

**Response Style and Rating Scales:  
The Effects of Data Collection Mode, Scale Format, and Acculturation**

by

**Mingnan Liu**

**A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Survey Methodology)  
in the University of Michigan  
2015**

**Doctoral Committee:**

**Research Professor Frederick G. Conrad, Co-Chair  
Assistant Research Scientist Sunghee Lee, Co-Chair  
Professor Norbert W. Schwarz, University of Southern California  
Professor Yu Xie**

© Mingnan Liu

2015

## Table of Contents

List of Tables .....	iv
List of Figures .....	v
List of Appendices .....	vi
Abstract .....	vii
Chapter 1 Introduction .....	1
References .....	9
Chapter 2 Examining Acquiescent and Extreme Response Styles between Face-to-Face and Web Surveys.....	12
2.1 Introduction.....	12
2.2 Literature review .....	13
2.3 Data and measures .....	18
2.4 Analytic approach .....	23
2.5 Model Specification .....	27
2.6 Results for mode effect .....	30
2.7 Discussion .....	37
References .....	41
Chapter 3 Comparing Extreme Response Styles between Agree-Disagree and Item Specific Scales .....	46
3.1 Introduction.....	46
3.2 Literature review .....	48
3.3 Data and measures .....	51
3.4 Analytical approaches and results.....	54
3.4.1 Between-subject experiment analysis.....	54
3.4.2 Between-subject experiment results .....	58
3.4.3 Within-subject experiment analysis.....	67
3.4.4 Within-subject experiment results .....	70
3.5 Discussion .....	75
References .....	80
Chapter 4 The Effect of Acculturation on Extreme Response Style: A Mediation Analysis among a Sample of Arab American Adults .....	89
4.1 Introduction.....	89
4.2 Literature Review.....	90
4.2.1 Extreme response style .....	90
4.2.2 Culture and response styles .....	91

4.2.3 Arabic culture, acculturation, and their impacts on response style.....	93
4.3 Data, Measures, and Models.....	97
4.3.1 Data source.....	97
4.3.2 Dependent variables.....	98
4.3.3 Independent variables.....	98
4.3.4 Mediation variable.....	99
4.3.5 Control variables.....	99
4.3.6 Analytical strategy.....	102
4.4 Results.....	104
4.5 Discussion.....	110
References.....	113
Chapter 5 Conclusion.....	123
References.....	126

## List of Tables

Table 2.1. Question Wordings, Means, and Standard Errors (S.E.) by Data Collection Mode, 2012 American National Election Studies.....	20
Table 2.2. Descriptive data for independent and control variables, 2012 American National Election Studies. ....	22
Table 2.3. Model Fit Statistics, 2012 American National Election Studies. ....	28
Table 2.4. Estimated Regression Coefficients and Standard Errors of ERS and ARS on the Likert Scale Items, 2012 American National Election Studies. ....	32
Table 2.5. Estimated Regression Coefficients (Log Odds) and Standard Errors of Mode, Race, Interaction, and Control Variables on Extreme Response Style (ERS) and Acquiescent Response Style (ARS), 2012 American National Election Studies. ....	35
Table 3.1. Proportions and standard errors of extreme responses for Agree/Disagree (A/D) and Item Specific (IS) scales, 2012 ANES.....	59
Table 3.2. Model fit statistics (BIC) for Agree/Disagree (A/D) and Item Specific (IS) scales, 2012 ANES pre-election and post-election.....	62
Table 3.3. Model fit statistics (BIC) for within-subject analysis, 2012 ANES pre-election and post-election.....	71
Table 4.1. Descriptive Data for Variables in the Analysis .....	101
Table 4.2. Mediation Effects of Acculturation on Extreme Response Style, 5-point scales. ....	106
Table 4.3. Mediation Effects of Acculturation on Extreme Response Style, 3-point scales. ....	109

## List of Figures

Figure 1.1. Illustration of four most frequently studied response styles.....	3
Figure 2.2. Interaction effects between mode and race/ethnicity on extreme response style, for a 60-69-year-old male respondent with high school or lower education, and whose household income is set at the sample mean.....	36
Figure 3.1. Experimental design and sample sizes. ....	53
Figure 3.2. Latent class model with one style factor and two content factors. Items 1-4 represent A/D1-A/D4 or IS1-IS4.....	57
Figure 3.3. Extreme response style coefficients on agree/disagree (A/D) and item-specific (IS) scales, 2012 ANES pre-election survey. See Appendix for response categories 1-5. ....	65
Figure 3.4. Extreme response style coefficients on agree/disagree (A/D) and item-specific (IS) scales, 2012 ANES post-election survey. See Appendix for response categories 1-5. ....	66
Figure 3.5. (a) Latent class model with one style factor and four content factors; (b) latent class model with two style factors and four content factors.....	69
Figure 3.6. Extreme response style coefficients for Group 1 (A/D in the pre-election and IS in the post-election), Model 2b (Table 3.3), 2012 ANES. See Appendix for response categories 1-5. .	73
Figure 3.7. Extreme response style coefficients for Group 2 (IS in the pre-election and A/D in the post-election), Model 2b (Table3), 2012 ANES. See Appendix for response categories 1-5. ....	74
Figure 4.1 Theoretical Multilevel Generalized Structural Equation Modeling of Acculturation, Interview Language, and Extreme Response Style (ERS). Double-circle refers to level-2 variable. ....	96

## **List of Appendices**

Appendix 3.1 Question wording and response options for agree-disagree and item specific questions. ....	86
Appendix 3.2 Means and standard errors of demographic variables for Agree/Disagree (A/D) and Item Specific (IS) scales by mode of data collection, 2012 ANES. ....	88
Appendix 4.1 Question wordings and response options. ....	120
Appendix 4.2 Proportions of Extreme Responses for Each Question by Acculturation Level. .	122

## **Abstract**

Rating scales are popular for measuring attitudes, but response style, a source of measurement error associated with this type of question, can result in measurement bias in important attitudinal measures. Although numerous research efforts have been devoted to this topic, there are still some overlooked areas. This dissertation intends to fill three gaps in the literature on response style. Chapter 1 investigates the effects of face-to-face and Web survey on acquiescent response style (ARS) and extreme response style (ERS) using the 2012 American National Election Studies (ANES) data. Using the latent class analysis approach, I find that: 1) both ARS and ERS exist in both face-to-face and Web survey; 2) face-to-face respondents demonstrate more ARS and ERS than Web respondents; 3) the effect of mode on ERS is larger for black respondents than for white and Hispanic respondents. Chapter 2 compares ERS with respect to the format of response scale, specifically agree-disagree (A/D) and item specific (IS) scales. This study analyzes a between- and within-subject experiment embedded in the 2012 ANES. Using latent class factor analysis, I reached the following three major findings: 1) ERS exists in both A/D and IS scale formats; 2) ERS shows a slightly different pattern between the two scale formats; 3) when analyzing ERS within subjects across two waves, there is only a single ERS latent class variable for both scale formats, after controlling for the correlation within respondents. Chapter 3 utilizes the 2003 Detroit Arab American Study to examine the impact of acculturation of Arab Americans on ERS. The results indicate that less acculturated respondents are more prone to ERS than more acculturated respondents, and this is especially true for the



5-point rating response scales as compared to 3-point rating scales. This phenomenon can be explained by the fact that less acculturated respondents identify more strongly with honor-based collectivist cultures that value decisive and assertive answers since this is a way of showing one's unambiguous attitude and standing, an important quality in such a culture. The language of the interview primes the relevant cultural norms and therefore mediates the relationship between acculturation and ERS.

## **Chapter 1**

### **Introduction**

Attitudinal questions, as distinct from factual questions, measure constructs that are not directly observable. One's attitude is regarded as an underlying phenomenon or construct and is sometimes called a latent variable, a variable that is neither manifest nor constant. Rating scales are widely used in survey research to measure the underlying construct of one's attitude. In a perfect world, we hope to accurately and reliably measure the underlying construct using survey questions, with only the random error. In reality, however, the measurement we have is usually far from perfect. Consequently, much research has been devoted to this type of question. One of the most fruitful lines of study is response style, a particular type of measurement error associated with rating scales. Response style describes the phenomenon that, regardless of the content of the survey question, respondents give answers based on some content-irrelevant criteria (Paulhus 1991). Since Cronbach (1941; 1942; 1946; 1950) developed and popularized the concept of response style more than 70 years ago, much attention has been given to this research area. Some review papers systematically synthesize the results of decades of research on response styles (Rorer 1965; Yang et al. 2010; Vaerenbergh and Thomas 2012).

To facilitate an easy understanding of this concept, I illustrate several frequently studied response styles in Figure 1.1. Acquiescent response style refers to the tendency to choose "agree" or "yes" responses to survey questions more frequently than other response options. Extreme response style refers to the tendency to choose the two endpoints of a response scale more frequently than other categories. Two response styles, midpoint response style and mild response

style, represent an approach opposite to extreme response style: with midpoint response style, respondents tend to choose the middle response option more often than others, while with mild response style, respondents tend to choose the intermediate options rather than the two extreme response options. Finally, disacquiescence response style is the opposite of acquiescence response style.

