes exhibit the same configuration of loadings in all groups. Here we assume that configural equivalence has been established and the researcher now seeks to investigate more restrictive forms of equivalence of measurement instruments.

$$E(Y_{ij} \mid F_{1i}, F_{2i}) = \beta_{0j} + \beta_{1j} F_{1i} + \beta_{2j} F_{2i} , \qquad [5]$$

where the expected value of the response Y_{ij}, conditional on the attitudes F_{1i} and F_{2i}, depends on the item parameters β_{0j} representing the intercept, the parameters β_{1j} representing the influence of F_{1i}, and the parameters β_{2j} representing the influence of F_{2i}. The expected value of the errors ε_{ij} is zero because they are assumed to be unrelated and normally distributed.

A weaker form of measurement equivalence is metric equivalence, which is attained when the groups differ in their perception of the origin of the item scale but perceive the distances between the item categories and/or the order of the item categories similarly. Thus, metric equivalence is defined as groups having different item intercepts and error terms but equal factor loadings given their attitudes:

$$E(Y_{ij} \mid F_{1i}, F_{2i}, g) = \beta_{0jg} + \beta_{1jg} F_{1i} + \beta_{2jg} F_{2i} \ . \qquad [6]$$

where the expectation of the response is conditional on the attitudes F_{1i} and F_{2i} and on group g to which individual i belongs. The subscript g of the parameter β_{0jg} denotes that the item intercept is group-specific; in other words, the intercepts are set free to vary across groups. In Figure 1b, the situation of metric equivalence is graphically represented for item 5 where a group-specific intercept is indicated by the dashed arrow representing a direct effect of the group variable G on Y_5.

Note that measurement equivalence is completely violated if the groups perceive the items completely different given their attitudes, resulting in group differences in the intercepts and the factor loadings:

$$P(Y_{ij} \mid F_{1i}, F_{2i}, g) = \beta_{0jg} + \beta_{1jg} F_{1i} + \beta_{2jg} F_{2i} \ . \qquad [7]$$

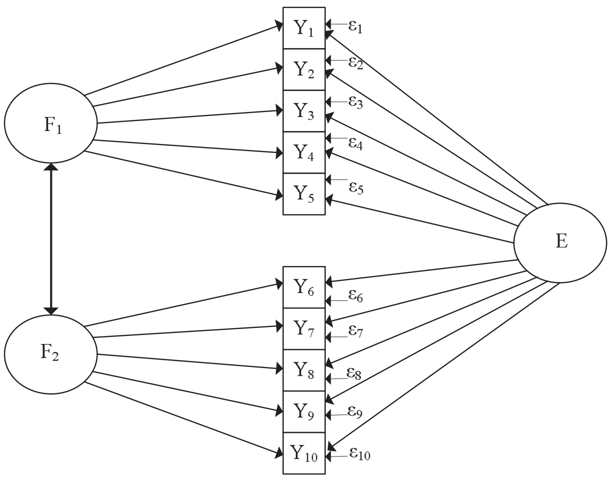
where the subscript g of the parameters β_{1jg} and β_{2jg} denotes that the factor loadings are group-specific in addition to the intercepts. In Figure 1c, measurement inequivalence is represented for item 5 by the dashed arrows representing a direct effect of G on Y_5 β_{05g} and a group-specific factor loading β_{15g}.

In the LCFA approach to multiple-group analyses, we test for equivalence of certain parameters by constraining them to be equal across groups, yielding a more parsimonious model. If the groups respond equivalently then this more parsimonious model is preferred. However, if they respond inequivalently more complex models are required to avoid misspecification. By comparing model goodness-of-fit values, we assess which specific form of equivalence is attained. If fixing the group-specific parameters to equality does not deteriorate the model fit, the more parsimonious model is accepted and a particular type of measurement equivalence is attained. Specifically, the measurements are scalar equivalent if the fit of the model in equation [5] does not deteriorate compared to model described in [6]; the measurements are metric equivalent if the fit of the model in equation [6] does not deteriorate compared to the model in [7]. Note that – although in Figures 3.1b and 3.1c inequivalence is depicted for only one item – multiple or all items can of course be simultaneously inequivalent across groups.

3.3 Extreme Response Style

Apart from measurement inequivalence an additional problem surfaces in multiple-group analyses if the groups differ in their style of responding: the confounding of group differences in the attitudes and the response styles (Eid, et al., 2003; Poortinga & Van de Vijver, 1987; Van de Vijver & Leung, 1997). A straightforward manner to deal with this problem is to explicitly control for the response style by including one or more latent variables that accurately measure the response styles (Billiet & McClendon, 2000; Cheung & Rensvold, 2000; De Jong, et al., 2008). Figure 3.2 illustrates this approach which is a latent variable model that simultaneously detects and corrects for the response style.

Figure 3.2. The latent variable model for the detection of a response style



In Figure 3.2, Y_1-Y_{10} indicate the item responses that relate to the latent variables representing the attitudes F_1 and F_2, and the extreme response style E. The response style and the attitude are disentangled by means of the model structure. Whereas the respondent's attitude only affects his or her answer to the items that reflect the same construct, the respondent's response style – by definition – affects the answers to all items regardless of their content (Hui & Triandis, 1989; Javeline, 1999; Johnson & Van de Vijver, 2003; Sudman, Bradburn, & Schwarz, 1996). The validity of the model is increased by including two weakly related attitudes: As ERS is unrelated to item content, it is expected that the response style is present across items treating diverse topics. Note that more substantive factors could be included to investigate whether the response style pertains to more items.

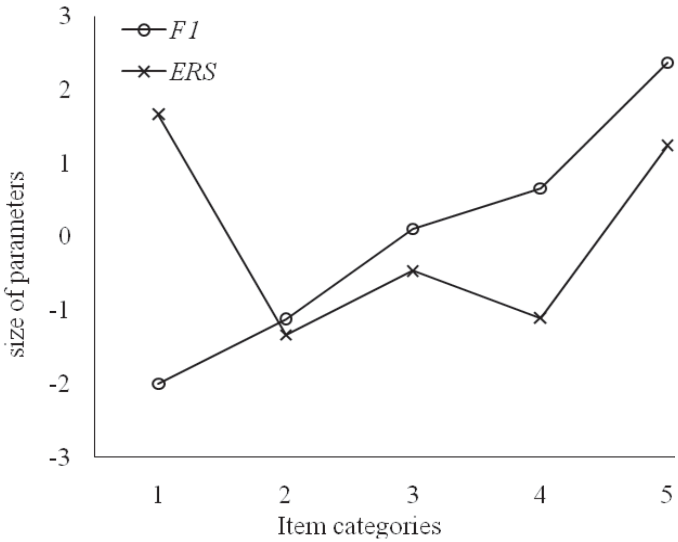
The model in Figure 3.2 was earlier applied within a Confirmatory Factor Analysis (CFA) framework to detect acquiescence (Billiet & McClendon, 2000); in this approach the latent variables as well as the observed variables are specified as continuous variables (Bollen, 1989; Jöreskog, 1971). A consequence of the continuous specification in CFA is that the observed variables are required to relate linearly to the latent variables. However, in the case of ERS, the model in Figure 3.2 cannot be applied within the linear CFA framework because ERS violates the assumption of linearity (see below). In this contribution, we relax this assumption of linearity by using a Latent Class Factor Approach (LCFA) where the observed responses are specified as nominal variables and the latent constructs as ordinal variables[12]. By modeling each item category separately, assumptions concerning the items as a whole are avoided and ERS influencing the item responses in a non-monotone manner can be detected by the model in Figure 3.2.

The non-monotonicity results from the particular response pattern that ERS causes among the item responses. The respondents subject to ERS are likely to select the extreme – positive and negative – categories more often than the other item categories, thereby leading to more observations in the extreme categories at both endpoints of the response scale (Moors, 2003). In contrast, the attitudes cause a monotone effect: the more positive the attitude of a respondent is, the more likely he or she is to select a positive answer and the more unlikely he or she is to select a negative answer. Therefore, an attitude induces a linear (and thus monotone) effect on the responses whereas ERS leads to a non-monotone

[12] The same model can be estimated with continuous latent variables without altering conclusions drawn in this chapter. In that case, one should make additional restrictions to fix the location and scale of the latent variables. We chose for ordinal specification of the latent variables to facilitate the estimation procedure.

relationship between the ERS factor and the responses. Figure 3 illustrates the non-monotone effect of ERS and the monotone effect of the attitude on the size of the category item parameters, which are in the case of the LCFA approach logit coefficients.

Figure 3.3. The size of the category-specific item parameters relating the responses to the factors F1 and ERS estimated with SPVA data (N=3576)



Note. The graph is based on the model parameters for the first item, estimated under the assumption of scalar equivalence and all observed variables are nominally specified with respect to all latent variables. Parameters are logit coefficients.

The x-axis in Figure 3.3 represent the item categories on the five-point response scale that runs from *totally agree* to *totally disagree* belonging to the item “In the Netherlands immigrants get many opportunities” which is part of the SPVA survey. Note that the same pattern appears for the other items. The y-axis describes the size of the parameters in the model that corrects for ERS illustrated in Figure 3.2. The dotted line shows the category-specific item parameters representing the influence of F_1, the crossed line illustrates the category-specific item parameters representing the influence of E.

Under the influence of the attitude, the size of the parameters increases along the item categories on the response scale. This illustrates the monotone manner in which the attitude relates to the observed item response. However, in the case of ERS, the size of the parameters decreases as well as increases along the response scale. This illustrates that ERS relates in a non-monotone manner to the item (see Figure 3.3). Because of this non-monotone pattern with respect to ERS, the item responses cannot be interpreted as responses of interval variables; that is, variables measured at an ordinal scale with equal distances between the item categories. Therefore, we model the item responses as nominal variables.

A nominal specification of the observed variables leads to a separate treatment of each response category: the response of individual i to item j is denoted by Y_{ij}, a response to a particular category by c, and the number of response categories by C. Note that for each attitude five items are included in the model, each having five categories and formulated as bipolar (the so-called Likert scales). The following multinomial logit model is used to model the relationship between the item responses and the attitudes:

$$P(Y_{ij} = c \mid F_{1i}, F_{2i}) = \frac{\exp(\beta_{0jc} + \beta_{1jc} F_{1i} + \beta_{2jc} F_{2i})}{\sum_{d=1}^{C} \exp(\beta_{0jd} + \beta_{1jd} F_{1i} + \beta_{2jd} F_{2i})} \quad . \tag{8}$$

The probability of choosing category c of item j by individual i, conditional on F_{1i} and F_{2i}, is explained by the item parameters β_{0jc} representing the intercept and the parameters β_{1jc} and β_{2jc} representing the monotone relationship between the substantive F_{1i} and F_{2i} and the items. The error ε_{ij} is multinomially distributed.

As is typical of these multinomial logit models, each category c of item j has its own parameters, indicated by the index jc of the parameters β_{0jc}, β_{1jc}, β_{2jc} and β_{3jc} (Agresti, 2002). In the case of these category specific parameters, the identification of the category parameters can be accomplished by effect coding where the parameters are restricted to sum to zero across categories for each item. Another possibility would be dummy coding where the parameters are fixed to zero for one category.

To correct for the response style, we include a separate latent factor E measuring ERS in the model as is depicted in Figure 3.2. This leads to the following multinomial logit model:

$$P(Y_{ij} = c \mid F_{1i}, F_{2i}, E_i) = \frac{\exp(\beta_{0jc} + \beta_{1j} c F_{1i} + \beta_{2j} c F_{2i} + \beta_{3jc} E_i)}{\sum_{d=1}^{C} \exp(\beta_{0jd} + \beta_{1j} d F_{1i} + \beta_{2j} d F_{2i} + \beta_{3jd} E_i)} \quad , \tag{9}$$

where response Y_{ijc} is conditional on the attitudes F_{1i} and F_{2i} and E_i. The influence of ERS on the response is explained by the parameters β_{3jc} representing the influence of E_i.

With respect to the attitudes, the parameters β_{1jc} and β_{2jc} are constrained to increase monotonically across the response scale by the restriction of the parameters as $\beta_{1j} \cdot c$ and $\beta_{2j} \cdot c$. A change from one category to the next (for example from 1 to 2) would denote an increase of β_{1j} by one since the difference between categories – denoted by c – equals one. In this way, a more parsimonious model can be estimated: only one parameter is needed for each item, assuming the distance between category 1 and 2 to be equal to the distances between the other adjacent categories. This adjacent-category ordinal logit specification can be implemented within the multinomial logit model so that the parameters reflecting ERS are specified nominally while the parameters[13] describing the substantive factors are constrained to monotonicity. Note that the model for the latent means and (co)variances is also an adjacent-category logit model as the factors are specified as ordinal variables (see Appendix B).

3.4 How ERS leads to Measurement Inequivalence

Metric or scalar equivalence may be violated in only one, a few or all items. If all items in both item subsets are affected, this may be caused by a difference in the style of responding between groups, because the response style affects all items simultaneously. In other words, the presence of a response style that differs across groups is likely to import measurement inequivalence in the data. Unfortunately, most comparative studies on measurement equivalence among culturally diverse populations focus on the detection of measurement inequivalence without correcting for ERS (Mullen, 1995; Myers, Calantone, Page Jr., & Taylor, 2000; Raju, Laffitte, & Byrne, 2002; Reise, Widaman, & Pugh, 1993; Steenkamp & Baumgartner, 1998).

To show which model parameters appear as inequivalent as a consequence of ERS, we generated a data set where groups are simulated to differ in ERS. Previous studies within the latent variable framework simulated group differences in ERS by generating the item intercepts to differ across groups (Meade & Lautenschlager, 2004a, 2004b). A disadvantage of this approach is that it assumes that ERS violates scalar equivalence and invalidates the possibility that ERS violates other forms of equivalence. Cheung and Rensvold (2000) explicitly examined how ERS affects the model parameters and simulated the group differences in ERS by generating group differences in the response patterns, with a group that was severely subject to ERS having many extreme answers. By running SEM models on this data set, they showed that the difference in the response patterns leads to inequivalent

[13] In this chapter, the effects of the latent variables on the item responses are referred to as factor loadings as is usual in CFA. Due to the discrete specification of the observed variables in LCFA, these effects actually are logit coefficients. Since the factor loadings and logit coefficients are conceptually equal and the only difference is the specification of the observed variables, we refer to these effects as factor loadings.

intercepts and factor loadings. More specifically, the group subject to a high level of ERS has higher loadings and lower intercepts than the group subject to a low level of ERS (Cheung & Rensvold, 2000).

As we discussed before, these results based on the SEM approach should be viewed with caution because ERS violates the assumption of linearity. Therefore, we generated a data set based on a latent variable model that detects and corrects for ERS by specifying the observed responses nominally with respect to the ERS factor and ordinally with respect to the substantive factors. To detect measurement inequivalence, we extend the model in [9] by including an exploratory group variable *g* as follows:

$$P(Y_{ij} = c \mid F_{1i}, F_{2i}, E_i, g) = \frac{\exp(\beta_{0jgc} + \beta_{1jg} c F_{1i} + \beta_{2jg} c F_{2i} + \beta_{3jc} E_i)}{\sum_{d=1}^{C} \exp(\beta_{0jgd} + \beta_{1jg} d F_{1i} + \beta_{2jg} d F_{2i} + \beta_{3jd} E_i)}. \quad [10]$$

The response Y_{ij} is explained by the item parameters β_{0jgc} representing the group specific category intercept, the ordinally-restricted parameters $\beta_{1jg}c$ and $\beta_{2jg}c$ representing the group specific influence of F_{1i} and respectively F_{2i}, and the unrestricted parameters β_{3jc} representing the influence of E_i on the item responses. Note that although the parameters β_{3jc} (representing the influence of the style factor) are restricted to equality across groups (no superscript *g*), this assumption could be relaxed to investigate how groups differ in ERS. We allowed for group differences in the latent group means of the factor that measures ERS. By comparing models that correct and do not correct for ERS, we examine which model parameters appear to be inequivalent as the result of these group differences in ERS.

We generate a data set by a latent factor model in which five 5-category variables are related to two continuous latent variables measuring one attitude and ERS. Three groups, each consisting of 1000 observations, are assumed to differ in their style of responding by specifying different latent means for each group (μ_g = 2, 0 and -2). To ensure that the group differences in the latent means of the attitude and ERS are not confounded, the groups do not differ in the attitude. Note that although only one attitude is included in the model, the effects of the attitude and ERS on the items cannot be confounded because the items are restricted to relate to the attitude in a monotone manner and allowed to relate to ERS in a non-monotone manner. This is accomplished by assigning values to the parameters describing the relationship between the manifest and latent variables. For each item, the category parameters are restricted to relate to the attitude in a monotone way by sequentially assuming the values -2, -1, 0, 1, and 2. These values restrict the effect of the attitude to be the same for each pair of categories as the inter-category distance is always 1. Furthermore, a respondent who scores highly on the dimension is likely to choose the positive outer category (see the SPVA example depicted in Figure 3). For the category item parameters relating to ERS, the values 1.5, -1, -1, -1, and 1.5 are assumed, indicating a non-monotone pattern. These values signify that respondents with a high score of ERS are more likely to select the outer categories than the categories 2, 3 or 4 (see also Figure 3).

We estimated various models on this generated data set to find out how the group differences in ERS import measurement inequivalence in the results. For this purpose we used the syntax module of the Latent GOLD 4.5 program[14] (Vermunt & Magidson, 2008), a program for the Maximum Likelihood estimation of latent class models[15] and other types of latent variable models (see Appendix C). A comparison of the results between the models that control and do not control for the response style informs us about how ERS affects the model parameters. To compare the models, we report in Table 3.1 both the log-likelihood values and the Bayesian Information Criterion (BIC) values. The latter fit measure introduces a penalty for the sample size and the number of parameters (Burnham & Anderson, 2004; Raftery, 1999). The best model in terms of fit and parsimony has the lowest value of BIC. Note that although we choose to simulate the data using continuous factors[16], we estimate the models with ordinal factors to preserve continuity with the other models presented.

[14] See Appendix C for the details of model specification using the syntax module of the Latent GOLD 4.5 program.

[15] A well-known problem with these models is the occurrence of local minima. Here, we deal with this problem by using 100 sets of starting values, 250 iterations using the Expectation-Maximization algorithm and a low minimum convergence criterion (1e-005).

[16] Theoretically, the style factor represents a continuous dimension; however, we approach the dimension as ordinal to avoid inappropriate assumptions about normal distribution of respondents on this dimension and to facilitate the estimation process. In the Latent Class Factor Approach the latent variables have three categories (the latent classes) that are restricted to be ordinally located with equal distances on an underlying continuous dimension (see Appendix B).

Table 3.1.

Model selection estimated with generated data (N=3000)

Model	Fit Statistics Log-Likelihood	BIC (based on LL)	Number of parameters
Without a correction for ERS (two factors)			
A) Measurement inequivalence	-15976.6	32585.8	79
B) Metric equivalence	-16011.5	32575.4	69
C) Scalar equivalence	-19569.0	39370.2	29
With a correction for ERS (three factors)			
A_{ERS}) Measurement inequivalence	-15055.7	30936.1	103
B_{ERS}) Metric equivalence	-15061.1	30866.8	93
C_{ERS}) Scalar equivalence	-15130.0	30684.4	53

Table 3.1 reports the fit statistics for the models estimated on the simulated data set. For the models that correct for ERS, the model of scalar equivalence has the best model fit (model C_{ERS}). This result is expected as the measurements are simulated to be scalar equivalent. More interesting is that among the models that do not correct for ERS the model of metric equivalence is preferred (Model B). Thus, group differences in ERS cause equivalent measurements to *appear* as group differences in parameters β_{0jc}; that is, as groups having unequal intercepts. These results show that correcting for ERS is crucial in making valid conclusions with respect to metric and scalar equivalence.

3.5 An Empirical Application

We now illustrate the importance of this finding with a dataset collected[17] in 2002 among the four largest ethnic minorities in the Netherlands, namely Turks, Moroccans, Surinamese and Antilleans. Response rates lie between 44% for Surinamese and Antilleans and 52% for Turks. The SPVA survey treats the Social Position and Utility Use of Ethnic Minorities by focusing on the cultural, economic and social life of ethnic minorities in the Netherlands (Dagevos, et al., 2003). In this application, we use two sets of five questions, each subset referring to an aspect of the cultural dimension; that is, family values and the attitude toward

[17] Since the data is collected among households, we only include the answers given by the heads of the households to have independent observations.

the Netherlands. One item subset contains three items that are negatively worded; the other subset contains one item negatively worded. 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 which now runs from a negative (1) toward a positive (5) response to the items. Descriptive statistics of the items are reported in Table 3.2.