Strongly agree

Strongly disagree

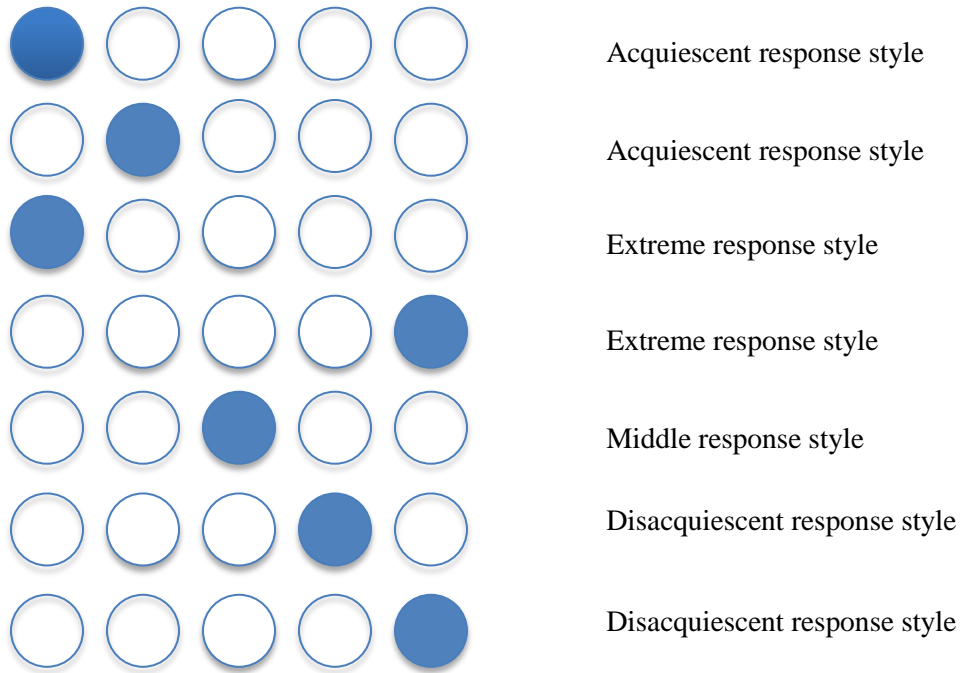


Figure 1.1. Illustration of four most frequently studied response styles.

Response style distorts the survey estimate in that the answers respondents provide is a blend of their true attitudes, or underlying constructs, and their consistent response behaviors that are independent of question contents. That is, response style can be viewed as a systematic measurement error component that introduces bias into the survey estimate. For example, acquiescent response style will artificially bias the estimate towards the positive end. Consequently, the survey estimate tends to overstate the positivity if the true attitude, which is in fact not as positive as it looks like. For another example, extreme response style drives the response away from the middle and therefore increases the variance of the estimate. Bias may also exist if the tendency of one end of the scale is higher than the other end. If more people are using the positive endpoint than the negative end, then a positive bias may exist. On the other hand, if people have a higher tendency of choosing the negative end of the scale, a negative bias will arise.

Given the popularity of rating scales, developing a good understanding of the response styles is an urgent task faced by survey researchers. Although there is a rich literature on response styles, it is still lacking in several respects. Response style is a type of measurement error. To improve the measurement of a rating scale, the first step is to achieve a solid understanding of the source of response style bias. The response style bias can come from a variety of channels, and this dissertation is designed to explore three understudied aspects of response style. In particular, this dissertation research expands the existing literature on response style in three directions, namely the impacts of data collection mode, the scale design format, and the respondent's acculturation level on response styles. Specifically, this dissertation has three research objectives:

- Objective 1: Assess survey mode effects on response styles and their differential effects across racial and ethnic groups.
- Objective 2: Compare the response style observed between two scale formats and the stability of the response style between the two scale formats overtime.
- Objective 3: Identify the relationship between response style and acculturation among a sample of Arab Americans.

This dissertation comprises three substantive chapters, one for each research objective. Chapter 2 of the dissertation assesses survey mode effects on response style and their differential effects across racial and ethnic groups. In this study, I compare the incidence of acquiescence response style (ARS) and extreme response style (ERS) between face-to-face and Web surveys using the 2012 American National Election Studies (ANES). The literature suggests that answering “agree” or “yes” is seen as a more polite conversational interaction compared to giving a negative response (Leary and Kowalski 1990; Couch and Keniston 1960). The positive or affirmative response creates a positive self-image. Therefore, ARS is mainly used out of consideration for image management and the fear of potential negative judgment from others. As for ERS, some earlier research has pointed out that extreme responses are considered more sincere than non-extreme responses (Marin, Gamba, and Marin 1992; Bachman and O’Malley 1984). Compared to extreme answers, which clearly express one’s attitudes and beliefs, middle responses make it more difficult for others to interpret responses. The more interpretable extreme opinions better promote social harmony, and create a polite, agreeable, likeable, and respectful interpersonal interaction atmosphere (Hui and Triandis 1989). The presence of the interviewer in

the face-to-face interview may stimulate the respondents to portray themselves as more polite (more ARS) and sincere (more ERS), in comparison to the self-administered Web survey.

Chapter 3 compares the ERS between two types of rating scales, namely the item-specific rating scale and the agree-disagree Likert scale. In contrast to the Likert scale, in which all questions use the same agree-disagree response options, the item-specific scale adopts response options that are specific to each question. Several earlier studies have shown that the item-specific rating scale reduces the incidence of ARS (Berkowitz and Wolkon 1964; Ross, Steward, and Sinacore 1995). However, not much is known about the ERS in the context of the item-specific scale. This chapter also utilizes data from the 2012 ANES, in which was embedded both a between-subject and a within-subject experiment. This chapter seeks to achieve two goals: First, do respondents show different patterns of ERS when using an IS scale compared to an A/D scale? Second, is ERS a stable trait over time, or it is a phenomenon mainly driven by survey design features, such as question formats?

Chapter 4 studies the impact of acculturation on ERS among Arab Americans using the 2003 Detroit Arab American Study. Two previous studies have explored the relationship between acculturation and response styles (Davis, Resnicow, and Couper 2010; Marin, Gamba, and Marin 1992). Both of them target the U.S. Hispanic population. This chapter expands this line of study in two directions. First, this chapter focuses on Arab Americans, an ethnic minority group with a unique cultural background. The Arabic culture is based on honor-based collectivism (Uskul, Oyserman, and Schwarz 2010). Since it is generally believed that response style is a reflection of a society's cultural norms, the Arabic culture should elicit a different pattern of ERS than the individualistic culture in the U.S. The acculturation level determines the degree to which the respondent and his or her response style reflect the norms of Arabic culture

or American culture. Second, the language of interview is treated as a mediating variable between the acculturation and response style. The interview language can prime one culture, and the activated culture will not only determine the respondent's communication style but also shape the cognitive construct. When interviewed in Arabic, the honor-based collectivist culture is likely to be activated and, as a result, respondents' response style will reflect the norms of that culture. Consequently, respondents will give more extreme responses. In contrast, less ERS will be observed from more acculturated respondents who speak English in the interview.

The contributions of this dissertation are both practical and theoretical. On the one hand, this dissertation will provide survey researchers and practitioners with practical applications of research on the choice of data collection mode and design of rating scales. The result from Chapter 2 will inform us the differential mode effect on response styles. If a survey is mainly composed of questions with rating scales, response style bias should be a factor for consideration when choosing the mode. A mixed-mode survey design may be more appropriate when the rating scale only constitutes a small portion of the survey. Similarly, designing a rating scale other than the agree/disagree Likert scale, as in Chapter 3, can also influence the rating scale data quality. This dissertation also explores the mechanism behind the response styles. For example, it is widely believed that response styles reflect the mainstream cultural norm of a society. A contrast between the norms can result in distinctly different response style patterns (for review, see Yang et al. 2010). Immigrants, such as the Arab American respondents examined in Chapter 4, are subject the cultural norms of both their host and home countries, and their acculturation levels can determine the extent to which they follow one cultural norm or the other. By establishing a relationship between the acculturation and response style, this chapter provides evidence on the hypothesized theoretical relationship between culture and response style.



I hope that this dissertation will provide more insights into the mechanism of response style as well as some practical information on the design of rating scales. However, the investigations of this dissertation are but small steps in an attempt fill a much larger void in our collective knowledge about response styles. My ultimate goal in writing this dissertation is to inspire further study on response style. This is especially important, as we increasingly rely on survey data to help us make predictions and decisions.

## References

- Bachman, Jerald G., and Patrick M. O'Malley. 1984. "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles." *Public Opinion Quarterly* 48 (2): 491–509.
- Berkowitz, Norman H., and George H. Wolkon. 1964. "A Forced Choice Form of the F Scale-Free of Acquiescent Response Set." *Sociometry* 27 (1): 54–65.
- Couch, Arthur, and Kenneth Keniston. 1960. "Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable." *The Journal of Abnormal and Social Psychology* 60 (2): 151–74.
- Cronbach, Lee J. 1941. "An Experimental Comparison of the Multiple True-False and Multiple Multiple-Choice Tests." *Journal of Educational Psychology* 32 (7): 533–43.
- . 1942. "Studies of Acquiescence as a Factor in the True-False Test." *Journal of Educational Psychology* 33 (6): 401–15.
- . 1946. "Response Sets and Test Validity." *Educational and Psychological Measurement* 6 (4): 475–94.
- . 1950. "Further Evidence on Response Sets and Test Design." *Educational and Psychological Measurement* 10: 3–31.
- Davis, R. E., K. Resnicow, and M. P. Couper. 2010. "Survey Response Styles, Acculturation, and Culture Among a Sample of Mexican American Adults." *Journal of Cross-Cultural Psychology* 42 (7): 1219–36.
- Hui, C. Harry, and Harry C. Triandis. 1989. "Effects of Culture and Response Format on Extreme Response Style." *Journal of Cross-Cultural Psychology* 20 (3): 296–309.

- Leary, Mark R., and Robin M. Kowalski. 1990. "Impression Management: A Literature Review and Two-Component Model." *Psychological Bulletin* 107 (1): 34.
- Marin, Gerardo, Raymond J. Gamba, and Barbara V. Marin. 1992. "Extreme Response Style and Acquiescence among Hispanics The Role of Acculturation and Education." *Journal of Cross-Cultural Psychology* 23 (4): 498–509.
- Paulhus, Delroy L. 1991. "Measurement and Control of Response Bias." In *Measures of Personality and Social Psychological Attitudes. Measures of Social Psychological Attitudes, Vol. 1*, edited by John P. Robinson, Phillip R. Shaver, Lawrence S. Wrightsman, 17-59, San Diego, CA: Academic Press
- Rorer, Leonard G. 1965. "The Great Response-Style Myth." *Psychological Bulletin* 63 (3): 129–56.
- Ross, Caroline K., Colette A. Steward, and James M. Sinacore. 1995. "A Comparative Study of Seven Measures of Patient Satisfaction." *Medical Care* 33 (4): 392–406.
- Uskul, Ayse K., Daphna Oyserman, and Norbert Schwarz. 2010. "Cultural Emphasis on Honor, Modesty, or Self-Enhancement: Implications for the Survey-Response Process." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith, 191-201, Hoboken, New Jersey: Wiley.
- Vaerenbergh, Yves Van, and Troy D. Thomas. 2012. "Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies." *International Journal of Public Opinion Research* 25 (2): 195–217.