Table 3.2.

Mean observed item response per ethnic group (N=3576)

Factors and items	Turks	Moroccans	Surinamese	Antilleans
Factor 1: Attitude toward NL				
Item 1: In the Netherlands immigrants get many opportunities	3.53 (1.06)	3.42 (1.07)	3.26 (1.11)	3.25 (1.15)
Item 2: The Netherlands is hostile to immigrants	2.80 (1.02)	2.46 (0.88)	2.39 (0.88)	2.52 (0.91)
Item 3: In the Netherlands your civil rights as an immigrant are respected	3.40 (0.91)	3.55 (0.86)	3.52 (0.86)	3.45 (0.84)
Item 4: The Netherlands is a hospitable country for immigrants	3.03 (0.97)	3.48 (0.92)	3.70 (0.88)	3.60 (0.91)
Item 5: The Netherlands is tolerant towards foreign cultures	3.83 (0.91)	3.58 (0.87)	3.83 (0.82)	3.69 (0.83)
Factor 2: Family values				
Item 6: A man and woman are allowed to live together without being married	2.55 (1.25)	2.12 (1.10)	3.90 (1.05)	3.95 (1.04)
Item 7: Married people with children should not divorce	3.12 (1.19)	2.67 (1.14)	2.64 (1.09)	2.42 (1.05)
Item 8: The best family is: two married parents with children	3.54 (1.09)	4.05 (0.93)	3.47 (1.22)	3.40 (1.21)
Item 9: A daughter aged 17 is allowed to live by herself	2.00 (0.93)	1.94 (0.95)	2.41 (1.04)	2.60 (1.14)
Item 10: The opinion of the parents should be important in the choice of a partner for their child	3.45 (1.04)	3.48 (1.09)	2.64 (1.11)	2.47 (1.07)
N (unweighted)	914	858	1022	782
Response Rate (%)	52	52	44	51

Note. Items 2, 7, 8 and 10 are formulated in reversed manner where a positive answer indicates a conservative attitude. For the other items a positive answer indicates a modern attitude. Standard deviations between parentheses.

We estimated various models for our data set; as in the generated data example, the model selection is based on log-likelihood and BIC values. We use a LCFA model that corrects for ERS by including an ERS factor and simultaneously tests for measurement equivalence described in [10].

As in the generated example, we specified six models of which three models correct for ERS. The first model depicts measurement inequivalence where the latent means, the latent (co-) variances, the intercepts and the factor loadings are simultaneously allowed to differ across groups (see Figure 3.1c). By using the situation of measurement inequivalence as a baseline model, we avoid inappropriate assumptions about measurement equivalence. To test for metric equivalence, this baseline model is compared to a more restrictive model where the factor loadings are restricted to equality across groups. Scalar equivalence is tested by additionally restricting the intercepts to equality across groups[18] (Byrne, et al., 1989; Vandenberg & Lance, 2000); if the model fit does not deteriorate significantly, the restrictions are confirmed to be appropriate. To investigate whether the conclusions with respect to metric and scalar equivalence are affected by ERS, these three models are re-estimated while controlling for ERS. In Table 3.3, the fit statistics are reported for all models.

[18] The model selection does not include partial equivalence models where only some of the factor loadings are restricted to equality because this chapter focuses on how the presence of a response style affects measurement equivalence. Since a response style is assumed to affect all items simultaneously, it presumably violates equivalence of all items simultaneously.

Table 3.3.

Model selection estimated with SPVA data (N=3576)

Model	Fit Statistics: Log-Likelihood	BIC (based on LL)	Number of parameters
Without a correction for ERS (two factors)			
D) Measurement inequivalence	-44153.0	90057.0	214
E) Metric equivalence	-44283.8	90073.1	184
F) Scalar equivalence	-45088.6	90700.8	64
With a correction for ERS (three factors)			
D_{ERS}) Measurement inequivalence	-41819.7	85758.5	259
E_{ERS}) Metric equivalence	-41901.3	85676.2	229
F_{ERS}) Scalar equivalence	-42564.0	86019.8	109

Comparing Models D, E and F in Table 3.3 illustrates that according to the BIC values Model D – the baseline model – is preferred: the model in which the factor loadings as well as the intercepts are allowed to differ between groups. The increase in BIC values of Models E and F compared to Model D confirm that the equality restrictions are inappropriate. However, this conclusion clearly alters when the same analyses are controlled for ERS in Models D_{ERS}, E_{ERS} and F_{ERS}. First, the model fit improves substantially between the models with and without a style factor which illustrates the necessity of introducing an ERS factor into the model. Controlling for ERS causes the model with unequal intercepts and equal factor loadings (Model E_{ERS}) to fit best. Thus, accounting for ERS yields a substantive reduction in the group differences in the factor loadings. The magnitude of the reduction is evaluated by inspecting the WALD statistics which allow to test whether β_{1jg} is equal across groups g for each item j belonging to factor F_1 (Buse, 1982; Vermunt & Magidson, 2005a).

Table 3.4.

WALD statistics for group differences in the intercepts and loadings for model D and D_{ERS}

	Intercepts		Factor loadings	
	Model D [a]	Model D_{ERS} [a]	Model D	Model D_{ERS}
Item 1	55.21***	131.87***	8.73*	18.68***
Item 2	209.20***	63.38***	4.42	6.13
Item 3	97.38***	36.37***	8.31*	5.42
Item 4	220.07***	71.68***	23.37***	3.76
Item 5	150.91***	146.37***	48.46***	46.73***
Item 6	550.63***	243.18***	20.75***	16.33**
Item 7	134.07***	63.75***	29.67***	27.08***
Item 8	100.34***	90.10***	13.01**	15.15**
Item 9	65.05***	46.66***	43.50***	32.66***
Item 10	302.62***	124.84***	16.71**	4.36

Note. Start values are used to ascertain that the four ethnic groups have positive factor loading parameters, which decreased the fit of Model D (BIC=90086) and Model D_{ERS} (BIC= 85724) somewhat.
* $p<.05$; ** $p<.01$; *** $p<.001$

Table 3.4 reports the WALD statistics for the group differences in the intercepts as well as the factor loadings of Model D and Model D_{ERS}. The large number of substantial reductions in the WALD statistics show that controlling for ERS decreases the group differences in both the intercepts and the factor loadings. However, the decrease in values of the WALD statistics is larger with respect to the intercepts than with respect to the factor loadings. These results indicate that the group differences are diminished substantially by controlling for the response style, and especially the group differences in the intercepts. The fact that the measurements are not scalar equivalent, even after controlling for ERS, is likely to be caused by unknown causes not taken into account in this model. These findings are in accordance with the results in the generated data example where controlling for ERS decreased the group differences in the intercepts. Therefore, we conclude that the ERS factor partly explains the group differences in the intercepts of the set of items that were taken from the SPVA data.

3.6 Conclusion

This chapter demonstrates that ERS imports inequivalence in measurements among groups if this response style is not explicitly controlled for. This conclusion is drawn from separate findings. First, correcting for ERS reduces the measurement inequivalence in both the item intercepts and the factor loadings. This conclusion holds in the case of secondary data as well as in the generated data set. Using a LCFA multiple-group analysis, we find that the presence of ERS violates metric and scalar equivalence in the models that do not control for ERS. In the generated data set the presence of group differences in ERS violates scalar equivalence. In the Dutch data set of the four largest minorities, the group differences in ERS violate scalar as well as metric equivalence. The group differences in the intercepts that remain after controlling for ERS are ascribed to unknown group differences not considered here.

In this chapter we focused primarily on the extent to which ERS leads to measurement inequivalence when comparing attitudes across culturally diverse groups. However, the model can be extended by including covariates to control for other possible socio-demographic or cultural group differences, for instance language proficiency, level of education or gender. Additionally, the assumption that ERS is measured equivalently across culturally diverse groups could be relaxed by allowing the factor loadings related to ERS to differ between groups. Finally, one could specify a more parsimonious model by assuming that ERS affects all items similarly.

To conclude, in this chapter we show that equivalent measurements could appear as inequivalent measurements because of a group difference in ERS for which the researcher does not control. Thus, to investigate metric and scalar equivalence adequately, one should control for ERS. We have shown that Latent Class Factor Analysis is a straightforward method to appropriately investigate measurement equivalence because it enables multiple-group analyses while simultaneously correcting for ERS.

Appendix B. The model for latent means and (co) variances

The model for the means and the (co)variances of latent variable k for group g can be represented as:

$$F_i \sim N(\mu_g, \Sigma_g), \qquad [11]$$

where the multivariate vector $\mathbf{F}_i$ is normally distributed with a vector containing group specific means $\boldsymbol{\mu}_g$ and a group specific co-variance matrix $\boldsymbol{\Sigma}_g$. Using dummy coding, the factor means are restricted to zero in one group. In the empirical application of the model using the SPVA data set this reference group is the Turks.

In the case of the regression model used for the ordinal latent variables, an adjacent-category ordinal logit model as is described in [5] is used:

$$P(F_i = k \mid g) = \frac{\exp(\gamma_{0k} + \gamma_{1g} \cdot k)}{\sum_{k'=1}^{3} \exp(\gamma'_{0k} + \gamma_{1g} \cdot k')} \quad , \qquad [12]$$

where the probability that respondent *i* belongs to class *k* of variable *F* is estimated given the respondent's group membership *g*. As one can see, the model has a similar structure as the model in equation [10] where the observed responses are modeled as a function of latent variables. In equation [11] the latent variables are modeled as a function of ethnicity. Note that the group structure in the latent means and co-variances described in equations [11, 12] applies to every estimated model in this chapter.