Yang, Yongwei, Janet A. Harkness, Tzu-Yun Chin, and Ana Villar. 2010. "Response Styles and Culture." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith, 203–223, Hoboken, New Jersey: Wiley.

## **Chapter 2**

### **Examining Acquiescent and Extreme Response Styles between Face-to-Face and Web Surveys**

#### **2.1 Introduction**

Likert scales are popular for measuring attitudes, but response style, a source of measurement error associated with this type of question, has been haunting researchers for decades. Response style describes the phenomenon that, rather than responding to the specific survey question, the respondent gives an answer that is based on some content-irrelevant criteria (Paulhus 1991). Acquiescence response style (ARS) and extreme response style (ERS) are two of the most frequently studied response styles in the literature. ARS refers to the tendency to choose “agree” or “yes” responses to survey questions more frequently than other response options (Baumgartner and Steenkamp 2001). ERS refers to the tendency to choose the two endpoints of a response scales more frequently than other categories (Paulhus 1991).

Facing this issue, researchers have developed several methods to identify and reduce the measurement error caused by response style. Earlier research suggests using theoretically heterogeneous items to measure ERS and using balanced scales with reversed worded items to measure ARS. More recently, another cluster of methods for examining response style has emerged, which falls within the latent structure model realm. To name just a few of these methods, confirmatory latent class analysis (Bolt and Newton 2011; Kieruj and Moors 2010; Kieruj and Moors 2013; Kankaraš and Moors 2009; Moors 2008), latent class regression analysis

(Moors 2010), confirmatory factor analysis (Billiet and Davidov 2008; Billiet and McClendon 2000; Welkenhuysen-Gybels, Billiet, and Cambré 2003), and representative indicators response style means and covariance structure (RIRSMACS) (Weijters, Schillewaert, and Geuens 2008; Thomas, Abts, and Weyden 2014) can measure and correct response style error using existing scales without necessarily creating additional special items such as heterogeneous items. (RIRSMACS is an exception.)

This paper will utilize the latent class analysis approach to study a largely overlooked area of response style, which is the data collection mode effect. In particular, this study will focus on both ARS and ERS and examine the effects of two data collection modes: face-to-face and Web surveys. In addition, we will also examine the differential mode effects on different racial/ethnic groups. More specifically, we will study whether the non-Hispanic white, non-Hispanic black, and Hispanic respondents will show more similar or different response style patterns in the two survey modes. As far as we know, there are only three studies that have examined the mode effects on response style, and none of these involve both face-to-face and Web interviews. Also, those studies only examined the overall mode effect without looking at the interaction effect between mode and race/ethnicity. Racial and ethnic differences of response styles have been well documented (Yang et al. 2010). In order to produce comparable data, it is important to evaluate the effect of modes on response styles for different racial/ethnic groups.

## **2.2 Literature review**

Although studies on mode effects on survey measurement and response style are both abundant in the literature, these two are not often studied in conjunction with one another. One of the earliest attempts compares telephone versus face-to-face interviews using a city survey

(Jordan, Marcus, and Reeder 1980). This study shows that telephone interviews produce more acquiescent (i.e., saying “agree” or “yes”), extreme, and “Don’t know” responses than do face-to-face interviews. The authors of this study conclude that telephone surveys produce worse data quality and attribute it to the lack of motivation in the telephone mode. A later study uses a university sample to examine electronic and paper mail surveys and finds that surveys completed in these two modes exhibit similar acquiescent and extreme response styles (Kiesler and Sproull 1986). More recently, Weijters, Schillewaert, and Geuens (2008) examine four types of response styles (i.e., acquiescence, disacquiescence (saying “disagree”), extreme, and midpoint) across three different data collection modes, namely paper and pencil, telephone, and Web using RIRSMACS method. The results show that telephone surveys produce the least midpoint but the most acquiescence response style. The higher level of obligation and engagement of the respondents are speculated to be the cause of the fewer nonsubstantive neutral responses in the telephone mode compared to the other two modes studied. Also, the prevalence of acquiescent response style is due to the higher time pressure associated with telephone surveys. As we will review next, time pressure makes respondents more likely to give a quick response, which fosters the acceptance rather than rejection of a statement. Although the method used in the Weijters et al. study both identifies and corrects for response styles, it has a major drawback: RIRSMACS requires the collection of data from additional items other than the scale of primary interest. Consequently, this method can substantially increase the cost of the survey and the burden on the respondents, which makes it impractical or inapplicable in many situations.

What contributes to acquiescent responses? In the literature, acquiescence has been discussed via two approaches: motivational and cognitive. The motivational approach regards acquiescence as a form of impression management, which means that people behave in certain

ways in order to create certain images of themselves for other people (Leary and Kowalski 1990). When answering questionnaires, the respondent adopts this yea-saying strategy to avoid creating a potential negative self-image resulting from negative responses (Couch and Keniston 1960). Some theorists even posit that, respondents may provide the acquiescent answers solely based on the response options without even reading the questions (Schlenker 1980; Leary and Kowalski 1990). In contrast, the cognitive approach argues that acquiescence is the consequence of the cognitive process of interpreting the questions and forming answers (Cronbach 1942). Gilbert (1991) describes a two-stage belief process, namely representation and assessment. In the representation stage, human minds comprehend information, during which acceptance of the statement is the premise. In order to comprehend, people usually accept the provided statement fast and automatically. In the assessment stage, one will reconsider the statement based on evidence or logic and will decide whether to reject it or not. These sequential cognitive processes make the acceptance happen faster with less effort while rejection happens slower and with more difficulty. The study by Knowles and Condon (1999) classifies people into three groups, yea-sayer, nay-sayer, and appropriate responders. While the response latencies of saying “yes” or “no” are similar for the latter two groups, it takes significantly less time for yea-sayers to say “yes” than “no”. This finding lends support to the Spinoza’s two-stage belief theory that acquiescence is the result of incomplete mental process.

We predict more ARS in face-to-face interviews than Web questionnaires because when facing an interviewer, respondents are more likely to provide agreeable answers; this may be motivated by the fear of potential negative judgment from the interviewer. Web surveys, in contrast, will elicit less acquiescent answers since the desire for image management is much lower given the absence of the interviewer. Also, in face-to-face surveys, respondents are under



greater time pressure, and they do not have entire control of the flow of the process, as compared to Web survey in which respondents can respond at their own pace. Consequently, it is reasonable to assume that respondents under the face-to-face condition process the items only at the comprehension stage, which leads to automatic acceptance. However, a Web survey imposes fewer time constraints and allows respondents to go through both stages. As a result, less acquiescence will be observed from responses to the Web survey. So, the two aforementioned mechanisms behind ARS lead to the same prediction.

When it comes to ERS, earlier research has pointed out that extreme responses are considered more sincere than nonextreme responses (Marín, Gamba, & Marín 1992; Bachman & O'Malley 1984a; Bachman, O'Malley, & Freedman-Doan 2010). The middle of the scale often indicates a lack of opinion, insufficient knowledge to answer the question, or an ambiguous opinion. Compared to extreme answers, which clearly express one's attitudes and beliefs, middle responses make it more difficult for others to interpret responses. The more interpretable extreme opinions better promote social harmony, and create a polite, agreeable, likeable, and respectful interpersonal interaction atmosphere (Hui and Triandis 1989). It is reasonable to expect that respondents are more likely to give extreme answers as a sign of politeness when interacting with an interviewer than doing a Web survey. These considerations lead to the hypothesis that, in face-to-face surveys as compared to Web surveys, respondents have a higher tendency to present themselves as being more sincere to the interviewer by choosing the more extreme response options.

This study will also examine the differential mode effect across three race/ethnicity groups, namely non-Hispanic white, non-Hispanic black, and Hispanic. Research in the U.S. has demonstrated that racial/ethnic minorities have different response styles than Caucasians. More

specifically, black and Hispanic respondents have a higher tendency toward ARS and ERS (Bachman and O'Malley 1984a; Bachman and O'Malley 1984b; Bachman and O'Malley 2013; Warnecke et al. 1997; Ross and Mirowsky 1984; Aday, Chiu, and Andersen 1980; Hui and Triandis 1989; Clarke III 2001; Gibbons, Zellner, and Rudek 1999; Marin, Gamba, and Marin 1992; Davis, Resnicow, and Couper 2010). People of relatively powerless social status are more likely to adopt acquiescence as an impression management strategy (Ross and Mirowsky 1984; Marin and Marin 1991). That is, acquiescence is the consequence of deference and nonresistance from respondents of lower socioeconomic status or minority racial/ethnic groups in order to present a positive self-image to the more powerful members in the society and seek their acceptance (Carr 1971). Respondents employ ARS as an impression management strategy in the survey setting particularly when facing an interviewer belonging to a more powerful group. For ERS, the honor-based collectivist cultural orientation of some minority groups will foster rather than dampen the use of extreme responses because being seen as sincere creates not only a positive representation of the respondent but also of his close relations and associates. The subgroup culture, which includes the strong family bonds and support present among both African Americans and Hispanics, can increase the need for a sincere portrayal of self because "every individual is regarded as a walking symbol of his family" (Madsen 1964, 46).

When the mode effect interacts with the sub-group cultures, we expect to discover more ARS and ERS from black and Hispanic respondents than from white respondents in the face-to-face interview since the need for positive self-presentation is more acute among the two minority groups when interviewers are present. In contrast, the need for positive self-presentation among the three groups of respondents may be more similar in the Web survey when the interviewers are removed from the data collection process and the respondents have no one to impress.

Therefore, the differential mode effect will lead to a significant interaction between mode and race/ethnicity. On the other hand, it is also possible that the cultural influence is so strong that even in the self-administered Web survey the minority respondents still provide more acquiescent and extreme answers than white respondents. In that case, there will be only two main effects for mode and race/ethnicity but no interaction effect.