Appendix C. Latent GOLD 4.5 syntax used for assessing measurement equivalence

We used the syntax module of Latent GOLD 4.5 to estimate models A to F and Model A_{ERS} to F_{ERS} from Table 1 and Table 3. The variables and equations sections of the syntax file for the most complex model D_{ERS} is 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|ethnicity + (~ord) F1|ethnicity + ERS;
        Y6 – Y10       <- 1|ethnicity + (~ord) F2|ethnicity + ERS;
```

.

In the variables section we provide the relevant information on the dependent, independent, and latent variables to be used in the analysis. 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 modelled as a conditional effect depending on the group. In other words, the association between F1 and F2 is group specific. 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 "lethnicity" yields a model without ethnic group difference in the intercepts and the factor loadings representing respectively scalar and metric equivalence, 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 factor.

Chapter 4*

Exploring the Response Process of Culturally Differing Survey Respondents with a Response Style: A Sequential Mixed-Methods Study

* This chapter has been conditionally accepted for publication as: Morren, M, Gelissen, J. P. T. M. and Vermunt, J. K. (cond. accepted). Exploring the response process of culturally differing survey respondents with a response style: a sequential mixed-methods study. *Field Methods*

4.1 Introduction

Studies on people's values, attitudes and opinions use the survey as the key instrument to measure these characteristics. Especially survey instruments intended for making comparisons among national cultures or among cultural groups within nations should be equally applicable across cultures to validly compare attitudes. Unfortunately, research has repeatedly shown that survey results are not always comparable across cultures (Johnson, et al., 1997; Van de Vijver & Leung, 1997; Van de Vijver & Phalet, 2004). One important explanation for this finding is that heterogeneous response behavior causes the responses to attitude questions – in particular those dealing with culturally sensitive topics – to be incomparable across diverse cultural groups (e.g. Harzing, 2006; Johnson, et al., 2005; Van Herk, et al., 2004). As an example, respondents from culturally diverse groups may have different reasons to choose 'totally disagree' on a five-point Likert scale of a survey question such as 'A man and woman are allowed to live together without being married' – apart from the traditional family values being measured by the item: Ethnic Moroccan people may not even consider some response categories because of their Islamic background whereas ethnic Turkish people may respond extremely because they seek to uphold their honor while answering this question.

To explore such group differences in response behavior one ideally needs statistical methods that simultaneously measure the underlying attitudes and detect the group specific response behavior. In this chapter, we use such a modeling approach, namely Latent Class Factor Modeling (Moors, 2003; Morren, Gelissen, & Vermunt, (in press a)). This modeling approach takes into account group differences in response styles as well as in attitudes. Strong points of this method are that strict statistical assumptions need not to be met for its application and that variables of different measurement levels can be included in a model. Such statistical modeling provides insight in the magnitude and the correlates of this differential response behavior; however, statistical analyses do not clarify how the response processes underlying the differential response behavior operate. In this chapter we present a novel mixed-method approach that integrates the strengths of statistical modeling and cognitive interviewing: In the first study, we estimate a statistical model on data from a large-scale nationally representative survey to detect the magnitude of group differences in responding. In the second study, we conduct cognitive interviews in a small purposive sample of respondents to study the response process in more detail. We integrate both data sets by inferring the response style of respondents in the purposive sample from the statistical model, and comparing the respondents' explanations for their response behavior in the cognitive interview accordingly.

This approach guarantees a strong mix of quantitative and qualitative methods: because identical survey questions are the basis for both methods, the sample strategy is based on quantitative results, quantitative analyses are used to infer response styles of interviewees, and qualitative results are used to clarify quantitative results. Generally, cognitive interviews are used as a pre-test measure to design survey items (Willis, 2005). In contrast, we implement cognitive interviews as a post-test measure to evaluate the response process related to particular survey items for culturally diverse respondents with differential response styles. In summary, in this particular application of the mixed method approach, we aim to find an answer to the following research questions:

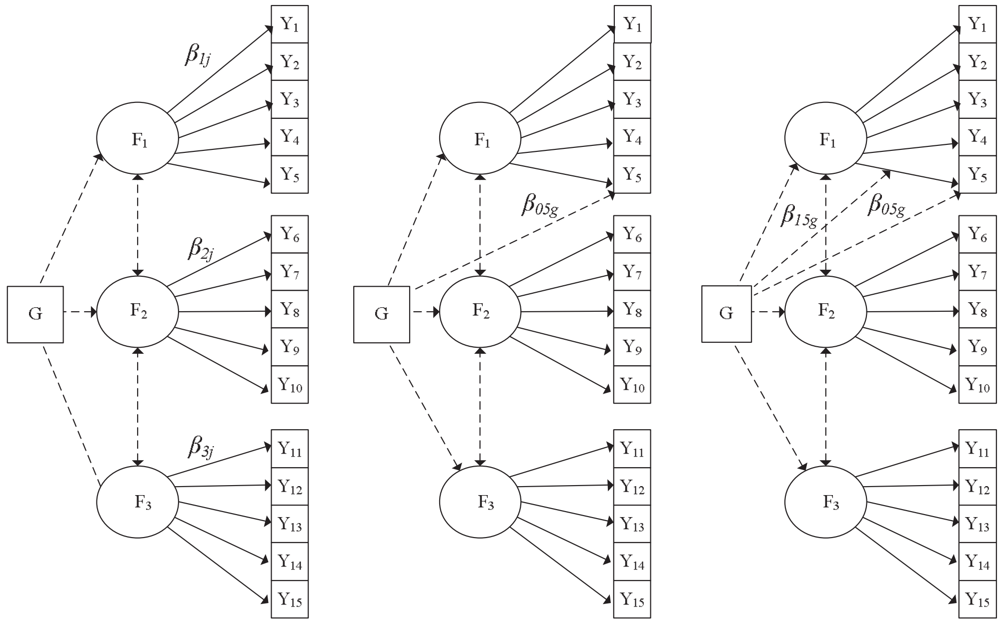
1. Which respondents endorse what type of response style, and in particular is response style usage systematically related to cultural background?
2. Is the response process for respondents with a particular response style systematically different from the response process of other respondents?

4.2 Study 1: Analysis of large-scale survey data to assess measurement equivalence and detect response style

4.2.1 Latent Variable Modeling to Assess Cross-Cultural Comparability

Within the Latent Variable Modeling (LVM) approach, attitudes are approached as complex theoretical constructs for which multiple empirical indicators, which reflect several important aspects of the attitude, are needed to obtain a valid and reliable measurement of the construct (Bollen, 2002; Skondral & Rabe-Hesketh, 2004). In this approach, an attitude is modeled as a latent – unobserved – variable (also called factor or trait) for which survey questions (hereafter called items) are used as the indicators or observed variables. Covariates are included in this basic modeling approach to investigate differences between groups as is illustrated in Figure 4.1:

Figure 4.1. The Latent Variable Model, measurement equivalence



4.1a: Scalar equivalence   4.1b: Metric equivalence   4.1c: Measurement inequivalence

Here, Y_1 to Y_{15} represent the item responses which are directly related to the latent variables measuring the attitudes F_1, F_2 and F_3, and indirectly related to the observed group variable G. The systematic variance among the item responses is captured by the factor loadings β_{1j}, β_{2j} and β_{3j}, which are the parameters for the direct relations between the latent variables and the item responses (index *j* refers an item, thus takes on values from 1 to 15). In the models depicted in Figure 4.1, it is assumed that the five items in the first item subset do not relate directly to the second and third attitude which is modeled by fixing the item parameters β_{2j} and β_{3j} to zero. Likewise, the parameters β_{1j} and β_{3j} are fixed to zero for items Y_6 to Y_{10} and parameters $\beta_{1\mathrm{j}}$ and β_{2j} are fixed to zero for items Y_{11} to Y_{15}.

This latent variable model assumes that each respondent with a certain attitude level will respond similarly to the items irrespective of his or her group membership, a condition referred to as equivalence. Note that measurement equivalence allows for group differences in attitudes but not for group differences in the measurement of these attitudes. There are two forms of measurement equivalence: Scalar equivalence and metric equivalence. Scalar equivalence is the most strict form and is obtained when respondents' responses are identically related to the attitudes across groups, i.e. if the item intercepts and the factor loadings are equal across groups (see Figure 4.1a), whereas metric equivalence is obtained if the groups have similar factor loadings (Steenkamp & Baumgartner 1998). As an example, in

Figure 4.1b a violation of scalar equivalence is indicated by the direct effect β_{05g} (dashed arrow) of G on item Y_5, and in Figure 4.1c, a violation of metric equivalence is indicated by the direct effect β_{05g} and the interaction effect on the relationship β_{15g} between item Y_5 and factor F_1 meaning that G moderates the association between the fifth item and the first factor.

We test measurement equivalence by comparing measurement inequivalence models to more restrictive models assuming metric or scalar equivalence. Thus, in our example, to test for scalar equivalence we fix the group-specific parameters β_{05g} and β_{15g} to zero; to test for metric equivalence only β_{15g} is fixed to zero. Even though in Figure 4.1 the group parameters are only indicated for item Y_5, the group variable G is usually assumed to affect all items simultaneously.

As the first step in the empirical analysis, we applied the above procedure to data collected[19] in 2002 among the four largest ethnic minorities in the Netherlands, namely Turkish, Moroccan, Surinamese and Antillean people. We include 15 items from this SPVA survey[20], each having five ordered response categories that range from 'totally agree' to 'totally disagree'. These items operationally define three attitudes, namely the attitude toward the Netherlands, the endorsement of traditional family values, and the autonomy of children within the family. Given the substantive nature of these constructs, the groups are expected to be subject to culturally specific sources of measurement error and possibly to derive their answers from diverse response processes. Table 4.1 reports descriptive statistics of all items included in the analyses for the four cultural groups:

[19] Since the data is collected among households, we only include the answers given by the heads of the households to secure independent observations.

[20] SPVA stands for Social Position and Utility Use of Ethnic Minorities. The survey maps the cultural, economic and social life of ethnic minorities in the Netherlands (Dagevos et al., 2003). We thank Data Archiving and Networked Services (DANS) for providing the data files.

Table 4.1.

Mean observed item response per ethnic group (N=3549)

Factors and items		Turks	Moroccans	Surinamese	Antilleans
Factor 1: Attitude toward the Netherlands					
Item 1	In the Netherlands immigrants get many opportunities	2.48 (1.06)	2.58 (1.07)	2.74 (1.11)	2.75 (1.15)
Item 2	The Netherlands is hostile to immigrants	3.20 (1.02)	3.53 (0.88)	3.60 (0.88)	3.48 (0.91)
Item 3	In the Netherlands your civil rights as an immigrant are respected	2.97 (0.97)	2.53 (0.91)	2.31 (0.89)	2.40 (0.91)
Item 4	The Netherlands is a hospitable country for immigrants	2.60 (0.90)	2.44 (0.85)	2.48 (0.86)	2.56 (0.84)
Item 5	The Netherlands is tolerant towards foreign cultures	2.17 (0.91)	2.43 (0.87)	2.16 (0.82)	2.31 (0.82)
Factor 2: Autonomy of the children					
Item 6	Children should live at home until marriage	2.30 (1.04)	2.24 (1.12)	3.07 (1.27)	3.41 (1.23)
Item 7	Elderly must be able to move in with their children	2.88 (1.13)	2.21 (0.96)	2.90 (1.15)	2.99 (1.17)
Item 8	Adult children should be able to move in with their parents	2.11 (0.88)	2.06 (0.85)	2.68 (1.08)	2.87 (1.11)
Item 9	Parents always have to be respected, even if they do not deserve it based on their behavior or attitude	1.89 (0.83)	1.80 (0.89)	2.39 (1.11)	2.30 (1.09)
Item 10	Older family members should have more influence in important decisions (for instance about moving) than younger ones	2.89 (1.15)	2.63 (1.12)	3.15 (1.12)	3.12 (1.09)
Factor 3: Family values					
Item 11	A man and woman are allowed to live together without being married	3.46 (1.25)	3.88 (1.11)	2.10 (1.04)	2.05 (1.04)
Item 12	Married people with children should not be allowed to divorce	2.88 (1.19)	3.33 (1.14)	3.36 (1.09)	3.59 (1.05)
Item 13	The best family remains to be: Two married parents with children	2.46 (1.09)	1.94 (0.93)	2.53 (1.22)	2.60 (1.21)

Item 14	A daughter aged 17 is allowed to live by herself	4.00 (0.93)	4.06 (0.95)	3.59 (1.04)	3.40 (1.14)
Item 15	The opinion of the parents has to be important in the choice of a partner for their child	2.54 (1.04)	2.52 (1.09)	3.36 (1.11)	3.53 (1.07)
N (unweighted)		905	854	1014	776
Response Rate (%)		52	52	44	51

Note. The standard deviations are in parentheses. The items 2, 12, 13 and 15 are formulated in a reversed manner where a positive answer indicates a conservative attitude. Totally agree is indicated by 1, agree by 2, neither agree nor disagree by 3, disagree by 4, totally disagree by 5.

Without being complete, explanations for violations of measurement equivalence among culturally diverse groups range from cultural differences in response styles (Baumgartner & Steenkamp, 2001; Clarke III, 2001), the socio-economic background (Light, Zax, & Gardiner, 1965), language use (Bachman & O'Malley, 1984), to acculturation (Marin, et al., 1992). To investigate these explanations, we separately include the following covariates as controls: Language use during the interview, generation of immigration, age, income and level of education in the Netherlands. Besides models with these substantive control variables, a model with a latent response style factor[21] as a control is also estimated (Moors, 2003). The latent style factor defined to be an ordinal latent variable with three categories[22]. No assumptions are made about the distribution of respondents along the underlying dimensions measured by the factors. In all models, the effect of attitudes on item responses is distinguished from the effects of the above control variables by estimating models in which only one item subset is affected by one attitude while all items are simultaneously affected by the control variables. By comparing the statistical fit of models with and without these covariates, we investigate to which extent the covariates form a reasonable explanation of the inequivalence. In our data set, only the model including the response style factor explains part of the measurement inequivalence; the models including the other control variables remain measurement inequivalent. Thus, the findings indicate that response style differences are important for explaining cultural differences and warrant further investigation. Therefore, in the following we only report and discuss the findings from the model including the response style factor as a control.

All models were estimated using the Latent GOLD 4.5 program (Vermunt & Magidson, 2008) which is developed for the maximum likelihood estimation of latent structure models. To compare between models we report in Table 4.2 both the log-likelihood values and the values of the Bayesian Information Criterion (BIC). The latter information criterion introduces a penalty for the sample size and the number of parameters (Burnham & Anderson, 2004; Raftery, 1999) and is the most widely used measure for model selection in Latent Class Analysis (Magidson & Vermunt, 2004). The best model in terms of fit and parsimony has the lowest value of BIC.

[21] See Appendix D for a detailed model specification.

[22] The three latent classes can be regarded as three ordered categories because they are scored -1, 0 and 1.

Table 4.2.

Model selection estimated with SPVA data (N=3549)

Model	Fit Statistics Log-Likelihood	BIC (based on LL)	Number of parameters
Without a correction for inequivalence			
A) Measurement inequivalence	-66075.41	134823.86	327
B) Metric equivalence	-66345.42	134996.02	282
C) Scalar equivalence	-67499.45	135832.70	102
Corrected for response style			
A_{RPS}) Measurement inequivalence	-61464.97	126134.32	392
B_{RPS}) Metric equivalence	-61588.42	126013.36	347
C_{RPS}) Scalar equivalence	-62594.19	126553.51	167

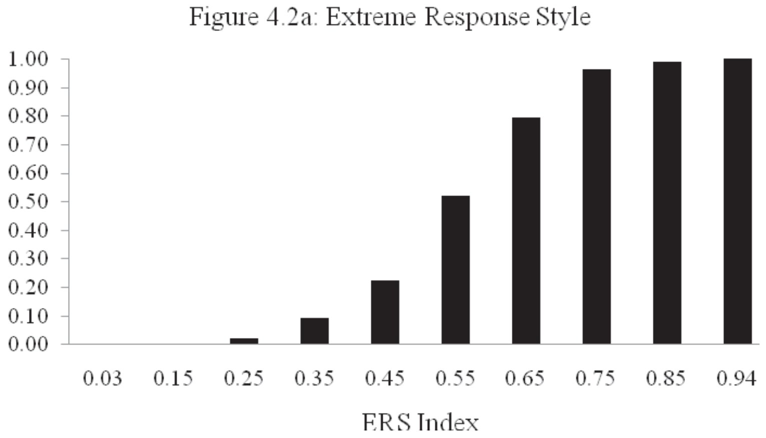
The findings indicate, first, that without controlling for a response style the ethnic minorities interpret the items differently given their attitude; close inspection of the parameters shows that these differences are visible in the intercepts as well as the factor loadings which indicates full measurement inequivalence (also indicated by Model A having the lowest BIC value). Furthermore, we find that the style factor accounts for the group differences in the factor loadings: The BIC values decrease between Model A_{RPS} and Model B_{RPS}, which indicates that the model assuming metric equivalence should be preferred. Even though group differences in the intercepts (not reported here) also decrease substantially once a control for response style is included, scalar equivalence is not attained. Note that the cultural group membership also influences response style directly: The Surinamese respond significantly less extreme than the other groups; the other groups do not differ significantly in their style of responding.

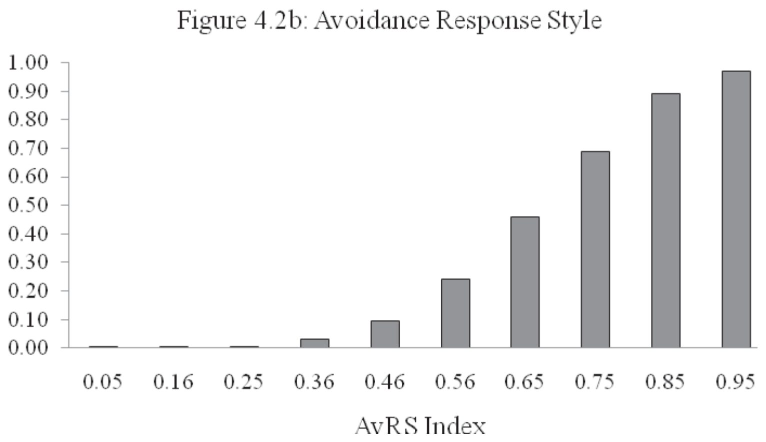
Based on a post-hoc interpretation of the parameters of the response style factor, we conclude that this style factor measures two related types of response style. First, a person who scores high on the style factor is more likely to select the extreme categories than other respondents with similar attitudes. We refer to this response style as the Extreme Response Style (ERS); i.e. the tendency to select extreme categories irrespective of the item content. Second, close inspection of the parameters shows that a person who scores low on the style factor avoids the extreme categories and selects the adjacent categories *agree* and *disagree*.

Here, this response style is referred to as the Avoidance Response Style (AvRS), and should not be confused with the tendency to select the midpoint category (Midpoint Responding, MPR) or to select the categories around the midpoint category (Response Range, RR) (see Baumgartner & Steenkamp, 2001).

As we have previously mentioned, the style factor is specified as a latent variable with three ordered classes of which Class 3 represents ERS and the class at the other endpoint (Class 1) represents AvRS. The middle class (Class 2) defines a position in between these two endpoints. We interpret this position as a mild response style and treat these respondents as not endorsing any response style. Based on the item responses of a respondent, we can compute the probability of belonging to each of these three classes. Typically we would assign a respondent to the most likely class. Practically this means that respondents who preferably select extreme categories are assigned to Class 3 (12%), and those who prefer to select adjacent categories are assigned to Class 1 (46%). The remaining respondents are assigned to Class 2.

Figure 4.2. The estimated class membership probability for the ERS (AvRS) class for different values of the ERS (AvRS) index (N=3549)





Note. The ERS (AvRS) index is the proportion of extreme (adjacent) responses across the 55 Likert items of the SPVA survey. The estimated class membership probabilities are computed per observed deciles of the ERS (AvRS) index.

To validate the ERS (AvRS) class from the Latent Class Factor Model with a response style (Model B_{RPS} in Table 4.2), we compared these with an ERS (AvRS) index based on all 55 5-point Likert items from de SPVA survey. As suggested by (Clarke III, 2001; Greenleaf, 1992b), and ERS index is obtained as the proportion of items answered in the extreme categories (1 or 5). Moreover, the AvRS index is computed as the proportion of items answered in the adjacent categories (2 or 4). Figure 4.2a depicts the estimated probability of belonging to the ERS class for different values (the 10 deciles) of the ERS index. As can be seen, the probability of belonging to the ERS class increases as the proportion of extreme responses in the 55-item set increases. Figure 4.2b shows a similar pattern for the AvRS index; that is, the probability of belonging to the AvRS class increases as the proportion of adjacent category response increases. These figures confirm that the categorical latent variable that we assume to capture ERS and AvRS is indeed doing so.

4.3 Study 2: Cognitive interviewing to explore the response process of individuals who use a response style

4.3.1 Data Collection

Cognitive interviews were held from December 2009 until May 2010 among 24 interviewees that were recruited via different contact persons, organizations, and personal contacts to avoid overlap in social background. Due to privacy restrictions, it was not possible to contact the original respondents from the large scale survey. We aimed for a heterogeneous sample as we found from Study 1 that the style of responding was not systematically related to socio-economic characteristics. We interviewed 7 Moroccan, 7 Turkish, 5 Antillean, and 5 Surinamese interviewees. From these persons, 12 belong to the second generation of immigrants, 13 are highly educated, 10 are female, and 8 are older than 40 and 4 are younger than 26. The interviews were conducted at home (14), at our university (5), at a public place (3) or at the workplace (2). The interviews lasted 45 minutes on average. During 8 interviews a research assistant was present who also helped with transcribing and coding. Five interviews were excluded from further analyses because 2 were used as a pilot, 2 were conducted among family of an interviewee, and once a translator was used. All interviews were conducted in Dutch.

We first ask the interviewees to answer particular survey questions that are also used in Study 1 in a similar fashion of a regular survey interview[23]. Based on the responses that interviewees give to the survey questions, we are able to apply the results obtained in Study 1 to determine whether they use a particular response style. After the interviewees have given their initial response, we probe them retrospectively about a selection of questions proven to elicit inequivalent answers across minorities based on the findings from Study 1. We also

[23] The respondent is asked to select one answer category that reflects his or her opinion best. The interviewer does not provide further explanation.

probe the questions to which the interviewees show interpretational or other problems in the first interview phase (see Appendix F for interview protocol). In general, the lower educated respondents[24] showed difficulties in understanding the purpose of the cognitive interviews. More specifically, they could not see why they should explain their answers given during the first interview phase. Presumably, a straightforward think-aloud interview would have been more suitable but we persisted with the two-phase interview to maintain comparability with the other interviews.

4.3.2 Coding scheme

All interviews were recorded and transcribed by the first investigator and a trained research assistant. To analyze the qualitative data from the interviews, a code list was developed by integrating data-driven and theory-driven codes. First, we approached the interview transcripts by open coding (Boeije, 2005; Miles & Huberman, 1994). However, we found that the resulting coding scheme was not helpful in answering the research questions posed earlier. Therefore, we coded the interviews again, now including thematic codes based on the theory of psychology of the survey response (Tourangeau, Rips & Rasinski, 2000). We paid special attention to elements indicated by previous research to the wide-spread occurrence of ERS in cross-cultural research: 1) socio-cultural characteristics of respondents such as language use, acculturation (Bachman & O'Malley, 1984; Gibbons, et al., 1999; Van Hemert, et al., 2001; Marin, et al., 1992), 2) type and content of questions (i.e. number of categories, labeling, language) (Bond & Yang, 1982; Hui & Triandis, 1989; Ralston, et al., 1995; Triandis & Marin, 1983; Yang & Bond, 1980), and 3) personal attributes such as intelligence, extraversion, or collectivist-individualist culture (Berg & Collier, 1953; Chun, et al., 1974; Light, et al., 1965). The integration of both coding schemes was considered satisfactory and after several debriefings among members of the research team this led to the final version of the code list (see Appendix G).

4.4 Classification of the interviewees in the cognitive interviews

We integrate quantitative and qualitative methods by calculating individual latent scores for the interviewees in the cognitive interviews based on their responses to the 15 items in Table 4.1. For two important reasons we do not use the best fitting Model B_{RPS} for classifying the interviewees but instead the most parsimonious model Model C_{RPS} assuming scalar

---

[24] We include primary education, the lower level of high school, and the lower level of professional education.