### **2.3 Data and measures**

The 2012 American National Election Studies (ANES) Time Series survey provides a unique data set for testing the mode effects on ARS and ERS. 2012 was the first year in which ANES conducted two parallel surveys, one through face-to-face interviews and another through Web, using an identical questionnaire. Respondents under each mode completed two interviews, one pre-election and one post-election. In total, the 2012 ANES completed 5916 interviews. The variables in our analysis all come from the post-election interview, of which 1929 surveys were conducted through face-to-face interview and 3581 through the Web. The response rates were 38% and 2% for face-to-face and Web surveys respectively (AAPOR RR1), and the post-election re-interview rates were 94% and 93% respectively.

The two modes used two independent samples, both targeting U.S. citizens aged 18 years or older by Election Day 2012. For the face-to-face interviews, an address-based, stratified, multi-stage cluster sample was selected. The face-to-face sample contained both a nationally-representative main sample and two oversamples, for blacks and Hispanics. All selected households completed a screening questionnaire and one person per household was randomly selected to participate in the face-to-face interview. For the Web survey, respondents came from the GfK KnowledgePanel, a nationally representative probability Web panel. Respondents were

initially recruited through either address-based sampling or random-digit dialing. During the recruitment stage, all household members were enumerated, and one person was selected from each household to become the panel member and participate in surveys.

The analysis examines three sets of balanced Likert scales from the survey to measure the ARS and ERS. Each 5-point scale ranges from disagree strongly (1) to agree strongly (5). The first two sets of scales contain four items, and the third scale contains six items. For each scale, half of the questions are positively worded, and the other half are negatively worded. The first set of questions asks about attitudes toward traditionalism, the second asks about the position of Blacks in the current society, and the third asks about attitudes toward egalitarianism. The question wordings of the 14 Likert scale questions, weighted means, and standard errors are presented in Table 2.1. The estimated means of 10 items and the estimated standard errors of 11 items are larger in the face-to-face than Web survey. This indicates more ARS and ERS in face-to-face than Web surveys, based on the bivariate analysis.

Table 2.1. Question Wordings, Means, and Standard Errors (S.E.) by Data Collection Mode, 2012 American National Election Studies.

	Face-to-face		Web	
	Mean	S.E.	Mean	S.E.
<i>Moral traditionalism</i>				
TRAD1: The world is always changing and we should adjust our view of moral behavior to those changes. <sup>ab</sup>	2.99	0.05	2.70	0.03
TRAD2: The newer lifestyles are contributing to the breakdown of our society.	3.53	0.04	3.50	0.03
TRAD3: We should be more tolerant of people who choose to live according to their own moral standards, even if they are very different from our own. <sup>a</sup>	3.60	0.04	3.35	0.03
TRAD4: This country would have many fewer problems if there were more emphasis on traditional family ties. <sup>a</sup>	3.96	0.04	3.74	0.03
<i>Position of blacks in society</i>				
BLAC1: Irish, Italians, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors. <sup>cb</sup>	3.82	0.04	3.76	0.03
BLAC2: Generations of slavery and discrimination have created conditions that make it difficult for blacks to work their way out of the lower class. <sup>ab</sup>	2.90	0.04	2.58	0.03
BLAC3: Over the past few years, blacks have gotten less than they deserve. <sup>ab</sup>	2.43	0.04	2.35	0.03
BLAC4: It's really a matter of some people not trying hard enough; if blacks would only try harder they could be just as well off as whites. <sup>b</sup>	3.36	0.04	3.40	0.03
<i>Equalitarianism</i>				
EQUA1: Our society should do whatever is necessary to make sure that everyone has an equal opportunity to succeed. <sup>ab</sup>	4.28	0.03	3.93	0.02
EQUA2: We have gone too far in pushing equal rights in this country. <sup>ab</sup>	2.81	0.04	2.94	0.03
EQUA3: One of the big problems in this country is that we don't give everyone an equal chance. <sup>ab</sup>	3.24	0.04	3.07	0.03
EQUA4: This country would be better off if we worried less about how equal people are. <sup>b</sup>	3.05	0.04	3.02	0.03
EQUA5: It is not really that big a problem if some people have more of a chance in life than others. <sup>ab</sup>	2.60	0.04	2.68	0.02
EQUA6: If people were treated more equally in this country we would have many fewer problems. <sup>ab</sup>	3.62	0.04	3.39	0.02

<sup>a</sup> Mean difference between face-to-face and Web is statistically significant at  $\alpha=0.05$ .

<sup>c</sup> Mean difference between face-to-face and Web is statistically significant at  $\alpha=0.1$ .

<sup>b</sup> Standard error difference between face-to-face and Web is statistically significant at  $\alpha=0.05$ .

The mode is dummy coded as 1=face-to-face and 0=Web. The race/ethnicity of respondent variable contains four categories: non-Hispanic white, non-Hispanic black, Hispanic, and other non-Hispanic. Considering the small frequency (144 face-to-face and 225 Web) and lack of information about the “other non-Hispanic” race group, it is excluded from further analyses. In both face-to-face and Web surveys, one of two interview languages, English or Spanish, was used at the respondent’s choice. As a result, 163 interviews were completed in Spanish, most of which were completed in the Web mode (143 cases). Although the interview language is of theoretical interest when examining response style, the small sample size in the face-to-face mode prevents us from conducting statistically sound analysis. Finally, race/ethnicity is categorized into three groups: non-Hispanic white (reference group), non-Hispanic black, and Hispanic. Demographic variables of the respondents are included in the analysis model as control variables. Sex is a dichotomous variable with male as the reference group. The education variable contains three levels: high school or less, some college, and college or more (reference). In the public data file, the household income contains 28 categories, ranging from under \$5000 to \$250,000 and higher. The respondent’s age contains 6 categories: under 29, 30-39, 40-49, 50-59, 60-69, and 70 and over (reference).

The demographic variables are listed in Table 2.2. The sample for these two modes came from two independent frames, both of which are representative of the whole population after weighting adjustments. Table 2.2 shows that the weighted distributions of demographic variables are indeed not significantly different across the two modes.

Table 2.2. Descriptive data for independent and control variables, 2012 American National Election Studies.

	Face-to-face		Web		
	%	S.E.	%	S.E.	
<i>Race/Ethnicity</i>					
White (non-Hispanic)	76.96	1.48	77.82	1.01	$\chi^2=0.318$ p=0.853
Black (non-Hispanic)	13.84	1.21	13.08	0.84	
Hispanic	9.2	0.89	9.1	0.71	
<i>Sex</i>					
Female	52.01	1.53	52.1	1.06	$\chi^2=0.002$ p=0.961
<i>Education</i>					
High school or less	40.6	1.51	40.11	1.08	$\chi^2=0.073$ p=0.964
Some post-high school	30.12	1.37	30.4	0.96	
Bachelor's and above	29.28	1.41	29.49	0.93	
<i>Age</i>					
<30	21.2	1.29	21.02	1.01	$\chi^2=4.807$ p=0.440
30-39	15.57	1.04	14.98	0.78	
40-49	17.28	1.18	16.71	0.78	
50-59	18.88	1.18	20.1	0.79	
60-69	14.28	1.12	16.27	0.69	
>=70	12.79	1.09	10.92	0.61	
	Mean (95% C.I.)		Mean (95% C.I.)		
Household income	14.8 (14.43, 15.18)		14.25 (13.99, 14.51)		

## 2.4 Analytic approach

Latent class analysis (LCA) is used to analyze the mode effects on response style. This analytical approach was first proposed by Moors (2003) for examining ERS (see also Moors 2008; Kankaraš and Moors 2009). Later, Morren, Gelissen, and Vermunt (2011) proposed a nominal-ordinal hybrid model that was modified by Kieruj and Moors (2013; see also Moors, Kieruj, and Vermunt 2014) in order to estimate ARS and ERS simultaneously. This study adopts this hybrid model. Specifically, all latent variables are analyzed as ordered equidistant discrete variables, and Likert scales are treated as both nominal and ordinal categorical variables. They are treated as ordinal variables when estimating the acquiescent latent class factor because, as the level of acquiescence increases, the respondent is more likely to choose a higher level of the response option (towards agree) than a lower level option (towards disagree) on the Likert scale. When estimating the extreme latent class variable, Likert scales are treated as nominal in order to capture the non-monotone (U-shape) relationship: respondents with higher levels of ERS are more likely to choose the two endpoints than the response options in between. This can be written as the following linear model for the logit of responding in category  $c+1$  versus  $c$ :

$$\log \frac{P(Y_{ij} = c + 1 | F1_i, F2_i, F3_i, E_i, A_i)}{P(Y_{ij} = c | F1_i, F2_i, F3_i, E_i, A_i)}$$

$$= (\beta_{0jc+1} - \beta_{0jc}) + \beta_{1j_1} F1_i + \beta_{2j_2} F2_i + \beta_{3j_3} F3_i + (\beta_{4jc+1} - \beta_{4jc}) E_i + \beta_{5j} A_i$$

Where  $Y_{ij}$  denotes the response of respondent  $i$  to Likert scale question  $j$ ,  $i=1, \dots, I$ ,  $j=1, \dots, 14$ ;

$F1_i$  denotes the “moral traditionalism” latent class variable;

$F2_i$  denotes the “position of black in society” latent class variable;

$F3_i$  denotes the “equalitarianism” latent class variable;



$E_i$  denotes the extreme response style latent class variable (ERS);

$A_i$  denotes the acquiescence response style latent class variable (ARS);

$\beta_{1j_1}$  denotes the effects on the adjacent category logit for the “moral traditionalism” latent class variable,  $j_1=1, 2, 3, \text{ or } 4$ ;

$\beta_{2j_2}$  denotes the effects on the adjacent category logit for the “position of black in society” latent class variable,  $j_2=5, 6, 7, \text{ or } 8$ ;

$\beta_{3j_3}$  denotes the effects on the adjacent category logit for the “equalitarianism” latent class variable,  $j_3=9, 10, 11, 12, 13, \text{ or } 14$ ;

$\beta_{4j_{c+1}} - \beta_{4j_c}$  denotes the nonmonotone (U-shape) relationship between extreme response style latent class variable and the Likert scale;

$\beta_{5j}$  denotes the effects on the adjacent category logit for the acquiescence latent class variable;

and

$c$  denotes the number of response options,  $c=1, 2, 3 \text{ or } 4$ .