equivalence. First, the number of parameters increases considerably when allowing for group differences in Model A_{RPS} and Model B_{RPS}. This makes the classification of respondents according to Model B_{RPS} more uncertain than the classification of respondents according to the more simple Model C_{RPS}. Second, great uncertainty exists with respect to the nature of the resulting sample as we use purposive sampling in the qualitative part of the study. This means that using a relatively simple and parsimonious model may be more appropriate. Based on these considerations, we decided to use the parameters from Model C_{RPS} for the classification of respondents in Study 2 (see Appendix D). The class assignments for the response style factor serve as the basis for the comparison among the respondents' explanations for their answers.

Table 4.3.

The classification of a selection of respondents in the cognitive interviews

Demographics					Response pattern		Response style
Minority	Education	Age	Immigration	Gender	# Extreme responses	# Adjacent responses	
Moroccan	intermediate	20	2	F	7	1	ERS
Antillean	low	65	1	F	10	5	ERS
Antillean	low	49	1	F	7	4	None
Surinamese	high	32	2	F	5	7	None
Turk	high	21	2	M	4	10	None
Moroccan	high	29	2	M	1	7	None
Moroccan	low	26	1	F	2	13	None
Surinamese	high	25	2	M	2	12	None
Turk	high	27	1	F	0	8	AvRS
Antillean	intermediate	42	1	M	0	14	AvRS

Note. Immigration stands for generation which immigrated, # stands for the number of extreme or midpoint responses to the 15 items.

Table 4.3 reports the estimated response style for a selection of the interviewees who participated in the cognitive interviews. Additionally, we report the number of extreme and adjacent responses in their response patterns and some demographics, such as the age, minority, and educational attainment. The first three interviewees in Table 4.3 are classified as ERS and are indeed more inclined to use extreme categories than the other interviewees. The last two interviewees in Table 4.3 are classified as using AvRS and are indeed more inclined to use the adjacent categories. Note that the classification of the interviewees is corrected for possible differences in attitudes.

4.5 Results

4.5.1 Extreme Responders and Avoiders

To understand why some respondents employ a response style and others do not, a clear understanding of the response process is crucial. Presumably, many respondents have problems with the attitude statements because they leave much space for interpretation. Based on our findings, we argue that respondents use response strategies to simplify the response process by narrowing down the possible interpretations to answer these questions. These response strategies are used as heuristic tools to solve a problem by finding the best possible answer, also referred to as an educated guess, an intuitive judgment or common sense (Newell & Simon, 1972; Tversky & Kahneman, 1974).

During the cognitive interviews, interviewees overcome the problems they encounter when responding to attitude statements by a) having a critical attitude toward the item wording or survey process in general, b) carefully weighing arguments in mapping the answers to the response scale, i.e. selecting a category, or c) excluding personal information in the arguments. The most important difference between extreme responders and avoiders on the one hand, and the respondents without a response style on the other hand, was that the first group applies these response strategies systematically to all questions. The use of language proves to be a complicating factor: Some respondents do not clearly distinguish between extreme and adjacent categories; others find extreme categories not useful.

4.5.2 Critical attitude

The findings indicate that avoiders are very critical towards the questions and to the survey process as a whole. Two avoiders focus on the auxiliary verbs used in each question; another avoider repeatedly expresses distrust toward the survey process in general and to the formulation of the questions specifically: “All these questions are trick questions”. Yet another avoider persistently argues from the start of the interview that she does not want to exclude the possibility of other norms and values. Compared to other interviewees the avoiders are more sensitive and extreme responders are less sensitive toward the auxiliary verbs that determine the grammatical mood of the sentence. The problem addressed about “have to” in an item wording is that – although one can relate to the item content – one disagrees with the obligation expressed by this auxiliary verb. The verbs ‘should’ or ‘ought to’ create a problem for some interviewees because they feel that they cannot agree or disagree to a recommendation expressed by this auxiliary verb: “Who are we to judge other people’s lives”. Lastly, the auxiliary verb “is allowed to” is viewed by some interviewees as a limitation to a person’s freedom to act according to their own standards. The differences in sensitivity toward these auxiliary verbs between avoiders, extreme responders and the other respondents are illustrated in Table 4.4.

Table 4.4.

Respondents’ inconvenience with the auxiliary verbs used in the questions

	Auxiliary words		
	Obligation[1]	Recommendation[2]	Possibility[3]
Avoiders			
T17	10	7	3
A22	14	5	5
S20	8	0	0
Extreme responders			
M3	3	0	0
A9	2	0	0
A10	0	0	0
Without response style			
T6	1	2	0
T7	4	0	0
A11	6	0	1

Note. The numbers reflect the percentage of words (to correct for the individual differences of the total number of words) used during the second phase of the interview treating the auxiliary verbs. [1] Obligation is expressed by the auxiliary ‘have to’ or ‘must’; [2] Recommendation is expressed by the auxiliaries ‘should’ or ‘ought to’; [3] Possibility is expressed by ‘could’, ‘might’ or ‘is allowed to’

Table 4.4 reports the percentage of words used by nine respondents to address their inconvenience with the auxiliary words in the question. On average, the extreme responders spend 1.9% of the interview on explaining the influence of auxiliary verbs, avoiders 15.2% and the other respondents 4.7%.

4.5.3 Careful vs. Abrupt Argumentation

In line with our previous findings, the avoiders respond more carefully than the extreme responders: They contemplate on the multiple meanings of words, and select a category, often reluctantly, after weighing their arguments. Two avoiders in our sample feel ambiguous toward the items and find it difficult to select a response category that reflects their opinion. One of them doubts a lot about selecting categories because she feels that many answers are possible. For example, she discusses her answer to item 2 as follows: “If you take everything in consideration, I believe that - during the last few years - we have evolved in that direction”. After hesitation, she selects the midpoint category as she cannot know how this process will proceed. The other avoider is annoyed with the vague formulation of the questions and responds to these questions by emphasizing that not all minorities are treated the same: “some cultures yes, other cultures no”. Notwithstanding, many respondents without a response style express similar problems with the item wording, and also elaborate on the precise meaning they attach to the words. However, unlike the extreme responders and avoiders, their arguments do not systematically translate in selecting the same response categories.

The extreme responders react rather emotionally and abrupt instead of carefully weighing their arguments. For instance, one extreme responder agrees or disagrees strongly with seven out of ten statements about family relationships because they are in (dis)agreement with what is prescribed in the Quran. In response to item 2, an extreme responder, a woman with a Moroccan background (age 20), becomes agitated about the aggressive language used at minorities by people in the street, and at a national level by certain Dutch politicians. Another extreme responder argues differently but with equal intensity to the same item: “If the Netherlands would be hostile toward foreigners, do you think I would get a house? No! Do you think I would get an allowance? No!” (woman, age 49, Antillean ethnicity). In general, the extreme responders give strong reactions to the item content without considering the precise item wording.

4.5.4 Personal Argumentation

Even though avoiders and respondents without a response style display sensitivity toward the wording of the questions, the findings indicate that they respond differently to the item content. Throughout the interview, one avoider shuns giving any personal information and another avoider only once relates to his personal situation but in a rather hypothetical way: He argues that he would take care of his parents if it were not the case that they lived in Curaçao. On the other hand, respondents without a response style quite often explain their answers using personal experiences. Below we compare two respondents in their answer to item 15:

> "Many questions you ask concern issues in our personal surroundings, parents often have very a -um...well yes, an opinion that interferes with their children's opinion, and therefore I believe that [the influence of parents in the children's choice] should not be important" (woman, 26, Moroccan ethnicity)

> "It plays a great role in our community, but I do not believe that that should be a leading role so um…so yes in my view they should have a voice but um…not the choice itself […]given my reasoning I think I agree" (woman, 27, Turkish ethnicity)

Although these answers look very similar, the ethnic Moroccan respondent thinks of people in her surroundings and gives several examples subsequently (not shown here). In contrast, the ethnic Turkish respondent (the avoider) refers to her personal surroundings in a more abstract way; she does not – explicitly – integrate her personal experiences into her judgment. We argue that the ethnic Turkish respondent simplifies the response process by systematically excluding her personal experiences from her answers. As the ethnic Moroccan respondent reflects on her personal experiences, she makes an effort to regard each question separately. These findings may be generalized to the other respondents in the sample as the percentages in Table 4.5 illustrate.

Table 4.5.

Respondents' personal, general and conditional arguments

	Argumentation	
	Personal[1]	General[2]
Avoiders		
T17	8	50
T12	2	91
S20	10	49
S21	25	25
A22	12	26
Without response style		
S14	13	52
S16	13	54
M4	17	58
T7	19	56
M1	25	33
S15	26	49
M2	32	61
A11	35	11
T6	36	22
T8	37	54
M5	56	34

Note. The numbers reflect the percentage of words (corrected for individual differences of the total number of words used during the interview) spend on the understanding of the auxiliary verbs in the questions.

[1] Personal arguments include comments about the personal situation or the cultural background

[2] General arguments refer to comments about abstract situations

As can be seen in Table 4.5, the avoiders spend the smallest amount of words on personal arguments among all respondents. This leads us to suggest that including personal arguments in the answers impedes the use of an avoidant response style. One interviewee illustrates this suggestion when he retrospectively changes his answer to 'totally agree': "I didn't think of a family situation at first, […] that is the difference I guess. So earlier I answered without considering my own family and when I think about it now, I come to a different conclusion". Two avoiders with a Surinamese background do not fit into the pattern by equally using personal and general arguments while avoiding extreme categories altogether (see: Language). The extreme responders are excluded from Table 4.5 because we could not detect a general pattern among the extreme responders: Some use more personal arguments, others use more general arguments.

4.5.5 Mapping

When we probe the respondents about mapping their answers to the response scale, that is selecting or avoiding certain categories, extreme responders respond differently than avoiders and interviewees without a response style as is illustrated in Textbox 4.1:

Text Box 4.1.

Respondents reflect on their own response behavior

> 'As Muslim […], I always totally agree [*with what Gods prescribes or forbids*]' and 'God created me […] so if he gives you advice and you say no, I don't agree, yes, that sounds funny' (Moroccan woman, 20, extreme responder)
>
> 'I never choose extreme categories […] I also think it is my character that I don't like to totally exclude things' (Turkish woman, 27, avoider)
>
> 'I am a free-thinking man' (Antillean man, 42, avoider)
> 'I decide not to use *totally agree* or *totally disagree* because this leaves room for discussion, or debate or eh…interpretation so to say' (Moroccan man, 29, no response style)
>
> 'I never choose *totally agree*, because in my point of view this means you can never disagree' (Turkish man, 20, no response style)

As can be read in Textbox 4.1, respondents mention methods to select or to avoid extreme categories: Extreme responders have rather fixed ideas about their response pattern while avoiders use a more nuanced language. These results suggest that, first, extreme responders reflect less often on their own response behavior than other respondents and second, respondents without a response style select extreme categories occasionally even though – similar to the avoiders – they prefer nuance.

Avoiders claim they do not like extreme categories because they do not want to exclude contradicting situations, arguments or people's behavior. Six interviewees without a response style argue similarly but nonetheless occasionally select extreme categories, mostly to family-related questions (10 out of 12) or using personal arguments (8 out of 10). Besides the personal arguments, other aspects of the response process seem to influence the tendency to select or to avoid extreme categories. First, in general extreme responders are less articulate about the reasons for selecting categories and after probing, two out of three extreme responders are unable to explain their extreme answers. With respect to avoiders a similar argument can be made; however, it seems that they do not want to consider the extreme categories while the extreme responders have difficulties clarifying the difference between the categories.

4.5.6 Language

Among the lower educated ethnic Antillean and Surinamese respondents, seven out of nine respondents do not clearly differentiate between extreme and adjacent categories. Presumably, this can be explained by cultural differences in language usage. The languages of these groups, Papiamento and Surinaams[25], are rather expressive. Being used to express themselves more extremely in their mother tongue may lead respondents to regard the difference between 'totally agree' and 'agree' less strongly than the other minorities. For example, one ethnic Surinamese man (25) argues: "Actually, I don't really care whether I agree or totally agree". Two ethnic Antillean women (49, 65) do not differentiate between the extreme and adjacent categories which could also be connected to their response style, as they are both extreme responders. In short, eight out of ten ethnic Surinamese and Antillean respondents either explain vaguely or do not explain at all how they perceive the difference between extreme and adjacent categories. In contrast, out of ten ethnic Moroccans and Turkish respondents, only one remarks that he does not perceive a clear difference between the categories.

4.6 Conclusion

In this contribution we have found that the incomparability of responses across culturally diverse groups can be partially explained by a response style. The response style itself was found to be systematically related to cultural groups. The results of Study 1, in which we analyzed a large dataset quantitatively, showed that this style factor measures a synthesis of avoiding extreme categories on the one hand, and a preference for extreme categories on the other hand. As this response style could explain much of the variance in responding across

[25] In Suriname many languages are spoken, depending on the ethnicity one belongs to. The most important language is Dutch, but most people also speak Surinaams (Sranantongo).

the cultural groups, we further investigated this response style in Study 2, using semi-structured interviews in which we focused on the cognitive aspects of the response process. We integrated the quantitative and qualitative results by classifying the participants in the cognitive interviews according to the types of responders inferred from a model estimated on the large dataset. Subsequently, we compared the arguments given in the cognitive interviews for the interviewees who exhibited a preference for extreme categories – the extreme responders –, for adjacent categories – the avoiders – or without a clear preference.

Using this sequential mixed-method design, we found that the response style is the outcome of an interaction between different response strategies. First, the extreme responders less precisely considered the item wording than the avoiders and the interviewees without a response style. In contrast, the avoiders more carefully considered the precise item formulation than the other respondents. Second, the avoiders carefully weighed arguments for and against when responding. In contrast, extreme responders and the interviewees without a response style weighed their arguments to a lesser degree. Third, some avoiders persistently shunned personal information in their argumentation. Lastly, and most importantly, even though the extreme responders, the avoiders and the interviewees without a response style all displayed these response strategies more or less, the extreme responders and avoiders systematically translated these response strategies into selecting certain categories across a diverse range of questions. In our view, avoiders use these response strategies to simplify the response process as they believe no unambiguous answers exist to these questions. Extreme responders employ response strategies to rigidly apply their ideas to surveys.

Finally, the findings suggest that extreme responding might be related to cultural heritage. Especially respondents with a Surinamese and Antillean ethnic background disregard the difference between extreme and adjacent categories. Therefore, they are more likely to be subject to a response style. In line with research by Gibbons et al. (1999) and Hui and Triandis (1989), we suggest that the cultural background influences the meaningfulness of language. In the Antillean and Surinamese culture, language is used in a more extreme way than in the Dutch culture. This could lead these respondents to overlook the difference between extreme and adjacent categories. However, this hypothesis should be further researched as these conclusions are based on a small sample and must be interpreted with care.

This chapter illustrates that Latent Class Factor models are a powerful and straightforward method to derive response profiles from large scale surveys, which can subsequently be used to classify respondents who were not initially included in the survey but rather purposely selected. Furthermore, cognitive interviews can be fruitfully used to evaluate items in a post-survey modus. In this case, we recommend integrating the findings from the cognitive interviews with the survey findings by classifying the interviewees based on the model estimates. We have shown that respondents have problems with the vague formulation of the attitude statements. Especially a complex syntax and auxiliary verbs cause problems as they increase the number of meanings that can be attached to the questions. Interviewees can agree with one meaning and simultaneously disagree with another meaning. In the design of attitude statements, we recommend avoiding auxiliary verbs, examples and subordinate clauses.

Appendix D. Latent Class Factor Analysis

Here we provide more details about the Latent Class Factor Model with a response style we used in our analysis. This model was proposed by Moors (2003). Recently, Morren et al. (in press a) extended this model by showing that it is better to treat the relationship between substantive factors and items differently from the relationship between response style factor and items. More specifically, in their relationship with the response style factor, the item responses are treated as nominal variables, yielding five category-specific parameters per item. This means that no assumptions are made about the form of these relationships. For the attitude factors, only one parameter is used per item because for this relationship the items are treated as ordinal variables. More specifically, we assume that

$$P(Y_{ij} = c \mid F_{1i}, F_{2i}, F_{3i}, R_i) = \frac{\exp(\beta_{0jc} + \beta_{1j} c F_{1i} + \beta_{2j} c F_{2i} + \beta_{3j} c F_i + \beta_{4jc} R_i)}{\sum_{d=1}^{C} \exp(\beta_{0jd} + \beta_{1j} d F_{1i} + \beta_{2j} d F_{2i} + \beta_{3j} d F_i + \beta_{4jd} R_i)}. \quad [13]$$

This is a hybrid between a multinomial and an ordinal logit model. The β's are the item parameters to be estimated: β_{0jc} is an intercept term for item j and category c, β_{1j},β_{2j} and β_{3j} are slope parameters corresponding to the three substantive factors, and β_{4jc} are the slope parameters for the response style factor denoted by R_i . The parameters β_{1j}, β_{2j}, and β_{3j} are multiplied by the category number c, which results from the ordinal specification for the relationships with the substantive factors. Note that some of these parameters are fixed to 0 because each item loads on only one substantive factor. The other model parameters are category specific.

Chapter 5*

Response Strategies and Response Styles in Cross-Cultural Surveys

* This chapter has been submitted for publication.

5.1 Introduction

In the last decades of the 20th century, many Western societies have transformed into multicultural societies as the result of a steady immigration flow. Ethnic minorities consist of a quarter of the US population and the prognosis is that by 2050 ethnic minorities will form the majority. In Europe, immigrants consist of 14% of the population on average (Pan & Pfeil, 2003) and 85% of the total Europe's total population growth results from immigration in 2005 (Munz, Straubhaar, Vadean, & Vadean, 2006). As these societies become more multi-cultural in nature, social scientists have become increasingly interested in the differences and similarities in values, attitudes and opinions that may exist between different groups of immigrants and between immigrant and natives.

To investigate these issues, surveys are usually the instrument of choice to gather attitudinal information on diverse populations. Obviously, when surveys are applied in a cross-cultural design, the issue of the cross-cultural comparability of survey findings becomes increasingly important (Van de Vijver & Leung, 1997). Some survey research only marginally pays attention to – or even completely ignores – that people may respond differently in surveys because they come from diverse cultural backgrounds. Overlooking this issue may lead to erroneous conclusions about group differences among culturally diverse populations. Fortunately, cross-cultural researchers more and more test the comparability of survey measurements empirically (Vandenberg & Lance, 2000). However, in order to adequately study the comparability of survey measurements, it is not sufficient to only establish whether a particular survey measurement constitutes an equivalent or inequivalent measurement across different cultural groups: Once measurements are found to be inequivalent, the causes for this should also be further investigated, so that cross-cultural survey measurements can be improved. Causes for inequivalence of measurements can be manifold, but in this chapter we focus on one particular issue which has not been systematically investigated in previous research; that is, the response strategy and response style that respondents may use within the framework of a cross-cultural survey. This is done against the backdrop of findings from a quantitative study on the assessment of measurement equivalence and the detection of response style for a large-scale cross-cultural survey. Specifically, this chapter sets out to answer the following research question: Do respondents participating in cross-cultural surveys differ in terms of their response style and response strategy when responding to attitude statements, and if so are these characteristics affecting the response process associated with a respondent's ethnicity and generation of immigration?