It is clear from this equation that the model will output one regression coefficient for each content latent class variable and for the ARS latent class variable, and it will output five regression coefficients for the ERS latent class variable since effect coding is used. In the model specification, the items were allowed to load only on their corresponding content latent class variables. For ARS and ERS latent class variables, all 14 items loaded on these two style latent class variables because, theoretically, response styles should affect all items regardless of the content. The three content latent class variables were allowed to be correlated with each other, but the two style latent class variables were specified as uncorrelated with the content latent class variables and uncorrelated with each other. All five latent class variables were regressed on the

mode and on other covariates, including race/ethnicity, sex, age, education level, and household income, in order to test the mode effect on response style latent variables. Figure 2.1 illustrates the full analysis model. For clarity of presentation, the plot only shows the response style variables regressed on the covariates. In the analysis model, three content variables are regressed on the covariates as well. All models were estimated using Latent Gold 5.0.

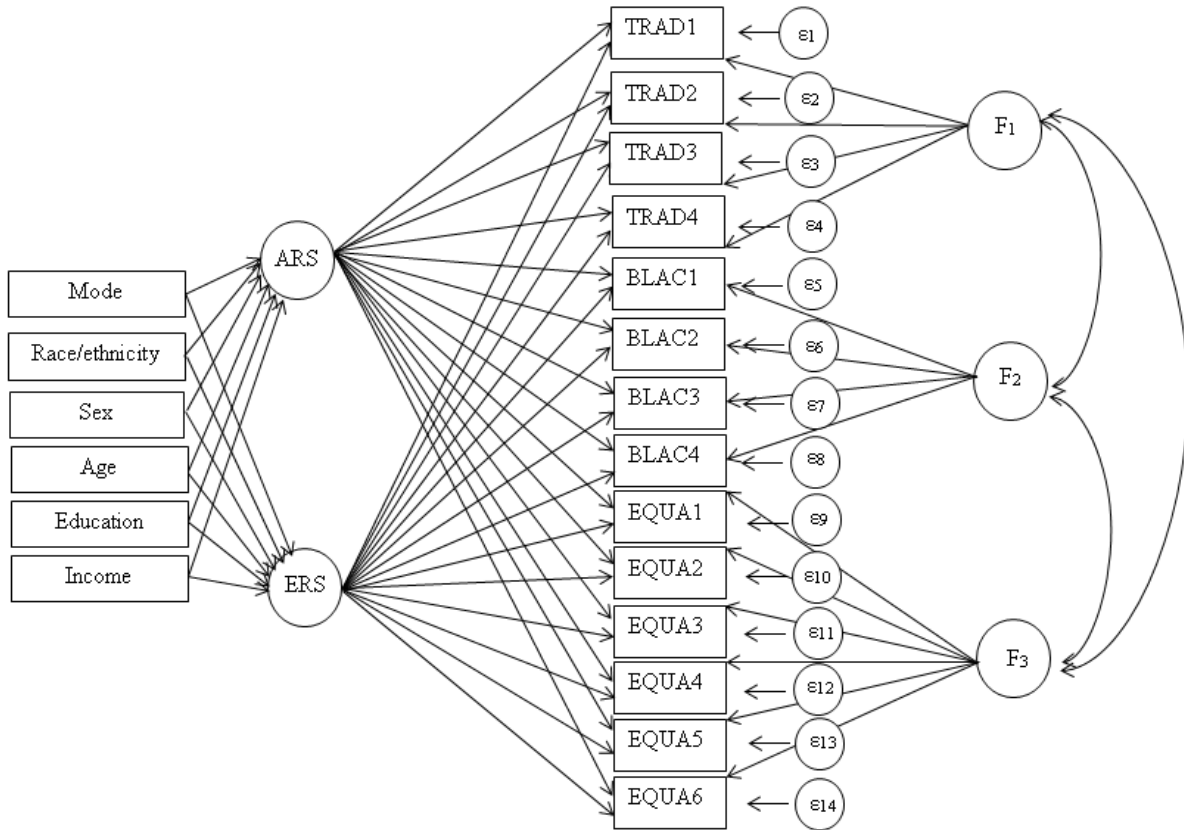


Figure 2.1. Latent class analysis model of acquiescent response style (ARS), extreme response style (ERS), and content latent class variables (F1, F2, F3), with covariates.

## 2.5 Model Specification

In order to find the best fit model for the data, we follow the following steps:

- 1) Determine whether adding the response style latent class variables improve model fit,
- 2) Decide the number of classes for the latent class variables, and
- 3) Examine whether the equality constraint on latent class variable coefficients improves the model fit.

In this study, we use the Bayesian information criterion (BIC) to compare models. The BIC reflects the model fit based on likelihood function and simultaneously penalizes less parsimonious models (i.e., models with more parameters). A more complex model tends to have a better fit but is less likely to be replicated and is sometimes uninterpretable. A smaller BIC value indicates a balance between model fit and model complexity.

The first step of the analysis is to determine whether the data really reflect response style: ARS, ERS, or both. In other words, we determine whether adding the response style latent class variables improves the model fit, in comparison to the model with only content latent class variables. Table 2.3 presents the model fit statistics of four different models: content only; content and ARS; content and ERS; and content, ARS, and ERS. It is clear that the BIC value decreases as the response style latent class variables are added to the model (Model 1 vs. 2 vs. 3) and that Model 4, with three content variables and two response style variables, has the smallest BIC value, which indicates the best model fit. Also, the model with ERS (Model 3) has a smaller BIC value than the one with ARS (Model 2), which suggests that ERS is a more important response style variables than ARS. (Moors, Kieruj, and Vermunt 2014) report similar findings.

Table 2.3. Model Fit Statistics, 2012 American National Election Studies.

	BIC	No. of parameters
Model 1: Content only (2-class)	152621	112
Model 2: Content+ARS (2-class)	145829	139
Model 3: Consent+ERS (2-class)	147837	184
Model 4: Content+ARS+ERS (2-class)	142416	208
Model 5: Content+ARS+ERS (3-class)	139761	213
Model 6: Content+ARS+ERS (4-class)	138905	218
Model 6a: Equality on all latent variables	156832	140
<b>Model 6b: Equality on style latent variables</b>	<b>139085</b>	<b>153</b>

After demonstrating that the inclusion of response style latent class variables indeed improves the model fit, we move on to determine the number of classes for the latent class variables. Recall that the content and style latent class variables are discrete ordinal variables with equidistance between classes. Table 2.3 summarizes the results from estimating 2-, 3- and 4-class models on all five latent class variables using the content+ARS+ERS model. The change in BIC value from 2-class (Model 4) to 3-class (Model 5) is substantial, and the rate of change decreases between 3-class (Model 5) and 4-class (Model 6) models. We also try to estimate the 5-class model, but due to model complexity and difficulty interpreting results, the result from 5-class model is not included. Considering these options, the 4-class model with three content variables and two style variables is the best intermediate model.

The next step is to examine whether adopting a more parsimonious model can further reduce the model complexity. One approach proposed by Billiet and McClendon (2000; see also Billiet and Davidov 2008) in their confirmatory factor analysis is to constrain the factor loadings on the style factor such that they are the same across all items. This approach is later used in the latent class analysis framework to constrain the coefficients of response style latent class variables to be identical across items (Kieruj and Moors 2010; Kieruj and Moors 2013; Moors, Kieruj, and Vermunt 2014).

We test three types of models: no constraints (Model 6, Table 2.3); equality constraints on both the content and style latent class variables (Model 6a); and equality constraints on the style latent class variables (Model 6b). After setting the equality constraints on all latent class variables, the model fit becomes even worse than for the model with no constraint. This suggests that the items are measuring different aspects of the substantive latent constructs and that the relationships between items and content latent class variables are of various strengths. Model 6b

has constraints on the style latent class variables only. The BIC value is only slightly larger than in the first model, but it has 65 fewer parameters, a much more parsimonious model. Besides, this approach is theoretically more meaningful because response style is a more stable personality trait than is attitude toward the question content (Naemi, Beal, and Payne 2009; Kieruj and Moors 2013; Weijters, Geuens, and Schillewaert 2010). Considering these options, the third model is preferred over the other two alternatives. At this point, the aforementioned analyses have led to the final model for examining the mode effect, which is a latent class analysis model with three content latent variables and two style variables, each with four classes and equal coefficients of both style variables across all items. The covariates are included in all models estimated so far.

## **2.6 Results for mode effect**

This section begins by presenting the estimated coefficients of the two style variables on Likert scale items. Then, we will present the estimated coefficients of covariates on the two response styles in order to examine the mode effect and the interaction effect between mode and race/ethnicity.

Table 2.4 presents the regression coefficients of ERS and ARS on the 16 Likert items. Since the coefficients are constrained to be the same across all items on the style latent class variables, there is only one set of coefficients for each. The estimated coefficients sum up to zero because of the effect coding of the Likert scale questions. For ERS, the coefficients can be interpreted in the same way as a multinomial logistic regression where the predictor is an equidistance ordered variable with four categories. The two endpoints have negative coefficients while the middle three options have positive coefficients, suggesting that respondents at the

lower level of this latent class variable tend to choose the endpoints of the scale more often than do respondents at the higher level of this latent class variable. In other words, a lower level of the latent class indicates pro-ERS while a higher level of the latent class represents avoid-ERS. To better understand the relationship between the ERS latent class variables and the response option choice, we can convert the coefficient (log odds) into an odds ratio. For example, as the latent ERS moves down one level, the odds of choosing “disagree strongly” over “disagree somewhat” change by a factor of 146.94 ( $=\exp(1.32+3.67)$ ), and the odds of choosing “agree strongly” over “agree somewhat” change by a factor of 192.48 ( $=\exp(1.18+4.08)$ ).

The estimated coefficient for ARS is the adjacent-category logit, and there is only one coefficient estimated due to the equality constraint. A positive coefficient indicates that, as the level of the latent ARS variable moves up, the respondent becomes more likely to choose the more positive response option. More specifically, as the level of ARS moves up one level, the respondent is 3.97 ( $=\exp(1.38)$ ) times more likely to choose category  $c+1$  over category  $c$ . Since the items were coded as “disagree strongly” (1) to “agree strongly” (5), this latent variable is indeed acquiescence.



Table 2.4. Estimated Regression Coefficients and Standard Errors of ERS and ARS on the Likert Scale Items, 2012 American National Election Studies.

Response style	Response option	$\hat{\beta}$	S.E.	
ERS	Disagree strongly	-3.67	0.18	***
	Disagree somewhat	1.32	0.13	***
	Neither agree nor disagree	5.24	0.26	***
	Agree somewhat	1.18	0.12	***
	Agree strongly	-4.08	0.19	***
ARS		1.38	0.07	***

\*\*\*  
 $p < .001$ .

The mode effect on the response style is presented in Table 2.5. The style latent class variables are dependent variables in this case. Since both style latent class variables are 4-level equidistant ordinal variables, the coefficients can be interpreted in the same way as an ordinal logistic regression. As a lower level of ERS latent class variable indicates pro-ERS and a higher level of ERS indicates avoid-ERS, a negative coefficient indicates a positive association between face-to-face and ERS. As for ARS, a positive coefficient indicates a positive association between face-to-face and ARS because a higher level of ARS latent class variable represents pro-acquiescence. For each response style, we fit three models to determine the mode effect. The results in Table 2.5 are weighted coefficients. Model 1 includes mode and race/ethnicity only. The result shows that face-to-face interviews are associated with higher instances of ARS and ERS than are Web interviews ( $\hat{\beta}=-2.60, p<.001$ ;  $\hat{\beta}=1.27, p<.001$ ). Since the literature has shown that response styles differ by respondents' demographics, in Model 2, we include several key demographic variables as covariates. Exactly the same pattern emerges: the face-to-face interview is associated with more ARS ( $\hat{\beta}=-2.68, p<.001$ ) and ERS ( $\hat{\beta}=1.23, p<.001$ ).

In order to test whether or not the mode effects on response styles are similar across the three racial/ethnic groups, we fit interaction models in which the mode interacts with the race/ethnicity of the respondents. Model 3 in Table 2.5 presents the regression coefficients for the interaction model for both ARS and ERS. The results lend clear support for the differential mode effect on ERS but not ARS, as the interaction coefficient of mode by black is statistically significant ( $\hat{\beta}=-3.49, p<.001$ ) for ERS. To facilitate the interpretation of the interaction effect, we plot the predicted log odds of ERS from the interaction model. A negative value indicates more ERS while a positive value indicates less ERS. We set control variables at constant and only vary the mode and race/ethnicity to predict the log odds of ERS. Figure 2.2 shows that, although the

mode effects on ERS are similar between Hispanic and non-Hispanic white respondents, the mode effect is larger for non-Hispanic black respondents. When interviewed face-to-face, black respondents are more likely to use extreme responses than are the other two groups, whereas in the Web survey, the use of ERS is similar across the three groups.

Table 2.5. Estimated Regression Coefficients (Log Odds) and Standard Errors of Mode, Race, Interaction, and Control Variables on Extreme Response Style (ERS) and Acquiescent Response Style (ARS), 2012 American National Election Studies.

	Model 1				Model 2				Model 3			
	ERS		ARS		ERS		ARS		ERS		ARS	
	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.
Face-to-face (vs. Web)	-2.60	0.31	1.27	0.29	-2.68	0.31	1.23	0.36	-2.23	0.33	1.44	0.47
Non-Hispanic black	-0.20	0.47	2.52	0.35	-0.56	0.48	2.5	0.46	-0.04	0.52	2.56	0.52
Hispanic	1.07	0.40	2.51	0.38	0.41	0.44	2.56	0.53	0.57	0.5	2.93	0.61
Non-Hispanic white	(Reference)											
Face-to-face × black									-3.49	0.95	-0.26	0.73
Hispanic									-1.03	0.9	-1.33	0.88
Female (vs. male)					0.1	0.23	-0.76	0.32	0.1	0.23	-0.75	0.32
<30					1.95	0.48	-1.74	0.55	1.97	0.49	-1.75	0.55
30-39					2.04	0.46	-1.97	0.58	2.04	0.46	-1.95	0.57
40-49					1.04	0.45	-2.14	0.55	1.05	0.45	-2.13	0.55
50-59					0.71	0.42	-2.37	0.47	0.7	0.42	-2.36	0.46
60-69					0.62	0.41	-1.27	0.47	0.6	0.42	-1.25	0.46
>=70	(Reference)											
High school or less					0.8	0.3	4.06	0.52	0.8	0.31	4.02	0.54
Some post-high school					-0.01	0.28	2.32	0.47	0	0.28	2.29	0.48
College and more	(Reference)											
Household income					-0.03	0.02			-0.03	0.02		

\*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ , #  $p < .1$ .

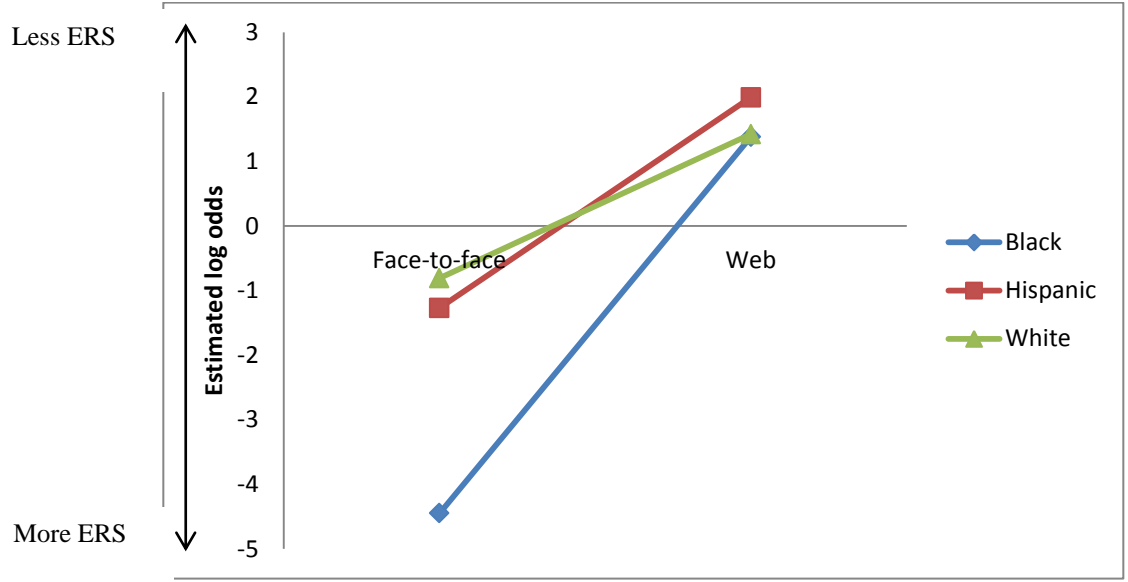


Figure 2.2. Interaction effects between mode and race/ethnicity on extreme response style, for a 60-69-year-old male respondent with high school or lower education, and whose household income is set at the sample mean.

## 2.7 Discussion

Although there is abundant literature on both data collection mode and response style, researchers rarely examine these two factors together, let alone the interaction effect of mode and race/ethnicity on response styles. Using the 2012 ANES data, this study examined the mode effect on acquiescent and extreme response styles. More specifically, the 2012 ANES has two independent samples: one for face-to-face surveys and another for Web surveys. This provides a unique opportunity to study the mode effect on response style using national representative samples. Using the latent class analysis approach, we discovered three main findings: 1) both ARS and ERS are present in the responses to three sets of Likert scale questions asking about “Moral traditionalism,” “Position of blacks in society,” and “Equalitarianism,” and of the two, ERS is the more dominant style variable; 2) face-to-face respondents demonstrate more ARS and ERS than Web respondents; and 3) the mode effect is larger for black respondents than for white and Hispanic respondents.

The larger reduction of BIC values of ERS than ARS suggests that ERS is the more influential response style in this set of questions. Previous studies have shown similar findings from other surveys questions (Kieruj & Moors 2013; Moors et al. 2014). The reason for the weaker ARS, however, is not clear, providing an opportunity for future research. One thing we do know is that controlling for ARS improves the construct equivalence in cross-cultural study (Welkenhuysen-Gybels, Billiet, and Cambré 2003).

The finding that face-to-face interviews elicit more ERS and ARS than do Web surveys confirms our theoretical expectations of both response styles. The desire to manage impressions in the presence of interviewers motivates the respondent to choose agreeable responses. In contrast, when taking a Web survey, the motivation of self-presentation is lessened by the

anonymity of the survey mode. Thus, we do not see as much acquiescence in the Web mode as we do in the face-to-face interview mode. A similar argument applies to ERS. Giving extreme answers is sometimes regarded as more informative and more opinionated in comparison to giving more moderate answers. Respondents, therefore, choose extreme answers in face-to-face interviews in order to portray themselves as genuine and to demonstrate to the interviewer that they are taking the task seriously. In taking the Web survey, respondents feel less pressure to give opinionated responses and therefore provide more moderate answers. Our findings are in line with the mode study by Weijters, Schillewaert, and Geuens (2008), which found a greater tendency toward ARS and ERS in telephone rather than online surveys. Both face-to-face and telephone surveys involve interviewers, but they interact with respondents more directly in the former mode than in the latter. Although this has not been studied, we speculate that more ARS and ERS would be observed in face-to-face interviews than in telephone interviews.