Since the 80s, survey researchers approach the survey response as the outcome of cognitive, communicative and social processes (Belson, 1986; Bradburn, Rips, & Shevell, 1987; Hippler & Schwarz, 1987; Schwarz & Sudman, 1996; Tourangeau, 1987; Tourangeau & Smith, 1996). Each respondent is assumed to go through five stages: Interpreting the question, retrieving information, generating the judgment, mapping the judgment to the response scale and editing the response. Although theoretical models on the response process occasionally discuss how culturally diverse respondents may differ in this response process (Hui & Triandis, 1989; Tourangeau, et al., 2000, pp. 210-213), they mainly focus on individual differences in responding. Whether immigrants, who come from different cultures and who belong to different generations of immigrants, use different response strategies in surveys has not been systematically studied before. Nonetheless, given the vast amount of cross-cultural differences in measurement errors and response styles (for an overview, see Groves, et al., 2004; Sudman & Bradburn, 1974; Van de Vijver & Tanzer, 1997), it is likely that such response strategies play a key role in the response process. For example, second-generation immigrants may have a higher educational attainment and language proficiency than the first-generation immigrants, and such differences between generations may relate to the response strategies that these respondents use when answering to survey questions. To explore such issues we have conducted a mixed method study, of which we report the design and findings in the remainder of this contribution.

5.2 An Integrated Mixed Methods Study

5.2.1 General Approach

The mixed method design allows us to integrate the strengths of statistical modeling and cognitive interviewing. We start with estimating a latent variable model on data from a large representative sample of the four largest minorities in the Netherlands to detect the magnitude of group differences in responding (Study 1). Then, we conduct cognitive interviews in a small purposive sample of interviewees from the same four cultural groups to study the response process in more detail (Study 2). We integrate both data sets by inferring the response style of interviewees in the purposive sample from the statistical model, and comparing the interviewees' explanations for their response behavior in the cognitive interview accordingly.

In the quantitative study, we use a Latent Class Factor Model (Magidson & Vermunt, 2001) – of which the details will be discussed in the next section – to assess whether minorities respond differently to a selection of survey-items from the large-scale survey given their attitudes and whether these response differences can be attributed to a differential response style (Kankaras & Moors, 2009; Moors & Wennekers, 2003; Morren, Gelissen, & Vermunt, in press b). In the second stage of the study, in cognitive interviews we present the same items that are analyzed in the statistical model to a purposively selected sample of members of these cultural minorities in similar fashion as a regular survey interview.[26] We probed the interviewees retrospectively about a selection of questions which were shown to elicit inequivalent answers across minorities based on the findings from the quantitative study. We also probed the questions to which the interviewees showed interpretational or other problems when they were presented to them for the first time during the interview. The interviewees' justifications of their answers is analyzed qualitatively and related to the cross-cultural differences found in the quantitative study.

5.2.2 Data collection

In this study, we compare the answers to the SPVA[27] survey collected in 2002 among the four largest minorities in the Netherlands[28], namely Turkish, Moroccan, Surinamese and Antillean people. We subjected 15 attitude statements on a Likert scale, each having five ordered response categories that range from *totally agree* to *totally disagree* to statistical analysis. These items operationally define three attitudes, namely the attitude toward the Netherlands, the endorsement of traditional family values, and the autonomy of children within the family. Given the substantive nature of these constructs, the minorities are expected to be subject to culturally specific sources of measurement error and possibly derive their answers to these attitude statements from systematically differing response processes. Table 5.1 reports descriptive statistics of all items included in the analyses for the four cultural groups:

[26] The respondent is asked to select one answer category that reflects his or her opinion best. The interviewer does not provide further explanation.

[27] SPVA stands for Social Position and Utility Use of Ethnic Minorities. The survey maps the cultural, economic and social life of ethnic minorities in the Netherlands (Dagevos et al., 2003). We thank Data Archiving and Networked Services (DANS) for providing the data files.

[28] Since the data is collected among households, we only include the answers given by the heads of the households to secure independent observations.

Table 5.1.

Mean observed item response per ethnic group (N=3549)

Factors and items	Turks	Moroccans	Surinamese	Antilleans
Factor 1: Attitude toward NL				
Item 1: In the Netherlands immigrants get many opportunities	2.48 (1.06)	2.58 (1.07)	2.74 (1.11)	2.75 (1.15)
Item 2: The Netherlands is hostile to immigrants	3.20 (1.02)	3.53 (0.88)	3.60 (0.88)	3.48 (0.91)
Item 3: In the Netherlands your civil rights as an immigrant are respected	2.97 (0.97)	2.53 (0.91)	2.31 (0.89)	2.40 (0.91)
Item 4: The Netherlands is a hospitable country for immigrants	2.60 (0.90)	2.44 (0.85)	2.48 (0.86)	2.56 (0.84)
Item 5: The Netherlands is tolerant towards foreign cultures	2.17 (0.91)	2.43 (0.87)	2.16 (0.82)	2.31 (0.82)
Factor 2: Autonomy of the children				
Item 6: Children should live at home until marriage	2.30 (1.04)	2.24 (1.12)	3.07 (1.27)	3.41 (1.23)
Item 7: Elderly must be able to move in with their children	2.88 (1.13)	2.21 (0.96)	2.90 (1.15)	2.99 (1.17)
Item 8: Adult children should be able to move in with their parents	2.11 (0.88)	2.06 (0.85)	2.68 (1.08)	2.87 (1.11)
Item 9: Parents always have to be respected, even if they do not deserve it based on their behavior or attitude	1.89 (0.83)	1.80 (0.89)	2.39 (1.11)	2.30 (1.09)
Item 10: Older family members should have more influence in important decisions (for instance about moving) than younger ones	2.89 (1.15)	2.63 (1.12)	3.15 (1.12)	3.12 (1.09)
Factor 3: Family values				
Item 11: A man and woman are allowed to live together without being married	3.46 (1.25)	3.88 (1.11)	2.10 (1.04)	2.05 (1.04)

Item 12: Married people with children should not be allowed to divorce	2.88 (1.19)	3.33 (1.14)	3.36 (1.09)	3.59 (1.05)
Item 13: The best family remains to be: Two married parents with children	2.46 (1.09)	1.94 (0.93)	2.53 (1.22)	2.60 (1.21)
Item 14: A daughter aged 17 is allowed to live by herself	4.00 (0.93)	4.06 (0.95)	3.59 (1.04)	3.40 (1.14)
Item 15: The opinion of the parents has to be important in the choice of a partner for their child	2.54 (1.04)	2.52 (1.09)	3.36 (1.11)	3.53 (1.07)
N	905	854	1014	776
Response Rate (%)	52	52	44	51

Note. The items 2, 12, 13 and 15 are formulated in reversed manner where a positive answer indicates a conservative attitude. For the other items a positive answer indicates a modern attitude (see Dagevos, et al., 2003). The standard deviations are in parentheses. Totally agree is indicated by 1, agree by 2, neither agree nor disagree by 3, disagree by 4, totally disagree by 5.

In addition, we held cognitive interviews from December 2009 until October 2010 among 24 interviewees who were recruited via unrelated contact persons, organizations, and personal contacts to avoid overlap in social background. We aimed for a heterogeneous sample as previous studies showed that the style of responding was systematically related to multiple socio-economic characteristics. We interviewed 7 ethnic Moroccan, 7 ethnic Turkish, 5 ethnic Antillean, and 5 ethnic Surinamese interviewees. Among them, 12 belong to the second generation of immigrants, 13 are highly educated, 10 are female, and 8 are older than 40 and 4 are younger than 26. The interviews were conducted at the interviewee's home (14), at Tilburg University (5), at a public place (3) or at the workplace (2). The interviews lasted 45 minutes on average (see Appendix F for interview protocol). During 8 interviews a research assistant was present who also helped with transcribing and coding. Five interviews were excluded from further analyses because 2 were used as a pilot, 2 were conducted among family of an interviewee, and 1 was conducted in another language using a translator. In general, some of the lower educated interviewees[29] displayed difficulties in understanding the purpose of the cognitive interviews. Presumably, a straightforward think-aloud interview would have been more suitable but we persisted with the two-phase interview to maintain comparability with the other interviews.

5.3 Study 1: Measurement Inequivalence and Response Style

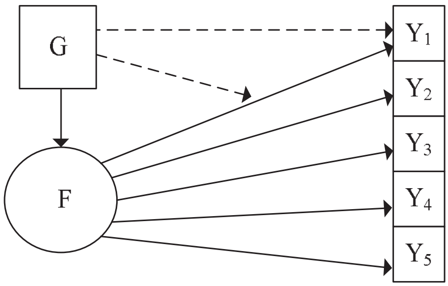
5.3.1 Latent Variable Approach

Within the Latent Variable Modeling framework, attitudes are defined as complex theoretical constructs for which multiple empirical indicators that reflect important aspects of the attitude are needed to obtain a valid and reliable measurement of the construct (Bollen, 2002; Skondral & Rabe-Hesketh, 2004). In this approach, an attitude is modeled as a latent – unobserved – variable (also called factor or trait) for which survey questions (hereafter called items) are used as the indicators or observed variables. One important assumption in cross-cultural research is measurement equivalence (or measurement invariance): Each respondent with a certain attitude level will respond similarly to the items irrespective of his or her group membership (Meredith, 1993). Note that measurement equivalence allows for group differences in attitudes but not for group differences in the indicators conditional on the

[29] Interviewees who only finished primary education, the lower level of high school, or the lower level of professional education.

attitudes. There is evidence for measurement inequivalence when particular model parameters significantly differ across groups (Vandenberg & Lance, 2000).

Figure 5.1. Measurement Inequivalence in a 1-Factor Model



Note. Measurement inequivalence in a 1-factor model containing 5 items. The group variable G relates to the factor (indicating different group means), to the item directly (indicating different group intercepts), and to the item indirectly via the relationship with the factor (indicating different group factor laodings).

Figure 5.1 depicts a latent variable based measurement model in which there is inequivalence with respect to the first item. As can be seen, latent variable F is related to items Y1 to Y5. The group variable G is related to F indicating a group difference in the attitude, but also directly related to the first item indicating group differences in the intercepts. Finally, G also moderates the association between F and the first item indicating group differences in the factor loadings. Note that usually measurement inequivalence occurs across several items simultaneously. The models that we test in this chapter are more complex than the model depicted in Figure 5.1. They contain three related attitudes (three latent factors) measured by five items each (see Appendix E). The unrestricted models allow for group differences in both the intercepts and the factor loadings in each of the fifteen items. The effect of attitudes on item responses is distinguished from the effect of a response style factor (RPS) by estimating models in which only each of the three item subset is affected by one attitude and in which all items are affected by a latent response style factor (Billiet & McClendon, 2000; Cheung & Rensvold, 2000; Moors, 2003; Morren, et al., in press a).

To test for cross-cultural differences in responding, we compare unrestricted models – that allow for all possible group differences – with models assuming model certain parameters to be equivalent across minorities. Table 5.2 reports the log-likelihood and BIC values for the most relevant models. The BIC values can be used to compare models with one another: The lower the BIC value the better the model is in terms of fit and parsimony. Note that the models without and with the RPS factor are nested.[30]

Table 5.2.

Model selection estimated with SPVA data (N=3549)

Model	Fit Statistics Log-Likelihood	 BIC (based on LL)	 Number of parameters
Without a correction			
A) Unrestricted model	-66075.41	134823.86	327
B) Equivalent factor loadings	-66345.42	134996.02	282
C) Equivalent intercepts	-67499.45	135832.70	102
Corrected for response style			
A_{RPS}) Unrestricted model	-61464.97	126134.32	392
B_{RPS}) Equivalent factor loadings	-61588.42	126013.36	347
C_{RPS}) Equivalent intercepts	-62594.19	126553.51	167

The BIC values in Table 5.2 show that there is evidence that ethnic minorities interpret the items differently after controlling for attitude differences. As indicated by the fact that Model A has a lower BIC value than Models B and C, these differences are visible both the item intercepts and the factor loadings. Moreover, inclusion of the style factor improves the model fit considerably; the models with a style factor have always a lower BIC value than their counterparts without a style factor. We also find that the style factor accounts for the group differences in the factor loadings; that is, after controlling for the RPS factor, the model with equal factor loadings (Model B_{RPS}) is preferred over the model with unequal factor loading (Model A_{RPS}). In summary, we find that there is measurement inequivalence between Dutch ethnic minorities and that inequivalence can partly be attributed to the response style factor.

For the purpose of the current study, we are mainly interested in the findings pertaining to the response style factor. Morren et al. (in press a) provide more details on how to interpret this response style factor. To avoid making too strong assumptions about the

[30] The former can be obtained from the latter either by fixing the model parameters for the RPS factor to 0 or by reducing the number of categories of the RPS factor to 1.

distribution of respondents along the underlying RPS dimension and to simplify its interpretation, we defined the RPS factor to be a discrete latent variable with three ordered categories[31] (Magidson & Vermunt, 2001; Vermunt & Magidson, 2005a). Based on a post-hoc interpretation of the parameters, it can be concluded that the first category (latent class) captures the tendency to prefer extreme categories (i.e. *totally agree* and *totally disagree*), whereas the third category captures the tendency to avoid extreme categories and select the adjacent categories (i.e. *agree* and *disagree*), given the attitudes. In other words, respondents who prefer extreme categories have an Extreme Response Style (ERS) and are likely to belong to the first category, whereas respondents who prefer adjacent categories have an Avoidant Response Style (AvRS) and are likely to belong to the third category. Respondents with no particular response style are likely to belong to the second category that measures a position in between the other two latent categories.

5.3.2 Effects of covariates

Cross-cultural differences in the preference for extreme response categories have been related to language use (Bachman & O'Malley, 1984; Gibbons, et al., 1999), a collectivist-individualist values dimension (Harzing, 2006; Johnson, et al., 2005) and acculturation (Van Hemert, et al., 2001; Marin, et al., 1992). In this chapter, we focus on acculturation which is often overlooked but nevertheless an important topic in cross-cultural measurement: The acculturation process –the settlement of immigrants into the receiving society– may influence the way in which people interpret survey questions thereby leading to measurement inequivalence (Van de Vijver & Phalet, 2004). Cultural minorities may either accommodate to the values of the culture corresponding to the language in which the survey is conducted (Harzing & Maznevski, 2002; Oyserman, Sakamoto, & Lauffer, 1998; Ralston et al., 1995), or affirm their ethnic background (Bond & Yang, 1982; Marin, Triandis, Betancourt, & Kashima, 1983). To statistically investigate whether acculturation and ethnicity are related to ERS and AvRS, we included 'generation of immigration and 'ethnicity' as covariates in our model.[32] Similar to the group variable G in Figure 1, these covariates are assumed to affect the response style measured by the latent variable RPS. The results are presented in Table 3.

[31] The three latent classes can be regarded as three ordered categories because they are scored -1, 0 and 1.

[32] Admittedly, this is a rather coarse indicator for acculturation, but unfortunately more sophisticated measures of acculturation were not available in SPVA data set.

Table 5.3.

Effects (logit coefficients) of ethnicity and generation of immigration on the response style in Model C_{RPS}

		Response style factor					
Model	Covariates	AvRS		No response style		ERS	
B_{RPS}	Ethnicity						
	Turkish	0.02	(0.07)	–0.25**	(0.05)	0.23**	(0.07)
	Moroccan	–0.03	(0.08)	0.00	(0.05)	0.03	(0.10)
	Surinamese	0.21**	(0.06)	0.13*	(0.05)	–0.34**	(0.08)
	Antillean	–0.19*	(0.08)	–0.11*	(0.05)	0.08	(0.09)
B_{RPS} +	Ethnicity						
immigration	Turkish	0.03	(0.06)	–0.24**	(0.05)	0.21**	(0.07)
	Moroccan	–0.08	(0.07)	–0.01	(0.05)	0.09	(0.09)
	Surinamese	0.21**	(0.06)	0.13*	(0.05)	–0.33**	(0.08)
	Antillean	–0.16*	(0.08)	0.12*	(0.05)	0.04	(0.09)
	Generation of immigration						
	First	0.05	(0.04)	-0.17**	(0.04)	0.12	(0.06)
	Second	–0.05	(0.04)	0.17**	(0.04)	–0.12	(0.06)

Note. Standard errors are shown in parentheses. According to the Log-Likelihood Ratio Test, including the variable generation of immigration improves the model fit of model B_{RPS} (ΔLL=7; Δdf =2, p = .029). * p < .05. ** p < .01.

Table 5.3 shows the model parameters related to each category of RPS[33] in Model B_{RPS} with and without the covariate 'generation of immigration'. In both models, we hold constant for differences in ethnic background. The parameters are logit coefficients subject to effect coding, which implies that they sum to 0 across latent classes and covariate categories. A negative (positive) value indicates that a certain combination is less (more) likely to occur than average. According to Model B_{RPS} Surinamese respondents are more likely to use AvRS and Antilleans less likely. Turkish respondents are less likely to belong to the category 'no response style', whereas Surinamese and Antillean people are more likely to belong to this category. Finally, Turkish people are more likely to use ERS as a response style, whereas Surinamese people are less likely to use ERS while responding to attitude statements. Controlling for generation of immigration (Model B_{RPS} + immigration) does not alter these

[33] The latent variables are operationalized by three ordinally related categories.

group differences in responding. Holding constant for differences in ethnic background, respondents belonging to the second generation[34] are more likely to use no response style. In summary, these findings indicate that both ethnicity and generation of immigration are related to the (non)usage of a response style.

5.3.3 Classification

Based on the model estimates resulting from the quantitative analysis (Study 1), we assign a response style to the interviewees in the qualitative study (Study 2). We have two reasons to classify the interviewees based on the model estimates of the most parsimonious model C_{RPS}. First, as the number of parameters increases, the classification of the respondents becomes more uncertain. Second, great uncertainty exists about the nature of the purposive sample in the qualitative study, which is why a relatively simple and parsimonious model may be more appropriate.

Table 5.4.

Classification of Interviewees According to Estimates of the Latent Class Factor Model and the Generation of Immigration

	ERS	No response style	AvRS
Generation of immigration			
First	2	2	4
Second	1	9	1

In Table 5.4, the interviewees are classified according to their response style and the generation of immigration. A few interviewees of the first generation (2 out of 8) and the majority interviewees of the second generation (9 out of 11) do not endorse a response style. Note that the quantitative analyses point in a similar direction (see the generational effect on the response style in Table 5.3).

[34] The Netherlands Institute for Social Research (SCP) assigns people, who were born abroad to the first generation, whose parents (at least one) were born abroad or who immigrated before the year of 6 to the second generation.

5.4 Study 2: Investigating Response Strategies in the Response Process

5.4.1 Cognitive Interviews

For Study 2, the recordings of the cognitive interviews were transcribed by the first investigator and a trained research assistant. To analyze the qualitative data from the interviews, a code list was developed which used the theory on survey response as a general thematic framework (see Appendix G). This analysis revealed an important pattern in the data: Interviewees express different aspects of an attitude when responding to the survey questions. First, to justify their answers some interviewees predominantly refer to their personal experiences or refer to abstract notions that apply to many people or situations. We refer to arguments as personal when the interviewee (a) relates his or her opinions to personal experiences, (b) emphasizes that something is only valid to him or herself, or (c) actually discusses personal behavior. An argument is regarded as general – or abstract – when the interviewee (a) talks in general terms, (b) perceives his or her own life in a distant manner or (c) is open to other opinions. Second, in giving personal arguments interviewees differ in the degree to which they relate explicitly to personal behavior: Some repeatedly interpret the questions as a behavioral inquiry while others refrain from revealing any information about personal behavior. Third, we distinguish two ways in which interviewees bring their arguments to the table. An interviewee with a convincing argumentation style seeks to persuade the interviewer of his or her norms and values, arguing in a firm manner without considering the relative value of the statements. Contrastingly, an interviewee endorsing a contemplative argumentation style argues thoughtfully, weighs arguments for and against, and carefully chooses words. Although most interviewees alternately employ both argumentation styles – depending on the type of question – throughout the interview, some interviewees use one style predominantly. Note that each argument is coded separately.

These individual differences in the argumentation allowed us to further distinguish three separate response strategies that differ with respect to general or personal arguments, behavioral information, and a convincing or contemplative argumentation style. Although many combinations are possible, we found that some occur more often others. In particular, we argue that interviewees who systematically exclude personal information in justifying their attitudes follow an attitude-detached response strategy, interviewees who repeatedly answer to questions using behavioral statements follow a behavioral response strategy, and interviewees who alternately use general and personal arguments follow an attitude-balanced response strategy. Note that these specific response strategies differ from response styles: A response style refers to the tendency to select or to avoid certain categories, irrespective of the item content whereas a response strategy refers to the type of arguments presented in justifying these responses.

Table 5.5.

Response Strategies and Use of Arguments and Argumentation Style

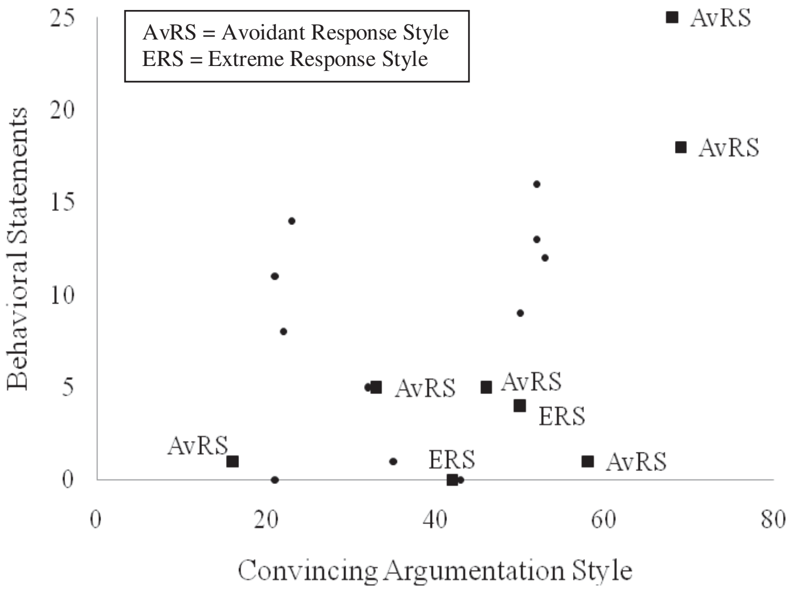
	Arguments			Argumentation Style	
	General	Personal	Behavioral	Convincing	Contemplative
Response Strategy					
Behavioral	Occasional	Frequent	Frequent	Frequent	Rare
Attitude-detached	Frequent	Rare	Rare	Occasional	Occasional
Attitude-balanced	Occasional	Occasional	Occasional	Occasional	Occasional

Note. The terms rare, occasional and frequent are assigned to the cells based on the percentages of words spent on these types of arguments. Occasional refers to diverging pattern of number of words used on these arguments across interviewees using this response strategy.

Table 5.5 illustrates the differences between the response strategies. We assign a response strategy to the interviewees in our sample based on the amount of words spent on these types of arguments and argumentation styles. Four interviewees have a behavioral response strategy, five interviewees have an attitude-detached response strategy, and nine interviewees have a balanced response strategy. Note that one interviewee could not be classified as he scored high on presenting general arguments in a convincing style. The interviewees differ in the degree to which they use these response strategies systematically. In the following, we describe the three response strategies and how they are related to the response style.

Behavioral response strategy. Interviewees who consistently answer using information about their own behavior are characterized by a behavioral response strategy. They interpret the attitude questions as if the interviewer asks about their actual personal behavior in certain situations. After the interviewer has asked them to repeat the question in their own words, they say something like: “You want to know if my parents have something to say about whether I move out?” Thinking of personal experiences can have two effects on responding: If the personal experiences support their general opinions about a subject, interviewees are likely to give a clear (possibly more extreme) answer; however, if the personal experiences contradict their opinions, they are likely to give an ambiguous (possibly less extreme) answer (Tourangeau et al., 2000, pp. 185). We find that interviewees who use personal experiences to support their attitudes often use a more convincing manner to present their arguments. An ethnic Turkish male (21) who scores high on behavioral response strategy, agrees to item 9: “I would always treat my parents very well, especially because of how they treated me until now, they raised me and uh, they made me a man”. With respect to the same item, other interviewees integrate personal information while simultaneously regarding alternative situations in which their attitudes might apply. An ethnic Surinamese female (27) – also agreeing to item 9 – argues: “I can always count on them, they do everything for me […] If I wouldn’t like their behavior or attitude, I would say something about it but that doesn’t mean I would respect them any less”. We illustrate these differences in presenting behavioral arguments in a convincing way in Figure 5.2.

Figure 5.2. The Use of Behavioral Statements and Convincing Argumentation Style



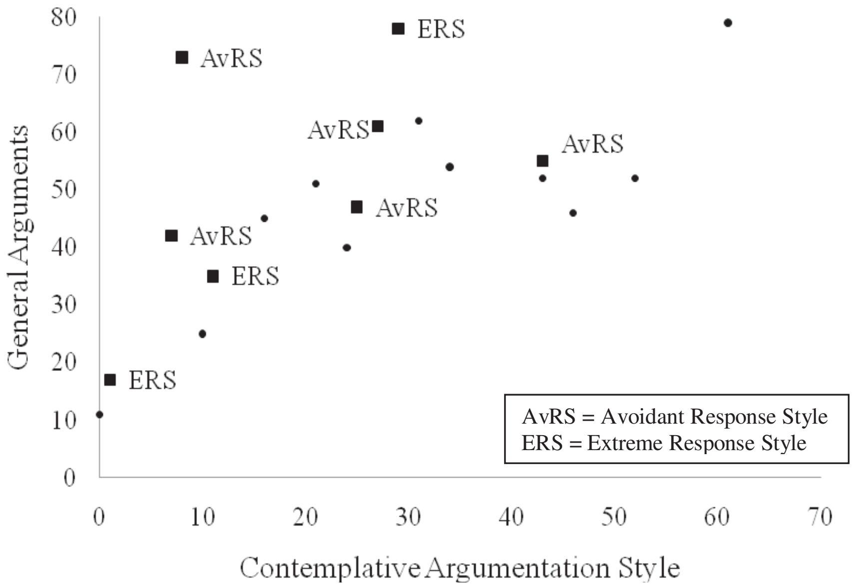
Note. The numbers on the axes represent the percentage of words spent on these arguments. Interviewees plotted on two dimensions related to the codes 'behavioral statements' and 'convincing argumentation style'. The response style of the interviewees is indicated by AvRS, ERS or not mentioned if no response style.

In Figure 5.2 we plot interviewees along two dimensions measuring the words used in arguments that either treat behavior or are posed in a convincing manner. The words are given in percentages and in relation to the total number of words spend on justifying the answers retrospectively. For example two interviewees located in the lower left corner score low in both behavioral statements and a convincing argumentation style. The other two respondents with a behavioral response strategy are somewhat more to the middle but remain in the upper right corner of the figure. Note that the conclusions we made with respect to behavioral arguments also hold for the less specific personal statements.

Remarkably, interviewees employing a response style score either high or low in behavioral statements. Six out of ten interviewees using 5% or less of their words on behavioral statements and both interviewees in the upper right corner systematically select or avoid extreme categories. Thus, most interviewees who use a response style either systematically exclude behavioral statements or intentionally interpret the questions as behavioral questions. We suggest that excluding personal information makes it easier to systematically translate ideas about surveys, the question topics, and the question format into particular response categories. For example, one interviewee retrospectively argues: “I didn't think of a family situation at first, […] that is the difference I guess. So earlier I answered without considering my own family and when I think about it now, I come to a different conclusion”. He changes his response from *agree* to *totally agree* after including his personal experiences. We also observe the opposite: Personal experiences that intensify attitude conflicts may lead to less extreme judgments. For instance, a single woman argues in agreeing to item 13: “To say ‘totally agree’ would imply that I do not approve of my own family situation”.

Attitude-detached response strategy. Interviewees, who shun personal information by avoiding their ethnicity or personal experiences, use a detached response strategy. They reveal only part of their attitudes through abstract, general reasoning. At some point, eight interviewees explicitly argue that they do not want to generalize their personal experiences, take their culture as abstract reference point, or give a general view on society. Two of them even declare in advance: “I will interpret the question generally, not personally”. Interviewees who use an attitude-detached response strategy use vague arguments, for example: “That is the traditional image that everybody longs for eventually, I think, to be together but also to have children to live for”. Some mention their own norms and values in a somewhat distant manner: “You should be there for your child, no matter what” or “That is the habit in our culture, yes, here in the Netherlands it is different”. Some of these interviewees explain why they use general arguments: “My personal opinion does not mean that everyone should have the same opinion”. To show the individual differences in the overlap between using general arguments and presenting them in a contemplative manner we plot the interviewees in Figure 5.3 with respect to which they use general arguments in a contemplative manner.

Figure 5.3. The Use of General Arguments and Contemplative Argumentation Style



Note. Interviewees plotted on two dimensions related to the codes 'general arguments' and 'contemplative argumentation style'. The response style of the interviewees is indicated by AvRS, ERS or not mentioned if no response style. The numbers on the axes represent the percentage of words spent on these arguments.

The top of Figure 5.3 depicts three interviewees who spend more than 70% of their words on making general arguments. The majority of the interviewees are plotted in the center of Figure 5.3 meaning that they use about 50% of their words on general arguments. Three interviewees using a detached response strategy are among these centered interviewees and the other two are located in the upper middle and upper right corner, the respondents using a behavioral response strategy are located in the lower left corner of Figure 5.3.

Attitude-balanced response strategy. Interviewees with an attitude-balanced response strategy form an opinion by integrating thoughts and experiences across several domains: Personal surroundings, the host country and their cultural background. First, we suggest that interviewees with an attitude-balanced response strategy focus on specific characteristics of the questions instead of responding similarly to all attitude questions. As they integrate several sources of information dependent on the topic of question, they react more balanced and are less likely to be subject to a response style than others. Only two out of nine interviewees using this response strategy are subject to a response style. These 'balanced' interviewees seem to choose for either personal versus general reflections or a convincing versus a contemplative argumentation style depending on the question content.

Response strategies, ethnicity and generation of immigration. After identifying the different response strategies, we examine whether the interviewees' ethnicity and their generation of immigration are systematically related to the response strategies that they use, as these characteristics were also used in the quantitative analysis. The evidence from the qualitative analysis suggests that particularly the generation of immigration is related to the response strategy used by the interviewees. Specifically, we find that the interviewees who are less acculturated than others predominantly use personal arguments or mainly present arguments in a convincing way. Table 5.6 reports the percentage words used during the interview in giving statements using the behavioral response strategy, averaged across the interviewees of the first generation (10), and across the interviewees of the second generation (9) of immigration.

Table 5.6.

Amount of words used in justifying answers, given in percentages

	Argumentation style		Arguments					
	All items		Items 1 to 5			Items 6 to 15		
	Convincing	Contemplative	Personal arguments	Behavior	Ethnic background	Personal arguments	Behavior	Ethnic background
Generation of immigration								
First	49	18	10	1	5	11	5	5
Second	33	35	5	0	2	17	5	0

Note. The percentages are obtained by dividing the words spent on these types of arguments by the total number of words spent on the retrospective justification of the answers.

Table 5.6 indicates differences between the first and second generation interviewees in three aspects. First, first generation' interviewees more often use a convincing argumentation style than a contemplative argumentation style. In contrast, second generation' interviewees use the contemplative and convincing argumentation style to similar degree. Second, interviewees belonging to the first generation rather use more personal than general arguments when answering questions about the attitude toward the Netherlands, whereas the interviewees belonging to the second generation answer personally to questions about family matters. Third, whereas interviewees belonging to the first generation use information related to both personal experience and the ethnic background when explaining their response, the second generation' interviewees avoid the ethnic background. This finding could indicate that the interviewees from the first generation more likely affirm their own ethnic background (Marin, et al., 1983), while interviewees of the second generation more likely accommodate their answers to the cultural majority (Ralston, et al., 1995). Table 5.6 also illustrates that the ethnic background is rarely referred to by the interviewees in a direct way which could also be related to the diffuse distinction between personal and cultural experiences (Smith, 1998). With respect to ethnicity, we find that the Moroccan interviewees of both first and second generation use an attitude-balanced response strategy. This finding is in accordance with Phalet and Schönpflug (2001) who attribute a more acculturated lifestyle to Moroccan than to Turkish respondents. Similarly, Stevens, Pels, Vollenbergh and Crijnen (2004) find that the majority of Moroccan respondents score high on identification with the Dutch as well as the Moroccan culture. This acculturation style is referred to by Berry (1990) as integration: Those who feel close to the values of the host society as well as their ethnic background. Presumably, ethnicity mainly plays a role in the response process through the mode of acculturation that differs across minorities.

5.5 Conclusion and discussion

In this chapter, we suggest that cross-cultural differences in responding are related to the extent to which respondents integrate their experiences of their personal, cultural and public life in the survey process. Similar to Krosnick (1991), we argue that respondents employ response strategies to deal with the difficulty of answering to an attitude statement. Especially for minority respondents, an attitude statement presents several decisions to be made: Should they focus on their personal situation, their cultural background, or the norms and values of the majority? Ideally, respondents have beliefs, impressions and prior judgments in each situation; they attribute importance to these beliefs accordingly and base their judgment on a balanced summary of these 'weighted' beliefs. Cross-cultural differences in responding may result if the respondents avoid the complexity of making a balanced judgment by focusing primarily on their personal situation, their cultural background or the degree to which they relate with the host culture. If their answers depend on their personal situation or ethnic background, cultural specific traditions, habits, or topics that are considered taboo become more important. These differences in perception of the topics referred to by the questions can ultimately lead to measurement inequivalence.

We have shown statistically that part of the response differences across the minorities in the Netherlands can be accounted for by the response style and the generation of immigration. Next, we have related these quantitative results to a qualitative sample of interviewees belonging to the same minorities. We assigned a response style to these interviewees based on the model estimates obtained using the large sample. We have questioned these interviewees about their answers, and related their way of justifying their answers to their response styles and generation of immigration. Our findings suggest that interviewees who exclude personal information or purposively relate all questions to their own behavior are more likely to use a response style than the other interviewees. We find that all ethnic Moroccan interviewees use a balanced response strategy, and that the ethnic Moroccan respondents in the quantitative sample are less likely to use any response style than the other minorities. Finally, we find that interviewees of the first generation are more likely to justify their answers using their own personal experiences, and are more likely to present these arguments in a convincing manner. Note that our cognitive interviews were based on a relatively small purposive sample of interviewees with a limited number of persons per ethnic group which makes the detection of patterns of response strategies more challenging.

We inferred three response strategies from our findings: The behavioral response strategy, the attitude-detached response strategy and the attitude-balanced response strategy. The great variation in the way in which respondents justify their answers illustrates that the understanding of the survey questions differ strongly across respondents. To improve especially cross-cultural survey designs, we suggest including a short introduction in which the researcher clarifies to the respondents whether the domain of interest is the host country, the cultural background, or the personal experiences. In this way, respondents who come from different cultural backgrounds may better understand what the researcher wishes to know. As a result, his or her responses may more validly reflect the construct that the researcher intends to measure and problems of measurement inequivalence may be reduced.

Appendix E. The Latent Class Factor Model

Here we provide more details about the Latent Class Factor Model with a response style we used in our analysis. This model was proposed by Moors (2003). Recently, Morren et al. (in press a) extended this model by showing that it is better to treat the relationship between substantive factors and items differently from the relationship between response style factor and items. More specifically, in their relationship with the response style factor, the item responses are treated as nominal variables, yielding five category-specific parameters per item. This means that no assumptions are made about the form of these relationships. For the attitude factors, only one parameter is used per item because for this relationship the items are treated as ordinal variables. More specifically, we assume that

$$P(Y_{ij} = c \mid F_{1i}, F_{2i}, F_{3i}, R_i) = \frac{\exp(\beta_{0jc} + \beta_{1j} c F_{1i} + \beta_{2j} c F_{2i} + \beta_{3j} c F_i + \beta_{4jc} R_i)}{\sum_{d=1}^{C} \exp(\beta_{0jd} + \beta_{1j} d F_{1i} + \beta_{2j} d F_{2i} + \beta_{3j} d F_i + \beta_{4jd} R_i)} \ . \qquad [14]$$

This is a hybrid between a multinomial and an ordinal logit model. The β's are the item parameters to be estimated: β_{0jc} is an intercept term for item j and category c, β_{1j}, β_{2j} and β_{3j} are slope parameters corresponding to the three substantive factors, and β_{4jc} are the slope parameters for the response style factor denoted by R_i . The parameters β_{1j}, β_{2j}, and β_{3j} are multiplied by the category number c, which results from the ordinal specification for the relationships with the substantive factors. Note that some of these parameters are fixed to 0 because each item loads on only one substantive factor. The other model parameters are category specific.

The Latent Class Factor Model with a response style can be estimated with the Latent GOLD software, a general package for latent variable modeling (Vermunt & Magidson, 2008). This program also provides estimates for $P(\text{RPS}_i|\text{Y})$, that is, the probability of having a particular response style given the provided responses. This feature is used in our Study 2 in which we determined the response style for each of the interviewees.

Conclusion

Researchers who use survey data to answer research questions on attitude differences between cultural groups too easily assume that empirical measurements of attitudes are naturally comparable, without considering the impact of a violation of this assumption on their conclusions. The research that is reported in this thesis contributes to the continuing development of quantitative and qualitative methods – and their possible combination – for investigating the cross-cultural comparability of attitude measurements. This thesis presents an innovative way of using interviews to determine how the response process differs between respondents with different response styles and cultural backgrounds.

The empirical findings that are reported in this thesis indeed emphasize the necessity of checking the comparability of attitude measurements. In particular, cross-cultural inequivalences in responding clearly occurred in a Dutch large-scale survey measuring attitudes of the four largest immigrant groups. When investigated in more detail using appropriate statistical methods, it was found these inequivalences distort conclusions about the actual group differences in attitudes. We also explored two specific causes of this incomparability, namely existence of culturally specific response styles and response strategies. Both response styles and response strategies relate to the way in which people deal with attitude statements; where the former refers to the respondent's tendency to select certain response categories and the latter to the argumentation a respondent uses to choose certain response categories. We investigated these response styles and strategies by an innovative approach to mixed method research in which we integrated statistical modeling with qualitative interviews. First, using three attitude scales from the SPVA survey held among the four largest Dutch minorities we illustrated how cross-cultural response differences can be detected by a latent variable modeling approach that was initially developed by Moors (2003, 2004) and which we extended in various ways. Second, to provide an explanation for the statistical findings, we held cognitive interviews among a small (unrelated to the SPVA respondents) sample of members of the four minorities. In this part of the research, we used the results from the statistical analysis to classify the interviewees of the cognitive interviews with respect to their response style.

The quantitative part of this study contributed to current survey-methodological research by demonstrating how comparability of measurements can be evaluated using an restricted version of Latent Class Factor Analysis (LCFA) approach, how cross-cultural differences in responding can be attributed to a differential response style, and which characteristics of the survey questions posed to respondents may lead them to use a response style and particular response strategies. Methodologically, the most important contributions are the use of a restricted LCFA approach in cross-cultural research, the step-by-step presentation of this approach which makes LCFA more accessible to researchers who are less familiar with latent variable modeling (Chapter 2), the finding that this response style affects the group differences in several parts of the model, and the exploration of the substantive meaning of the response style factor (Chapter 2 and 3).

However, there are also various limitations associated with the quantitative findings presented in this thesis. One of these is that in our quantitative analyses we always estimated and compared a limited number of LCFA models. Model selection was based on the BIC, which means that we selected the model that performs best in terms of fit and parsimony among the sets of investigated models. Whereas according to the BIC there was evidence that the more complex models with an extreme response style factor had to be preferred, we were not able to test the overall fit of estimated models. This is the result of the fact that the frequency table we are analyzing is extremely large (and thus sparse), which means that goodness-of-fit measures can no longer be applied. The consequence of not being able to check the overall fit of the estimated model is that important aspects of the data may have remained undetected. For example, it may be that in addition to an ERS factor, the responses are affected by other style factors such as acquiescence. Another example is that the effect of the ERS factor may differ across cultural groups, which is something we did not investigate. In other words, while the encountered extreme response style factor was in agreement with our expectations, easy to interpret, and validated using an ad hoc ERS scale, reality is probably slightly more complex than we showed in our quantitative analyses.

In the qualitative part of the study, we explored possible reasons why some respondents use a response style and others do not, and discovered that respondents use, when asked, different arguments to justify their answers to attitude statements. This latter finding is conceptualized as the usage of response strategies. We found that minority respondents employ different response strategies to respond to attitude statements and that this appears to be related to the generation of immigration to which the respondents belongs. Especially the respondents who immigrated to the host country themselves are susceptible to answering to attitude statements in a behavioral manner based on their personal situation and the customs shared with their ethnic background. Ethnic minorities who belong to the second generation are more likely to make a more balanced argument that includes their experiences with the norms and values of the host society. Although some differences in response strategies across minorities are found (see Chapter 5), most findings were related to more general characteristics of the interviewees such as the generation of immigration. This finding could be - the results of two factors: First, the qualitative findings were gathered using a relatively small, purposive sample and only a limited number of persons per group is included in the analyses, which makes it less likely that systematic cultural differences between minorities come to the fore. Second, the presence of a native interviewer may have impacted on some interviewees to be less candid about their cultural background as they would have been with an interviewer of comparable background.

We have shown that the use of cognitive interviews has potential for revealing underlying mechanisms which may lead to a better understanding of response style usage. In addition, we have found that cognitive interviews are an adequate tool for uncovering different strategies that respondents use to answer to attitude questions and for suggesting which characteristics of attitude questions are important in these processes. However, cognitive interviews also have limitations, as we already pointed out in the introductory chapter. One of the main limitations is the need for interviewees who are capable of perceiving their own response process, expressing these thoughts into words, and reporting these thoughts in an unbiased and systematic manner. Not every interviewee has such capabilities. Consequentially, the reliability of the data depends on the quality of the interviewees. Even though most of our interviewees were very willing to present their thoughts to us, we do recognize this particular limitation of this interviewing method. Similar to regular qualitative interviews, we approached this data as a reflection of the way people present themselves and not necessarily as a direct observation of their cognitive processes. Thus, the verbal reports obtained by the cognitive interviews mainly show us how people justify their answers and reveal how people filter their answers, but they only rarely give us an unbiased account of inner thought processes.