This study detected the interaction effect between mode and race/ethnicity on ERS for the first time. We found that the mode has a larger effect among black respondents than Hispanic and white respondents. It is not entirely clear to us why there is an absence of interaction effects on ARS. One possible explanation is the overall ARS is less salient than ERS in this set of questions. Therefore, there is less variation between racial/ethnic groups in terms of mode effect on ARS. Also, the reason for the lack of interaction effect between mode and Hispanic is not entirely clear to us. As Figure 2 shows, the coefficients for Hispanic and white respondents are similar in both modes. One possible explanation is the relative high level of acculturation of the Hispanic respondents since ANES respondents are all U.S. citizens. Also, we excluded the Spanish interviews from the analysis. We would expect a different pattern if we have included the Spanish interviews as those Spanish-speaking respondents should be generally less

acculturated and carry higher level of Hispanic culture. However, the small sample size of Spanish interview prevents us from drawing any conclusion.

This study advances our knowledge of the motivations behind both ARS and ERS. Although we study the mode effects on response style, what is actually revealed is a social and psychological consequence. The unique interactions present in survey situations provide us with an opportunity to empirically test the theories of ARS and ERS. The greater incidence of response style bias in the face-to-face survey as compared to the Web survey supports the notion that response style is driven by the interactive dynamics between the dyad of interviewer and respondent. When one party is removed from the interaction (i.e., the interviewer from the Web survey), a different set of cognitive and social norms, such as the survey response process, will emerge and influence respondents' behavior.

This study raises perhaps more questions than it answers. First, we only know that the Web survey mode reduces ARS and ERS for the three topics we examined. Whether the findings will be replicable using other questions is unknown. Second, what is the implication of data collection mode? Do these findings suggest that Web surveys provide better data quality since they lead to decreased response style bias? We know this is not a simple yes/no question. Our findings provide some evidence that the measurement capability of Likert scales is better in Web surveys than in face-to-face interviews, but we cannot carry these findings too far. A variety of factors must be considered when choosing the data collection mode for a particular survey. Third, are these findings applicable to other racial/ethnic groups beyond the ones included in this study? Some groups have distinct response styles reflecting unique cultural norms that are quite different from those of the three groups that we examined. Various interview modes may have different impacts on other groups, an important consideration given the increasing interest in



cross-cultural survey research. Finding a mode that reduces response style bias across different cultural groups will improve the comparability of survey data.

## References

- Aday, Lu Ann, Grace Y. Chiu, and Ronald Andersen. 1980. "Methodological Issues in Health Care Surveys of the Spanish Heritage Population." *American Journal of Public Health* 70 (4): 367–74.
- Bachman, Jerald G., and Patrick M. O'Malley. 1984a. "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles." *Public Opinion Quarterly* 48 (2): 491–509.
- . 1984b. "Black-White Differences in Self-Esteem: Are They Affected by Response Styles?" *American Journal of Sociology* 90 (3): 624–39.
- Bachman, Jerald G., and Patrick M. O'Malley. 2013. "Response Styles Revisited: Racial/Ethnic And Gender Differences In Extreme Responding." Accessed December 16. <http://monitoringthefuture.org/pubs/occpapers/occ72.pdf>.
- Baumgartner, Hans, and Jan-Benedict EM Steenkamp. 2001. "Response Styles in Marketing Research: A Cross-National Investigation." *Journal of Marketing Research* 38 (2): 143–56.
- Billiet, Jaak B., and Eldad Davidov. 2008. "Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design." *Sociological Methods & Research* 36 (4): 542–62.
- Billiet, Jaak B., and McKee J. McClendon. 2000. "Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items." *Structural Equation Modeling: A Multidisciplinary Journal* 7 (4): 608–28.
- Bolt, Daniel M., and Joseph R. Newton. 2011. "Multiscale Measurement of Extreme Response Style." *Educational and Psychological Measurement* 71 (5): 814–33.

- Carr, Leslie G. 1971. "The Srole Items and Acquiescence." *American Sociological Review* 36 (2): 287–93.
- Clarke III, Irvine. 2001. "Extreme Response Style in Cross-Cultural Research." *International Marketing Review* 18 (3): 301–24.
- Couch, Arthur, and Kenneth Keniston. 1960. "Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable." *The Journal of Abnormal and Social Psychology* 60 (2): 151–74.
- Cronbach, Lee J. 1942. "Studies of Acquiescence as a Factor in the True-False Test." *Journal of Educational Psychology* 33 (6): 401–15.
- Davis, R. E., K. Resnicow, and M. P. Couper. 2010. "Survey Response Styles, Acculturation, and Culture Among a Sample of Mexican American Adults." *Journal of Cross-Cultural Psychology* 42 (7): 1219–36.
- Gibbons, Judith L., Jennifer A. Zellner, and David J. Rudek. 1999. "Effects of Language and Meaningfulness on the Use of Extreme Response Style by Spanish-English Bilinguals." *Cross-Cultural Research* 33 (4): 369–81.
- Gilbert, Daniel T. 1991. "How Mental Systems Believe." *American Psychologist* 46 (2): 107–19.
- Hui, C. Harry, and Harry C. Triandis. 1989. "Effects of Culture and Response Format on Extreme Response Style." *Journal of Cross-Cultural Psychology* 20 (3): 296–309.
- Jordan, Lawrence A., Alfred C. Marcus, and Leo G. Reeder. 1980. "Response Styles in Telephone and Household Interviewing: A Field Experiment." *Public Opinion Quarterly* 44 (2): 210–22.

- Kankaraš, Miloš, and Guy Moors. 2009. "Measurement Equivalence in Solidarity Attitudes in Europe Insights from a Multiple-Group Latent-Class Factor Approach." *International Sociology* 24 (4): 557–79.
- Kieruj, Natalia D., and Guy Moors. 2010. "Variations in Response Style Behavior by Response Scale Format in Attitude Research." *International Journal of Public Opinion Research* 22 (3): 320–42.
- . 2013. "Response Style Behavior: Question Format Dependent or Personal Style?" *Quality & Quantity* 47 (1): 193–211.
- Kiesler, Sara, and Lee S. Sproull. 1986. "Response Effects in the Electronic Survey." *Public Opinion Quarterly* 50 (3): 402–13.
- Knowles, Eric S., and Christopher A. Condon. 1999. "Why People Say Yes: A Dual-Process Theory of Acquiescence." *Journal of Personality and Social Psychology* 77 (2): 379–86.
- Leary, Mark R., and Robin M. Kowalski. 1990. "Impression Management: A Literature Review and Two-Component Model." *Psychological Bulletin* 107 (1): 34–47.
- Madsen, William. (1964), *Mexican-Americans of South Texas*, New York: Holt, Rinehart and Winston.
- Marin, Gerardo, Raymond J. Gamba, and Barbara V. Marin. 1992. "Extreme Response Style and Acquiescence among Hispanics The Role of Acculturation and Education." *Journal of Cross-Cultural Psychology* 23 (4): 498–509.
- Marín, Gerardo, & Barbara VanOss Marín. (1991), *Research with Hispanic Populations*, California: SAGE Publications.

- Moors, Guy. 2003. "Diagnosing Response Style Behavior by Means of a Latent-Class Factor Approach. Socio-Demographic Correlates of Gender Role Attitudes and Perceptions of Ethnic Discrimination Reexamined." *Quality and Quantity* 37 (3): 277–302.
- . 2008. "Exploring the Effect of a Middle Response Category on Response Style in Attitude Measurement." *Quality & Quantity* 42 (6): 779–94.
- . 2010. "Ranking the Ratings: A Latent-Class Regression Model to Control for Overall Agreement in Opinion Research." *International Journal of Public Opinion Research* 22 (1): 93–119.
- Moors, Guy, Natalia D. Kieruj, & Jeroen K. Vermunt. (2014), "The Effect of Labeling and Numbering of Response Scales on the Likelihood of Response Bias," *Sociological Methodology* [online]. Available at <http://smx.sagepub.com/content/early/2014/01/13/0081175013516114.abstract>
- Morren, Meike, John P.T.M. Gelissen, and Jeroen K. Vermunt. 2011. "Dealing With Extreme Response Style In Cross-Cultural Research: A Restricted Latent Class Factor Analysis Approach: Extreme Response Style In Cross-Cultural Research." *Sociological Methodology* 41 (1): 13–47.
- Naemi, Bobby D., Daniel J. Beal, and Stephanie C. Payne. 2009. "Personality Predictors of Extreme Response Style." *Journal of Personality* 77 (1): 261–86.
- Paulhus, Delroy L. 1991. "Measurement and Control of Response Bias." In *Measures of Personality and Social Psychological Attitudes. Measures of Social Psychological Attitudes, Vol. 1*, edited by John P. Robinson, Phillip R. Shaver, Lawrence S. Wrightsman, 17-59, San Diego, CA: Academic Press.

- Ross, Catherine E., and John Mirowsky. 1984. "Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health." *Journal of Health and Social Behavior* 25 (2): 189–97.
- Schlenker, Barry R. (1980), *Impression management: The self-concept, social identity, and interpersonal relations*, California: Brooks/Cole.
- Thomas, Troy D., Koen Abts, and Patrick Vander Weyden. 2014. "Response Styles and the Rural–Urban Divide." *Educational and Psychological Measurement* 74 (1): 97–115.
- Warnecke, Richard B., Timothy P. Johnson, Noel Chávez, Seymour Sudman, Diane P. O’rourke, Loretta Lacey, and John Horm. 1997. "Improving Question Wording in Surveys of Culturally Diverse Populations." *Annals of Epidemiology* 7 (5): 334–42.
- Weijters, Bert, Maggie Geuens, and Niels Schillewaert. 2010. "The Individual Consistency of Acquiescence and Extreme Response Style in Self-Report Questionnaires." *Applied Psychological Measurement* 34 (2): 105–21.
- Weijters, Bert, Niels Schillewaert, and Maggie Geuens. 2008. "Assessing Response Styles across Modes of Data Collection." *Journal of the Academy of Marketing Science* 36 (3): 409–22.
- Welkenhuysen-Gybels, Jerry, Jaak Billiet, and Bart Cambré. 2003. "Adjustment for Acquiescence in the Assessment of the Construct Equivalence of Likert-Type Score Items." *Journal of Cross-Cultural Psychology* 34 (6): 702–22.
- Yang, Yongwei, Janet A. Harkness, Tzu-Yun Chin, and Ana Villar. 2010. "Response Styles and Culture." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith, 203–223, Hoboken, New Jersey: Wiley.