This having said, we believe the cognitive interviews were a valuable asset in combination to the quantitative methods for several reasons. First, the qualitative findings suggested a plausible explanation for the research findings obtained by the statistical modeling. Second, assigning response styles to the interviewees based on the response profiles allowed us to integrate the quantitative findings to the qualitative findings. Unfortunately, privacy regulations prohibited us to use the same respondents in both parts of the research and we were forced to use different samples. Combining quantitative and qualitative methods made us more sensitive to the unique advantages of each method: Statistical modeling allowed us to investigate the generality of the findings while the cognitive interviews permitted us to delve into why these findings occurred in the first place. The mixed method approach also gave us more insight to attribute a meaning to the response style factor. In particular, at first we claimed that the factor measures extreme response style (see Chapter 2 and 3), i.e. the tendency to select extreme answer-categories, irrespective of the content of a survey-item. However, upon closer examination we found that the factor detects two response tendencies, a preference for extreme categories and a tendency to avoid extreme categories by preferring adjacent categories. In Chapter 4 and 5, we described why these two tendencies should be considered separately and referred to them as two distinct response styles: Extreme response style (ERS) and avoidant response style (AvRS).

To conclude, the findings induced us and hopefully other cross-cultural researchers to look into a new direction. Besides paying attention to the detection of and correction for response differences across groups, we emphasize the need for exploring ways in which attitudes can be measured. First, we suggest paying extra care to the formulation of questions, especially when they are meant for respondents from culturally diverse backgrounds because they might have a culturally specific interpretation of the question topics. Based on our findings that ethnic respondents have different frameworks on which they can base their answers, we advice researchers to include an introduction to the survey questions that indicates whether the researcher is interested in the personal behavior of the respondent, his or her ethnic background, or the extent to which the respondent approves of the norms and values of the host society. Second, we stress that although the influence of ERS and AvRS might be less visible when assuming linearity in factor models, the possibility of the presence of these response styles should persuade researchers to think very carefully about the measures they use. If particularly the use of Likert scales imports such distortions in the data, researchers should pay attention to how they could improve these measures to avoid response styles. Third, we suggest exploring the relationship between measurement inequivalence and the style of acculturation further. In our view, this thesis presents only a start in considering the culturally specific characteristics of the response process.

Appendix F. Interviewer protocol

Cognitieve interviews	
atum: Tijd:	Naam: Telefoon:

Introductie

Dit interview is deel van een groter onderzoek dat wordt uitgevoerd door de Universiteit van Tilburg. Wij zijn geïnteresseerd in hoe mensen vragen beantwoorden en of de culturele achtergrond hier een rol in speelt. De vragen die ik u ga stellen zijn letterlijk overgenomen uit een landelijke vragenlijst (SPVA). Uit voorgaand onderzoek bleek dat deze vragen soms problemen opleveren door de zinsopbouw of de onderwerpen. We hebben vragen geselecteerd over de beeldvorming van Nederland, familiewaarden en over de autonomie van kinderen in de familie. Als u problemen hebt om over een van deze onderwerpen te praten, kunt u dat nu aangeven. Uw privacy wordt gegarandeerd, uw gegevens worden vertrouwelijk behandeld en komen niet terug in de analyse. Heeft u er problemen mee als het interview wordt opgenomen? Het interview zal alles bij elkaar zo’n drie kwartier in beslag nemen. Heeft u nog vragen?

A: retrospective CI

INTERVIEWER Deze vragen worden letterlijk voorgelegd aan de respondent waarin deze wordt gevraagd te kiezen uit de vijf antwoordcategorieën. Geen verdere toelichting geven over de vragen, zelfs niet als de respondent daarom vraagt. Als de respondent moeite heeft met de antwoordcategorieën, mag de interviewer de respondent KAART A geven.

SPVA Vragen		Antwoordcategorieën				
1: Autonomie kind		helemaal mee eens	mee eens	mee eens/ niet mee eens	niet mee eens	helemaal niet mee eens
a	Kinderen kunnen het beste thuis blijven wonen tot zij gaan trouwen	1	2	3	4	5
b	Als ouders bejaard zijn, moeten ze bij hun kinderen kunnen inwonen	1	2	3	4	5
c	Ouders zouden hun volwassen kinderen in huis moeten nemen als die daar om vragen	1	2	3	4	5
d	Je moet je ouders altijd respecteren, ook wanneer ze dit door hun houding of gedrag niet verdienen	1	2	3	4	5
e	Bij belangrijke beslissingen (bijvoorbeeld over verhuizen) horen oudere familieleden meer invloed te hebben dan jongere	1	2	3	4	5

2: Beeldvorming		helemaal mee eens	mee eens	mee eens/ niet mee eens	niet mee eens	helemaal niet mee eens
a	In Nederland krijg je als buitenlander alle kansen	1	2	3	4	5
b	Nederland staat vijandig tegenover buitenlanders	1	2	3	4	5
c	In Nederland worden je rechten als buitenlander gerespecteerd	1	2	3	4	5
d	Nederland is een gastvrij land voor buitenlanders	1	2	3	4	5
e	Nederland staat open voor buitenlandse culturen	1	2	3	4	5

3: Familie waarden		helemaal mee eens	mee eens	mee eens/ niet mee eens	niet mee eens	helemaal niet mee eens
a	Een man en een vrouw mogen ongehuwd samenwonen	1	2	3	4	5
b	Gehuwden met jonge kinderen mogen niet scheiden	1	2	3	4	5
c	De beste gezinsvorm is nog altijd: twee getrouwde ouders met hun kinderen	1	2	3	4	5
d	Een dochter van 17 jaar mag zelfstandig wonen	1	2	3	4	5
e	De mening van de ouders moet een belangrijke rol spelen bij de keuze van een partner voor hun kind	1	2	3	4	5

4	Ik heb liever vrienden over de vloer dan familieleden	1	2	3	4	5

B: Think aloud
Introductie Bij de volgende vragen, willen we weten hoe u tot uw antwoord komt. Ik zou van u willen vragen hardop te denken, en me te vertellen waar u aan denkt bij het kiezen van uw antwoord. Ik wil deze methode eerst even met u oefenen dmv het volgende voorbeelden.
INTERVIEWER In dit gedeelte wordt bij elke vraag de ruimte gelaten aan de respondent om zoveel mogelijk te vertellen wat er in zijn/haar hoofd omgaat bij het beantwoorden. Als de respondent echt niet weet wat hij/zij kan vertellen, dan kun je helpen door te proben met de probes in Appendix A. Het is niet de bedoeling de respondent teveel te proben. Selecteer drie vragen waarmee respondent moeite heeft en daar extreem, niet-extreem en midden heeft geantwoord, maar niet de 'onderstreepte' vragen zoals bij (a).

Voorbeeldvragen	
1	Wat heeft u voor het laatst gegeten als avondmaal?
2	Hoeveel deuren heeft uw huis?

INTERVIEWER Bij a) worden vragen herhaald die al gesteld zijn en bij b) wordt een nieuwe vraag gesteld. Selecteer bij c) (zie hieronder) uit gedeelte A een drietal vragen waarbij de respondent extreem, niet-extreem en in het midden heeft geantwoord (verschillende item subsets?).

Checklist onderwerpen CI		Antwoord	Begrip	Aannames	Inform atie
a	Nederland is een gastvrij land voor buitenlanders [E]				
	Nederland staat vijandig tegenover buitenlanders				
	De mening van de ouders moet een belangrijke rol spelen bij de keuze van een partner voor hun kind				
	Als ouders bejaard zijn, moeten ze bij hun kinderen kunnen inwonen				
	De beste gezinsvorm is nog altijd: twee getrouwde ouders met hun kinderen				
	Je moet je ouders altijd respecteren, ook wanneer ze dit door hun houding of gedrag niet verdienen				
b	Ik heb liever vrienden over de vloer dan familieleden				
c	[interviewer: hier nummer vraag invullen]				
	[interviewer: hier nummer vraag invullen]				
	[interviewer: hier nummer vraag invullen]				

C: Extreme Response Style (gebruik KAART B)	
INTERVIEWER Kies 3 tot 5 vragen uit waarmee de respondent wel of juist helemaal geen moeite had in B (+ de stelling: 'Een man en een vrouw mogen ongehuwd samenwonen') [E].	
a	Kunt u de categorieën makkelijk/moeilijk interpreteren?
b	Wat betekent de middencategorie voor u?
c	Als de schaal meer opties zou bevatten, zou u dan nog steeds de [extreme/ niet-extreme/ midden] categorie

	kiezen? [KAART B LATEN ZIEN]	
D: Vignettes		
Introductie In het volgende gedeelte, leggen we u enkele vragen over fictieve personen. Zoals bij de voorgaande vragen, vertel precies de redeneringen die u volgt bij het kiezen van een antwoord uit de antwoordcategorieën.		
Instructie interviewer De vignetten worden een voor een door de interviewer opgelezen en de respondent kan meelezen van de kaart. Na elke vignet wordt de vraag gesteld door de interviewer waarbij de naam en geslacht wordt ingevuld nav het geslacht en etniciteit van de persoon (zie Appendix B)		
Self-assessment vraag (zie eerste gedeelte)		
Vignette vraag *Zie instructie INTERVIEWER*		
In hoeverre bent u het er mee eens/oneens dat [...] samenwoont met [haar/zijn] partner?		Kaart B
a	[...] woont ongehuwd samen. [Haar/Zijn] partner had even geen plek om te wonen en dus was [hij/zij] bij [hem/haar] komen wonen.	
b	[...] woont ongehuwd samen met [haar/zijn] partner. Vlak nadat ze elkaar hadden ontmoet, zijn ze gaan samenwonen omdat ze bij elkaar wilden zijn.	
c	[…] woont ongehuwd samen. [haar/zijn] partner had even geen plek om te wonen en […] had nog een slaapkamer over.	
d	[…] woont ongehuwd samen. Na een serieuze relatie van een aantal jaar zijn ze gaan samenwonen ondanks dat hun ouders er niet mee eens zijn.	
e	[…] woont ongehuwd samen. Na een serieuze relatie van een aantal jaar hebben ze na lang praten besloten om te gaan samenwonen met instemming van beide ouders.	
f	[...] woont ongehuwd samen. Ze hebben zich onlangs verloofd in het bijzijn van familie en vrienden.	
g	[…] woont ongehuwd samen. Ze zijn onlangs getrouwd voor de kerk/imam maar het huwelijk is nog niet officieel voor de Nederlandse wet.	
h	[...] woont ongehuwd samen. Ze gaan volgende week trouwen en zijn alvast verhuisd zodat ze na het feest meteen op vakantie kunnen.	

E: Vragen over beleving survey proces	
Wat vond u van de vragen die aan bod kwamen?	
Wat vindt u van het face-to-face interview?	
F: etnische achtergrond informatie	
In dit gesprek hebben we het gehad over hoe u zelf denkt over verschillende onderwerpen. We willen graag ook weten hoe er volgens u in uw gemeenschap gedacht wordt over de onderwerpen waarover wij hebben gesproken.	
De beeldvorming van Nederland? Familie waarden? Autonomie van kinderen in de familie/opvoeding?	
G: acculturation [kaart C, D, E]	
Welke taal/talen spreekt u over het algemeen? **[kaart C]**	
Welke taal/talen sprak u als kind thuis? **[kaart C]**	
Welke taal/talen spreekt u thuis? **[kaart C]**	

Welke taal/talen spreekt u met uw vrienden? **[kaart C]**
Welke taal/talen wordt gebruikt in de TV programma's die u meestal kijkt? **[kaart C]**
In uw vrije tijd, gaat u vooral om met: **[kaart D]**
Als u naar muziek luistert, hoe vaak is dit T/M/S/A muziek? **[kaart E]**
Als u belangrijke gebeurtenissen viert, hoe vaak zijn dit T/M/S/A tradities? **[kaart E]**
Hoe vaak eet u T/M/S/A eten? **[kaart E]**
Met welke groep(en) van mensen voelt dat u veel van uw denkbeelden en waarden mee deelt? **[kaart D]**
Naar uw mening, welke groep(en) begrijpen uw ideeën (uw manier van nadenken) het beste? **[kaart D]**
Bij welke groep(en) van mensen voelt u zichzelf het meeste thuis? **[kaart D]**
Welke cultuur bent u trots om deel van te zijn? **[kaart F]**
Mijn geloof is een belangrijk deel van mijzelf **[kaart A]**

H: Persoonlijke informatie				
Leeftijd				
Opleidingsniveau				
Tot welke bevolkingsgroep rekent u uzelf?	Turkse / Marokkaanse / Surinaamse / Antilliaanse / Nederlandse (meerdere antwoorden mogelijk)			
Generatie (Welk familielid is in buitenland geboren?)	1^e (zelf + 1 ouder)	2^e (1 ouder)	3^e (1 grootouder)	weet niet
I: stel vast hoe goed de respondent Nederlands spreekt.	Goed	Voldoende	Matig	Slecht
J: Respondenten				
Kent u mensen in uw omgeving die mee zouden willen werken aan een interview?				
Afsluiting				
Ik wil u bedanken voor het interview. U zal op de hoogte worden gebracht over de resultaten door middel van een kort rapport. Als u uw interview wil inzien, kan dit natuurlijk. Heeft u nog vragen of opmerkingen?				

Appendix A: Probes	
Begrip/ interpretatie	
a	Kunt u me in uw eigen woorden vertellen waar deze vraag over ging?
b	Wat betekent [*woord*] voor u zoals het is gebruikt in deze vraag?
c	Kunt u me vertellen waar u aan dacht toen ik u vroeg over [*onderwerp*]?
Veronderstellingen	
a	Tot op welke hoogte heeft deze vraag betrekking op u?
b	Kunt u me meer vertellen daarover?
c	Zou u zeggen dat dit meestal hetzelfde is, of dat het ergens vanaf hangt?
d	Kunt u me meer vertellen over uw mening?
Informatie	
a	Waar denkt u aan bij [*onderwerp*]?
b	Hoe vaak denkt u na over [*onderwerp*]?
c	Wanneer heeft u voor het laatst nagedacht over [*onderwerp*]?
d	Hoe makkelijk of moeilijk is het [*onderwerp*] te herinneren?
Antwoordcategorieën	
a	Hoe makkelijk/moeilijk was het om een antwoordcategorie te kiezen?
b	Aan welke situaties denkt u bij helemaal mee eens, en zijn deze anders als bij helemaal niet mee eens?
Appendix B: NAMEN BIJ VIGNETTES	
T	Vrouw [Asli / Fatma / Zaide / Birsu]
	Man [Mehmet / Emir / Umut / Usuf]
M	Vrouw [Samira / Dunya / Yoessra / Deheb / Merjam]
	Man [Mohammed / Rachid / Moenier / Fouad / Jamany]
S/A	Vrouw [Yanella / Chayenne / Kayleigh / Rudesha / Chanesha / Dyonne]
	Man [Harvey / Miley / Cecilio / Lorenzo / Ramsey]

KAART A:

Helemaal mee eens	Mee eens	Mee eens/ niet mee eens	Niet mee eens	Helemaal niet mee eens
1	2	3	4	5

KAART B:

Helemaal mee eens	Mee eens	Beetje mee eens	Mee eens/ niet mee eens	Beetje niet mee eens	Niet mee eens	Helemaal niet mee eens
1	2	3	4	5	6	7

KAART C:

Altijd Turks	Meestal Turks	Beide hetzelfde	Meestal Nederlands	Altijd Nederlands
1	2	3	4	5

KAART D:

Alleen mensen met Turkse achtergrond	Vooral mensen met Turkse achtergrond	Beide achtergronden	Vooral mensen met Nederlandse achtergrond	Alleen mensen met Nederlandse achtergrond
1	2	3	4	5

KAART E:

De hele tijd	De meeste tijd	De helft van de tijd	Soms	Nooit
1	2	3	4	5

KAART F:

Alleen Turkse cultuur	Vooral Turkse cultuur	Beide culturen	Vooral Nederlandse cultuur	Alleen Nederlandse cultuur
1	2	3	4	5

Appendix G. The list of codes.

Three layers of information are distinguished in the data. First, the data is analyzed with respect to the phase of the response process. According to the coding scheme, we assess which part of the respondent's argumentation is related to interpretation, retrieving information, making a judgment, mapping a judgment or editing his or her answer. Second, we code the type of the arguments respondents make during the interview by differentiating between arguments based on general opinions, on personal experiences or feelings, or the conditions people define under which their arguments are deemed to be relevant. Next to these information sources, we code the extent to which their arguments relate to the attributed meaning of actual words used in the question or whether they are related to the formulation of the question or the survey method as a whole. Lastly, we coded the remarks about mapping of the judgments to the response categories. To illustrate these codes to the other researchers, we developed a codebook including a definition, inclusion and exclusion criteria and an example text for each code (not presented here due to limited space).

1. Understanding: Wording

1. Model auxiliary verbs
    a. Have to/ must/ have got to (obligation)
    b. Ought to/ should (recommendation)
    c. Could/ might/ is allowed to (possibility)
2. Adjectives
    a. Married parents
    b. Important role
    c. Young children
    d. Best family
3. Ambiguous words
    a. Hostile
    b. Rights
    c. Respect
    d. Hospitable
    e. The Netherlands
    f. Parents
    g. Foreign cultures
4. Hidden assumptions
    a. Foreigners means immigrants
    b. Compare friends with family members
    c. Seventeen years is immature
    d. Without children not a family
    e. Daughter different than son
    f. Extended or nuclear family

2. Understanding: Syntax

1. Misinterpretation of question
    a. Subordinate clause not considered
    b. Confuses questions
    c. Negative question
    d. Example influences answer
2. Aware of misunderstanding
    a. Request to repeat
    b. Request to explain
    c. Repeats in own words
    d. Corrects oneself

3. Judgment: Argumentation

1. Flawed argumentation
    a. Opinion changes
    b. Contradictory opinions held simultaneously
2. Agree
    a. In general
    b. Conditionally
    c. Personal situation
    d. No alternative sit. (not fully agree)
3. Disagree
    a. In general
    b. Conditionally
    c. Personal argument
    d. No alternative situations
4. Consistency
5. Moral dilemma

4. Mapping

1. Uncertainty in mapping
    a. In-exclusive meaning of categories
    b. Inadequate categories
    c. Doubt
    d. Reverses response scale
    e. Meaning categories changes
2. Mismatch categories-question
3. Certainty in mapping
    a. Without explanation
    b. Conditional opinion
    c. Absolute opinion

5. Behavior

1. Memory remarks
2. Reflects
    a. Silently
    b. Aloud
3. Suspicion dissatisfaction fatigue
4. Expresses strong feelings

6. Response process

1. Evaluation of question format
2. Evaluation of response process
3. Evaluation of response scale

References

Algemeen Bureau voor de Statistiek (ABS). (2005). *Zevende algemene volks- en woningtelling in Suriname. Landelijke resultaten: Demografische en sociale karakteristieken* [Seventh general population census in Surinam. National results: Demographic and social characteristics]. Retrieved from http://www.statistics-suriname.org/publicaties/census7-resultaten.pdf.

Agresti, A. 2002. *Categorical data analysis*. New York, NY: John Wiley & Sons.

Arce-Ferrer, A. 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*(3), 374-392.

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review, 95*(3), 529–546.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-White differences in response styles. *Public Opinion Quarterly, 48*(2), 491-509.

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143-56.

Belson, W. A. (1986). *The validity in survey research*. Aldershot, UK: Gower.

Berg, I. A., & Collier, J. S. (1953). Personality and group differences in extreme response sets. *Educational and Psychological Measurement, 13*(2), 164-169.

Berry, J. W. (1990). Psychology of acculturation. Understanding individuals moving between cultures. In R. W. Brislin (Ed.), *Applied cross-cultural psychology* (Vol. 14, pp. 232-253). Newbury Park, CA: Sage Publications.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology. Research and applications.* (2nd ed.). Cambridge, UK: Cambridge University Press.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling, 7*(4), 608-628.

Bock, D. R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*(1), 29-51.

Boeije, H. (2005). *Analyseren in kwalitatief onderzoek* [Analysis in qualitative research]. Den Haag, NL: Boom Onderwijs.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: John Wiley & Sons.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review*

Psychology, 53, 605-634.

Bond, M. H., & Yang, K. -S. (1982). Ethnic affirmation versus cross-cultural accommodation: The variable impact of questionnaire language on Chinese bilinguals from Hong Kong. *Journal of Cross-Cultural Psychology, 13*(2), 169-185.

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*(5), 157-161.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference. Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261-304.

Buse, A. (1982). The likelihood ratio, wald, and lagrange multiplier tests: An expository note. *The American Statistician, 36*(3), 153-157.

Byrne, B. M. (1989). Multi-group comparisons and the assumption of equivalent construct validity across groups: Methodological and substantive issues. *Multivariate Behavioral Research, 24*(4), 503-523.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling, 13*(2), 287-321.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*(2), 155-175.

Centraal Bureau voor Statistiek (CBS). (2005). *Jaarrapport Integratie 2005* [Annual report integration 2005]. Retrieved from http://www.cbs.nl/nl-NL/menu/themas/dossiers/allochtonen/publicaties/publicaties/archief/2005/2005-jaarrapport-integratie.htm

CBS. (2010). *Jaarrapport Integratie 2010* [Annual report integration 2010]. Retrieved from http://www.cbs.nl/nl-NL/menu/themas/dossiers/allochtonen/publicaties/publicaties/archief/2010/2010-b61-pub.htm

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *Journal of Cross-Cultural Psychology*, *31*(2), 187-212.

Chun, K.-T., Campbell, J. B., & Yoo, J. H. (1974). Extreme response style in cross-cultural research. A reminder. *Journal of Cross-Cultural Psychology, 5*(4), 465-480.

Clarke III, I. (2001). Extreme response style in cross-cultural research. *International*

Marketing Review, 18(3), 301-324.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10*(1), 3-31.

Dagevos, J., Gijsberts, M., & Van Praag, C. (2003). *Rapportage minderheden 2003. Onderwijs, arbeid en sociaal-culturele integratie* [Minorities report 2003. Education, Employment and social-cultural integration]. Den Haag, NL: SCP.

Data Archiving and Networked Services (DANS) (2002). *Sociale positie en voorzieningengebruik van allochtonen 2002* [Social Position and Utility Use of Ethnic Minorities]. [Data file and code book]. Retrieved from http://easy.dans.knaw.nl/dms?command=home

Davis, D. W. (1997). The direction of race of interviewer effects among African-Americans: Donning the black mask. *American Journal of Political Science, 41*(1), 309-322.

De Jong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research, 45*(1), 104 - 115.

Dohrenwend, B. S., Colombotos, J., & Dohrenwend, B. P. (1968). Social distance and interviewer effects. *The Public Opinion Quarterly, 32*(3), 410-422.

Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multi-group latent class analysis. A primer. *Journal of Cross-Cultural Psychology* 34:195-210.

Ellis, B. B., & Mead, A. D. (2000). Assessment of the measurement equivalence of a Spanish translation of the 16PF questionnaire. *Educational & Psychological Measurement, 60*(5), 787-807.

Gibbons, J. L., Zellner, J. A., & Rudek, D. J. (1999). Effects of language and meaningfulness on the use of extreme response style by Spanish-English bilinguals. *Cross-Cultural Research, 33*(4), 369-381.

Goodman, L. A. (1981). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, *68*(2), 347-355.

Green, D. P., & Citrin, J. (1994). Measurement error and the structure of attitudes: Are positive and negative judgments opposites? *American Journal of Political Science*, *38*(1), 256-281.

Greene, J. C., & Caracelli, V. J. (2003). Making paradigmatic sense of mixed methods practice. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 91-110). Thousand Oaks, CA: Sage Publications.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29*(2), 176-188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly, 56*, 328-351.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley & Sons.

Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin, 69*(3), 192-203.

Harzing, A.-W. (2006). Response styles in cross-national survey research. *International Journal of Cross Cultural Management, 6*(2), 243-266.

Harzing, A. -W., & Maznevski, M. (2002). The interaction between language and culture: A test of the cultural accomodation hypothesis in seven countries. *Language & Intercultural Communication, 2*(2), 120-139.

Heine, S. J., Lehmann, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective likert scales? The reference-group effect. *Journal of Personality & Social Psychology, 82*(6), 903-918.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage Publications.

Hippler, H.-J., & Schwarz, N. (1987). Response effects in surveys. In H.-J. Hippler, N. Schwarz & S. Sudman (Eds.), *Social information processing and survey methodology*. New York, NY: Springer-Verlag

Hui, H. C. and Triandis, H. C. (1985). The instability of response sets. *Public Opinion Quarterly, 49*(2), 253-260.

Hui, H. C., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology, 20*(3), 296-309.

Hurtado, A. (1994). Does similarity breed respect: Interviewer evaluations of Mexican-Descent respondents in a bilingual survey. *The Public Opinion Quarterly, 58*(1), 77-95.

Javeline, D. (1999). Response effects in polite cultures: A test of acquiescence in Kazakhstan. *The Public Opinion Quarterly, 63*(1), 1-28.

Johnson, T. P. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika, 68*(4), 563-583.

Johnson, T. P., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*(2), 264-277.

Johnson, T. P., O'Rourke, D., Chavez, N., Sudman, S., Warnecke, R., Lacey, L., & Horm, J. (1997). Social cognition and responses to survey questions among culturally diverse populations. In L. Lyberg, P. Biemer, M. Collins, E. D. Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey measurement and process quality*, *Wiley series in probability and statistics* (pp. 87-113. New York, NY: John Wiley & Sons.

Johnson, T. P., & Van de Vijver, F. J. R. (2003). Social desirability in cross-cultural research. In J. Harkness, F. J. R. Van de Vijver & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 195-204). Hoboken, New Jersey: John Wiley & Sons.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409-426.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*(3), 381 - 389.

Jöreskog, K. G. (2005). Structural equation modeling with ordinal variables using LISREL. Retrieved from: http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf

Kankaras, M., & Moors, G. (2009). Measurement equivalence in solidarity attitudes in Europe insights from a multiple-group latent-class factor approach. *International Sociology, 24*(4), 557-579.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98*(1), 191-207.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

Krosnick, J. A. (1999). Survey research. *Annual Review Psychology, 50*, 537-567.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*(140), 1-55.

Light, C. S., Zax, M., & Gardiner, D. H. (1965). Relationship of age, sex, and intelligence level to extreme response style. *Journal of Personality & Social Psychology, 2*(6), 907-909.

Lucassen, J., & Penninx, R. (1994). *Nieuwkomers, nakomelingen, Nederlanders. Immigranten in nederland 1550-1993* [Newcomers, offspring, the Dutch. Immigrants in the Netherlands]. Amsterdam: Het Spinhuis.

MacLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: John Wiley & Sons.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan, *The sage* *handbook of quantitative methodology for the social sciences* (pp. 175-198). Thousand Oaks, CA: Sage Publications.

Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology, 31*(1), 223-264.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among hispanics: the role of acculturation and education. *Journal of Cross-Cultural psychology, 23*(4), 498-509.

Marin, G., Triandis, H. C., Betancourt, H., & Kashima, Y. (1983). Ethnic affirmation versus social desirability. Explaining discrepancies in bilinguals' responses to a questionnaire. *Journal of Cross-Cultural Psychology, 2*(14), 173-186.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Meade, A. W., & Lautenschlager, G. J. (2004a). A comparison of item response theory and confirmatory factor analytical methodologies for establishing measurement equivalence/ invariance. *Organizational Research Methods, 7*(4), 361-388.

Meade, A. W., & Lautenschlager, G. J. (2004b). A monte-carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling, 11*(1), 60-72.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543.

Miles, M. B., & Huberman, M. A. (1994). *An expanded sourcebook. Qualitative data analysis*. Thousand Oaks, CA: Sage Publications.

Millsap, R. E. (1995). Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behavioral Research, 30*(4), 577-605.

Moors, G. (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 & Quantity, 37*(3), 277-302.

Moors, G. B. D., & Wennekers, C. H. W. (2003). Comparing moral values in Western European countries between 1981 and 1999. *International Journal of Comparative Sociology, 44*(2), 155-172.

Moors, G. (2004). Facts and artefacts in the comparison of attitudes among ethnic minorities. A multi-group latent class structure model with adjustment for response style behavior. *European Sociological Review, 20*(4), 303-320.

Morgan, D. L., & Krueger, R. A. (1993). When to use focus groups and why. In D. L.

Morgan (Ed.), *Successful focus groups. Advancing the state of the art* (pp. 3-19). Newbury Park, CA: Sage Publications.

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (in press a). Dealing with extreme response style in cross-cultural research: A restricted latent class factor approach. *Sociological Methodology, 41*.

Morren, M., Gelissen, J. P. T. M., & Vermunt, J. K. (in press b). The impact of controlling for extreme responding on measurement equivalence in cross-cultural research. *Methodology.*

Mullen, M. R. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies, 26*(3), 573-596.

Munz, R., Straubhaar, T., Vadean, F., & Vadean, N. (2006). *The costs and benefits of European immigration* (Policy Report 3). Hamburg, DE: Hamburg Institute of International Economics.

Myers, M. B., Calantone, R. J., Page Jr., T. J., & Taylor, C. R. (2000). Academic insights: An application of multiple-group causal models in assessing cross-cultural measurement equivalence. *Journal of International Marketing, 8*(4), 108-121.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: New Jersey Prentice-Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259.

O'Cathain, A., Murphy, E., & Nicholl, J. (2007). Integration and publications as indicators of "yield" from mixed methods studies. *Journal of Mixed Methods Research, 1*(2), 147-163.

Oskamp, S., & Schultz, P. W. (2005). *Attitudes and opinions* (3rd ed.). Mahwah, NJ: Erlbaum.

Oyserman, D., Sakamoto, I., & Lauffer, A. (1998). Cultural accommodation: Hybridity and the framing of social obligation. *Journal of Personality & Social Psychology, 74*(6), 1606-1618.

Pan, C., & Pfeil, B. S. (2003). *National minorities in Europe. Handbook.* Vienna, AT: Braumuller.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightman (Eds.), *Measures of personality and social psychological attitudes* (Vol. 1, pp. 17-60). San Diego, CA: Academic Press.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I.

Braun, D. N. Jackson & D. E. Wiley, *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Erlbaum.

Peabody, D. (1962). Two components in bipolar scales: Direction and extremeness. *Psychology Review, 69*(2), 65-73.

Phalet, K., & Schönpflug, U. (2001). Intergenerational transmission of collectivism and achievement values in two acculturation contexts. *Journal of Cross-Cultural Psychology, 32*(2), 186-201.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.

Poortinga, Y. H., & Van de Vijver, F. J. R. (1987). Explaining cross-cultural differences: bias analysis and beyond. *Journal of Cross-Cultural Psychology, 18*(3), 259-282.

Raftery, A. E. (1999). Bayes factors and BIC. Comment on “A critique of the bayesian information criterion for model selection”. *Sociological Methods & Research, 27*(3), 411-427.

Ralston, D. A., Cunniff, M. K., & Gustafson, D. J. (1995). Cultural accommodation. The effect of language on responses of bilingual Hong Kong Chinese managers. *Journal of Cross-Cultural Psychology, 26*(6), 714-727.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory. *Journal of Applied Psychology, 87*(3), 517-529.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552-566.

Ross, C. E., & Mirowsky, J. (1984). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health & Social Behavior, 25*(2), 189-197.

Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association, 96*(453): 20-31.

Russo, E. J., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal reports. *Memory & Cognition, 17*(6), 759-769.

Saris, W. E. (1998). The effects of measurement error in cross cultural research. In J. Harkness (Ed.), *Cross-cultural survey equivalence. ZUMA- Nachrichten Band 3* (pp.67-87). Mannheim, DE: ZUMA.

Saris, W. E., & Aalberts, C. (2003). Different explanations for correlated disturbance terms in MTMM studies. *Structural Equation Modeling, 10*(2), 193-213.

Schwarz, N., & Sudman, S. (1996). *Answering questions. Methodology for determining cognitive and communicative processes in survey research*. San Francisco, CA: Jossey-Bass Publishers.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage Publications.

Skondral, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.

Smith, E. R. (1998). Mental representation and memory. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 391-445). New York, NY: McGraw-Hill.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90.

Stevens, G. W. J. M., Pels, T. V. M., Vollebergh, W. A. M., & Crijnen, A. A. M. (2004). Patterns of psychological acculturation in adult and adolescent Moroccan immigrants living in the Netherlands. *Journal of Cross-Cultural Psychology, 35*(6), 689-704.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys. A review and synthesis*. Chicago, IL: Aldine Publishing Company.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers. The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass Publishers.

Thurstone, L. M. (1928). Attitudes can be measured. *The American Journal of Sociology, 33*(4), 529-554.

Tourangeau, R. (1987). Attitude measurement: A cognitive perspective. In H.-J. Hippler, N. Schwarz & S. Sudman (Eds.), *Social information processing and survey methodology*. New York, NY: Springer-Verlag.

Tourangeau, R. (2003). Cognitive aspects of survey measurement and mismeasurement. *International Journal of Public Opinion Research 15*(1), 3-7.

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

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions. The impact of data collection mode, question format and question content. *Public Opinion Quarterly, 60*(2), 275-304.

Triandis, H. C. (1990). Theoretical concepts that are applicable to the analysis of ethnocentrism. In R. W. Brislin (Ed.), *Applied Cross-Cultural Psychology* (Vol. 14). Newbury Park, CA: Sage Publications.

Triandis, H. C., & Marin, G. (1983). Etic plus emic versus pseudoetic: A test of a basic assumption of contemporary cross-cultural psychology. *Journal of Cross-Cultural Psychology* 14:489-500.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185:1124-1131.

Van de Vijver, F. J. R. (1998). Towards a theory of bias and equivalence. *ZUMA-Nachrichten Spezial*, 41-65.

Van de Vijver, F. J. R. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31*(1), 33-51.

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research* (Vol. 1). Thousand Oaks, CA: Sage Publications.

Van de Vijver, F. J. R., & Phalet, K. (2004). Assessment in multicultural groups: The role of acculturation. *Applied psychology: An international review, 53*(2), 215-236.

Van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology 47*(4), 263-280.

Van Hemert, D. A., Baerveldt, C., & Vermande, M. (2001). Assessing cross-cultural item bias in questionnaires. Acculturation and the measurement of social support and family cohesion for adults. *Journal of Cross-Cultural Psychology, 32*(4), 381-396.

Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales. Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*(3), 346-360.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.

Vermunt, J. K., & Magidson, J. (2004). Latent class analysis. In E. S. Lewis-Beck, A. Bryman & T. F. Liao (Eds.), *The sage encyclopedia of social sciences research methods* (pp. 549-553). Thousand Oaks, CA: Sage Publications.

Vermunt, J., & Magidson, J. (2005a). Factor analysis with categorical indicators: A comparison between traditional and latent class approaches. In A. Van der Ark, M. A.

Croon & K. Sijtsma (Eds.), *New developments in categorical data analysis for the social and behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Vermunt, J. K., & Magidson, J. (2005b). *Technical guide for Latent GOLD 4.0: Basic and advanced*. Belmont MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2008). *LG-syntaxTM user's guide: Manual for Latent GOLD 4.5 syntax module*. Belmont, MA: Statistical Innovations.

Wallace, R. A., & Wolf, A. (1998). *Contemporary sociological theory. Expanding the classical tradition* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy Marketing Science, 36*(3), 409-422.

Welkenhuysen-Gybels, J., Billiet, J. & Cambré, B. (2006). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology, 34*(6), 702 - 722.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.

Willis, G. B., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology, 5*(3), 251-267.

Wooley, C. M. (2009). Meeting the mixed methods challenge of integration in a sociological study of structure and agency. *Journal of Mixed Methods Research, 3*(1), 7-25.

Yang, K.-S., & Bond, M. H. (1980). Ethnic affirmation by Chinese bilinguals. *Journal of Cross-Cultural Psychology, 11*(4), 411-425.

Summary

Many studies have shown that the cross-cultural comparability of attitudes can be jeopardized by differences in the response processes among culturally diverse respondents. Using a Latent Class Factor Analysis (LCFA) approach, we investigated the extent to which answers to attitude questions which were included in three attitude scales in the SPVA survey[35] are comparable across the four largest minorities in the Netherlands. The LCFA model offered the possibility to do exploratory analyses by including an unrestricted style factor. The results showed that part of the inequivalence could be attributed to an Extreme Response Style (ERS), i.e. the tendency to select (or to avoid) extreme categories on a 5-point response scale. To understand the occurrence of ERS, we held cognitive interviews among interviewees that were purposively selected among the ethnic groups. We integrated both data sets by inferring the response style of respondents in the purposive sample from the statistical model, and comparing the respondents' explanations for their response behavior in the cognitive interview accordingly.

In general, we found that the response style is the outcome of an interaction between the systematical application of response strategies and the interference of cultural traits such as language use and the acculturation style. In Chapter 2, we show how these differences in responding can distort the conclusions about how the groups differ in their attitudes. 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, the groups not only differ in their attitudes but also in ERS.

To examine how extreme responding affects the cross-cultural comparability of survey responses, in Chapter 3 we propose and apply a multiple-group latent class approach where groups are compared on basis of the factor loadings, intercepts and factor means in a Latent Class Factor Model. In this approach a latent factor measuring the response style is explicitly included as an explanation for group differences found in the data. Findings from two empirical applications that examine the cross-cultural comparability of measurements show that group differences in responding import inequivalence in measurements among

[35] SPVA stands for Social Position and Utility Use of Ethnic Minorities. The survey maps the cultural, economic and social life of ethnic minorities in the Netherlands (Dagevos, et al., 2003).

groups. Controlling for the response style yields more equivalent measurements.

In Chapter 4, we show how a differing response style can lead to differences in the response process. We present a mixed methods approach that integrates quantitative and qualitative methods to analyze why the four largest minorities in the Netherlands – Turks, Moroccans, Antilleans and Surinamese – respond differently to items treating cultural topics. First, we conduct Latent Class Factor Analyses on a large nationally representative data set to assess whether these minorities respond differently to the items and to distinguish respondents with different types of response styles. Then we purposely select interviewees from the same cultural groups and classify them according to the response profiles derived from the quantitative study. Moreover, we ask interviewees in cognitive interviews to explain their answers to the same set of items. We find that a response style is related to the extent to which a respondent a) considers the item content, b) weighs arguments for and against an answer, and c) systematically applies response strategies across a range of items.

Depending on the generation of immigration, and the response style, respondents use arguments differently to justify their answer, which leads us to distinguish three response strategies (Chapter 5). The quantitative results indicate that group differences between minorities can partially be explained by response styles. Some respondents prefer to select the extreme categories (Extreme Response Style) whereas others prefer adjacent categories (Avoidance Response Style). The qualitative study shows that extreme responders pay less attention to the precise item wording than other respondents, use more abrupt argumentation and less often reflect on their response behavior. Avoiders have a critical attitude toward surveys, weigh arguments carefully, avoid integrating personal experiences in their answers, and/or use nuanced language. Many respondents use these response strategies but respondents with a response style apply them more systematically.

In summary, we argue that the cross-comparability of measurements can not merely be assumed. We show how group differences in responding to attitude statements can be detected and how part of these response differences can be accounted for by group differences in a response style. Furthermore, we argue that these group differences in responding and response style occur because of the characteristics of the attitude statements and the respondents. Finally, we find evidence that the generation of immigration affects the way in which respondents belonging to a minority deal with attitude statements.

Nederlandse samenvatting

Wereldwijd steunen veel onderzoeken op data verkregen uit vragenlijsten. Of deze vragen wel meten wat ze zouden moeten meten (validiteit) en of de vragen altijd op dezelfde manier beantwoord worden (betrouwbaarheid) speelt een belangrijke rol in het gebruik van deze data. Daarnaast wordt er steeds vaker gekeken naar verschillen en overeenkomsten tussen groepen mensen met verschillende culturele achtergronden. Culturele verschillen in interpretatie, reactie op de onderzoeksmethode of taalgebruik kunnen een extra verstorende factor vormen op de validiteit en betrouwbaarheid.

In dit onderzoek hebben wij gekeken naar de mate waarin de vier grootste minderheden in Nederland – Turken, Marokkanen, Antillianen en Surinamers – verschillen in hun antwoorden op vragen gesteld in de SPVA survey over familie waarden, de houding van minderheden ten opzichte van Nederland, en de mate waarin kinderen in een gezin autonoom worden behandeld. Daarbij maakten we gebruik van latente klassen factor analyse (LCFA), een methode die de mogelijkheid gaf om exploratieve analyses te doen met behulp van een ongerestricteerde stijl factor. Uit de resultaten kunnen we concluderen dat de verschillen in antwoorden gedeeltelijk verklaard werden door verschillen in het gebruik van een extreme responsstijl (ERS), de tendens om extreme antwoordcategorieën (zoals helemaal mee eens) te kiezen of te vermijden. Om te begrijpen hoe het komt dat de minderheden verschillen in deze responsstijl, hebben we cognitieve interviews gehouden onder mensen die we doelgericht hebben geselecteerd uit de vier minderheidsgroepen. De twee datasets hebben we geïntegreerd door een responsstijl toe te kennen aan de geïnterviewden met behulp van het statistische model en hun antwoorden te vergelijken.

Over het algemeen vonden we dat de responsstijl het resultaat is van een interactie tussen het systematisch toepassen van de responsstrategieën en de invloed van culturele eigenschappen zoals het gebruik van taal en de acculturatie stijl. In hoofdstuk 2 hebben we laten zien dat de verschillen in antwoorden de conclusies over de meningen van minderheden kan verstoren. Naast het geven van een overzicht van de beschikbare statistische methoden om ERS te hanteren, beargumenteren we dat de latente klassen factor aanpak (LCFA), geïntroduceerd door Moors (2003), verschillende voordelen heeft vergeleken met andere methoden. Moors' methode betreft het definiëren van een latent variabele model waarin, naast factoren die de meningen meten, een factor is opgenomen die de responsstijl meet. In LCFA, de geobserveerde antwoorden kunnen behandeld worden als nominale antwoorden wat nodig is voor het meten van ERS. We vinden sterke aanwijzingen dat ERS aanwezig is en dat de groepen niet alleen verschillen in hun meningen maar ook in hun responsstijl.

Om te bekijken hoe de extreme responsstijl de crossculturele vergelijkbaarheid van survey antwoorden beïnvloed, stellen we in hoofdstuk 3 een multiple groep latente klassen aanpak voor waarin groepen worden vergeleken op basis van de factor ladingen, de intercept en de factor gemiddelden. In deze aanpak wordt een latente factor opgenomen als een verklaring voor de groepsverschillen in de data. Bevindingen van twee empirische toepassingen waarin de crossculturele vergelijkbaarheid onderzocht wordt laten zien dat de groepsverschillen in ERS onvergelijkbare metingen tot gevolg hebben. Controleren voor deze responsstijl geeft meer vergelijkbare metingen.

In hoofdstuk 4 laten we zien hoe verschillen in een responsstijl kan leiden tot verschillen in een respons proces. We presenteren een nieuwe aanpak waarin statistische methoden met kwalitatieve interviews gecombineerd worden om te analyseren waarom de vier minderheden verschillend reageren op vragen over culturele onderwerpen. Ten eerste, we hebben een LCFA analyse toegepast op een grote dataset om te kijken in hoeverre minderheden verschillend reageren op de vragen en in welke mate respondenten verschillen in hun responsstijl. Vervolgens hebben we respondenten geselecteerd van dezelfde culturele minderheden en hen geclassificeerd op basis van de respons profielen die naar voren kwamen in de kwantitatieve studie. Daarnaast vragen we respondenten in de cognitieve interviews om hun antwoorden (op dezelfde vragen als geanalyseerd in hoofdstuk 2 en 3) toe te lichten. Uit de resultaten kunnen we opmaken dat de responsstijl gerelateerd is aan de manier waarop een respondent a) reageert op de inhoudelijke betekenis van een vraag, b) zijn voor- en tegenargumenten afweegt, en c) systematisch deze en andere antwoordstrategieën toepast in een grote set van vragen.

Afhankelijk van de generatie waarin een familie naar Nederland is geïmmigreerd en hun responsstijl, gebruiken respondenten verschillende argumenten om hun antwoorden te verantwoorden. Deze verschillen vormden de aanleiding om drie verschillende antwoordstrategieën te onderscheiden (hoofdstuk 5). De kwantitatieve resultaten geven weer dat groepsverschillen tussen minderheden gedeeltelijk kan verklaard worden door de aanwezigheid van een responsstijl. Sommige respondenten prefereren extreme categorieën (ERS) terwijl andere respondenten vaker de aangrenzende categorieën (AvRS) kiezen zoals ‘mee eens’. De kwalitatieve studie laat zien dat respondenten met ERS minder aandacht besteden aan de manier waarop de vraag is geformuleerd is dan de andere respondenten, vaker korte argumenten geven, en minder vaak reflecteren op hun eigen responsgedrag. Respondenten die juist aangrenzende categorieën preferen, hebben een kritische houding ten opzichte van vragenlijstonderzoek, wegen hun argumenten zorgvuldig, vermijden hun persoonlijke ervaringen te integreren in hun antwoorden, en/of gebruiken genuanceerde taal.

Veel respondenten gebruiken deze antwoordstrategieën maar respondenten met een responsstijl passen ze systematischer toe.

Samenvattend, we stellen vast dat de crossculturele vergelijkbaarheid van metingen niet zomaar kan worden aangenomen. We laten zien hoe groepsverschillen in het beantwoorden van attitudevragen kunnen worden gedetecteerd en hoe deze groepsverschillen gedeeltelijk verklaard kunnen worden door groepsverschillen in een responsstijl. Een andere bevinding is dat deze groepsverschillen in het beantwoorden van surveyvragen voorkomen door specifieke eigenschappen van de vragen en de respondenten. Tenslotte, we suggeren dat de generatie van immigratie de manier beinvloedt waarop respondenten uit minderheden omgaan met surveyvragen.