## Chapter 3

### Comparing Extreme Response Styles between Agree-Disagree and Item Specific Scales

#### 3.1 Introduction

When designing survey questions to measure opinions and attitudes, one of the first ideas that comes to a researcher's mind is probably a Likert scale. As Bradburn, Sudman, and Wansink (2004) point out, attitudes comprise three major components, namely cognitive, evaluative, and behavioral. The cognitive component is one's beliefs about the target object. The evaluative component is one's evaluation of the object. The behavioral component is the connection between attitude and behavior. A Likert scale can be used to measure both the respondent's evaluation (agree vs. disagree) of the rating object and the strength of the evaluation (strongly vs. somewhat). That is, it measures the intensity of one's attitude toward an object (Krosnick and Abelson 1992). Furthermore, it may seem easy to design Likert scales since researchers have only to create questions describing the objects they want to measure without varying the response options. These advantages make Likert scales very popular among social scientists and marketing researchers. For example, the Marketing Scales Handbook and other similar references provide numerous citations to publications using Likert scales (Bearden and Netemeyer 1999; Bruner, Hensel, and James 2001).

Although it is popular in many fields, the data quality yielded by Likert scales in the agree-disagree (A/D) format is controversial among researchers (for examples, see Clark and

Clark 1977; Fowler and Cosenza 2008; Billiet and McClendon 2000; DeVellis 2011; Revilla, Saris, and Krosnick 2013). Some survey methodologists have even argued that “researchers will have more reliable, valid, and interpretable data if they avoid the agree-disagree question form” (Fowler 2013). One of the concerns about the A/D scale is the acquiescent response style bias, defined as the tendency to choose “agree” or “yes” responses more frequently than other response options (Baumgartner and Steenkamp 2001). Recognizing the measurement issues involved with the A/D format, researchers have developed other question formats to measure attitudes. In this study, we will focus on one of these questions types, the item specific (IS) question. The IS question asks the respondent to choose an option that best describes his or her attitude. Distinct from the A/D format, which offers the same response options for all questions, an IS question uses response options that are specific to the question contents (see Appendix A for examples of IS questions).

As we will review below, some efforts have been taken to compare the A/D and IS rating scales. However, we notice that although some research has examined extreme response style (ERS) using the A/D scale, no work has yet examined the ERS in the context of the IS scale. ERS is another common measurement bias in rating scales and refers to the tendency to select the two endpoints of a response scale more frequently than the intermediate ones (Paulhus 1991). This study will fill that gap by examining the ERS of A/D and IS scales using experimental data in a panel study. We also try to answer two research questions. First, do respondents show different patterns of ERS when using an IS scale compared to an A/D scale? Since ERS is regarded as a type of measurement error, a scale that yields lower levels of ERS bias is preferred. Second, is ERS a stable trait over time or it is a phenomenon mainly driven by survey design features, such as question formats?



### **3.2 Literature review**

The literature presents a number of problems associated with questions in the A/D format. First of all, to design an effective scale, researchers must make a series of decisions about the format of the question, including the number of response options (Alwin and Krosnick 1991; Revilla, Saris, and Krosnick 2013), whether to label the whole scale or only the endpoints (Krosnick and Berent 1993), whether to use numeric or verbal labels (Alwin and Krosnick 1991), and whether to provide middle options (O’Muircheartaigh, Krosnick, and Helic 2000). These decisions can potentially influence the reliability as well as the validity of the data. (For a more comprehensive review, see Krosnick and Presser 2010.) Second, the choice of the rating object is arbitrary, and different wordings are subject to differential measurement errors (DeVellis 2011). The common practice is to phrase the statement such that it aligns unambiguously with one end of a continuum, although the endpoint of a latent continuum may also be arbitrary (Fowler 1995). Third, the cognitive process of answering an A/D question is different from that used to answer most other survey questions. The cognitive burden is higher than if the attitude is approached more directly through IS format. Rather than the well-known four-step cognitive process for survey response (Tourangeau, Rips, and Rasinski 2000), A/D format requires a more involved set of cognitive steps (Carpenter and Just 1975; Clark and Clark 1977; Trabasso, Rollins, and Shaughnessy 1971). The fourth problem associated with the A/D format is that it is not clear what the distance is between the two adjacent options (Fowler and Cosenza 2008). In analysis, researchers often dichotomize Likert scale items to agree vs. disagree and hence information is lost from the supposedly ordinal scale (Fowler 2013). Last but not least, A/D format suffers from acquiescent response style bias (Schuman and Presser 1981; Billiet and McClendon 2000; Billiet

and Davidov 2008). Such response style can artificially increase the estimated mean and make the responses look more positive than they really are.

Knowing the measurement errors associated with the A/D scale, some survey methodologists prefer to use IS formats over A/D formats because IS formats are “much simpler, direct and informative” (Fowler and Cosenza 2008; see also Converse and Presser 1986; Fowler 1995). In contrast to the A/D format, in the IS format, “the categories used to express the opinion are exactly those answers that we would like to obtain for this item” (Saris et al. 2010). For this reason, it is reasonable to assume the cognitive burden of answering an IS question is lower than that of answering an A/D question. The respondent needs only to judge and map his or her attitude on the response scale without having to map both his or her true attitude and the rating object onto the response continuum in order to answer A/D question.

Empirical studies comparing the measurement quality of A/D and IS formats are relatively limited and the findings are mixed. Most of the studies examine the acquiescent response style and the reliability and validity of the A/D and IS scales. Berkowitz and Wolkon (1964) are among the first to examine these two types of rating scales and discover that the reliability is similar across all scale types. Later, Ray (1979) tests a total of six A/D and IS scales, and correlates the estimates’ external validation data<sup>1</sup>. He reports mixed findings: two types of the A/D scale and one IS scale have higher validity than the other scales. Patient satisfaction rating is another area in which A/D versus IS formats have been studied. Counte (1979) compares patient satisfaction rating scales and finds that IS scale explains more variance than do

---

<sup>1</sup> The external validation data in the Ray (1979) study is the peer ratings of the achievement motivation of the respondents. The validity is operationalized as the correlation between the self-reporting and the peer rating.

the A/D measures. Similarly, Ross, Steward, and Sinacore (1995) evaluate seven patient satisfaction rating scales, three of which are A/D Likert scales. They detect acquiescence bias for all scale types<sup>2</sup>. They also find that the internal consistency of satisfaction measures is higher for respondents who give either no acquiescent responses or a moderate amount, as compared to those with a high level of acquiescent responses. Saris et al. (2010) expand this line of study to an international setting (i.e. 14 European countries) using a multitrait-multimethod approach and demonstrate that, for two out of three questions, the item specific format shows better reliability, validity, and quality than the A/D format<sup>3</sup>. However, the quality advantage for the IS question is not universal across all countries examined. More recently, Cibelli and Callegaro (2011) report the findings from a Google AdWords customer satisfaction survey conducted in five languages; in each language the A/D format elicits more acquiescence bias than does the item specific format.

Previous research has also shown that the A/D Likert scales are susceptible to ERS bias (Greenleaf 1992; Vaerenbergh and Thomas 2012; Weijters, Schillewaert, and Geuens 2008; Morren, Gelissen, and Vermunt 2011). This is true regardless of the design features of the A/D scale, including scale length (5- to 11-points), middle response category, fully labeled vs. end labeled, numerical labels, agreement scale vs. bipolar scale, although the magnitudes are slightly different (Kieruj and Moors 2010; Kieruj and Moors 2013; Moors 2003; Moors, Kieruj, and Vermunt 2014). The presence of extreme response style is stable over time in longitudinal

---

<sup>2</sup> The acquiescent response bias in Ross et al. (1995) is measured through pairs of questions for which the content is the same but the direction of the question wording is reversed.

<sup>3</sup> The validity in Saris et al. (2010) refers to the internal validity as measured by the MTMM model.

studies and consistent across items in the same survey, suggesting that it is a stable personality trait among respondents (Weijters, Geuens, and Schillewaert 2010a; Weijters, Geuens, and Schillewaert 2010b)<sup>4</sup>. It is unknown whether the IS scale also suffers from ERS bias and to what extent compared to the A/D scale. The goal of this study to evaluate and compare the ERS bias between these two scale formats through experimental data.

### **3.3 Data and measures**

This study analyzes the experimental data embedded in the 2012 American National Election Studies (ANES). The questions under study measure the general population's political efficacy. Political efficacy refers to the "feeling that individual political action does have, or can have, an impact upon the political process" (Campbell, Gurin, and Miller 1954, 187). The ANES has measured the general population's political efficacy for decades using the A/D format. Given the importance of this measure, it is critical to achieve a thorough understanding of the measurement of these questions. In 2012, the ANES includes four A/D items to measure individuals' sense of efficacy. The first two items measure one's internal efficacy and the last two items measure one's external efficacy. The internal efficacy questions ask about how people think about their ability to adequately understand and effectively participate in politics, and the

---

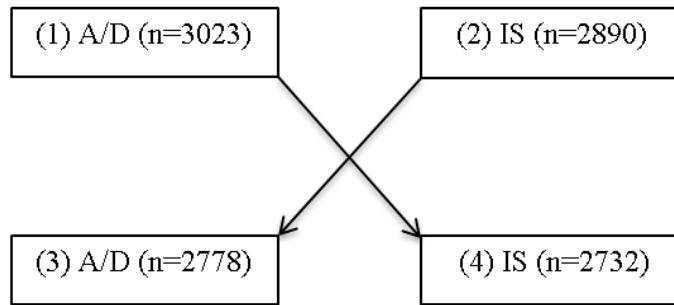
<sup>4</sup> The ERS in Greenleaf (1992) is operationalized as the proportion of the extreme answers to 16 questions with low inter-item correlations. The ERS in Weijters, Schillewaert, and Geuens (2008) is measured through representative indicators response style means and covariance structure (RIRSMACS) method. The ERS in Weijters et al. (2010a) is estimated by the multiple-indicators, multiple-covariates (MIMIC) Model. The ERS in Weijters et al. (2010b) is modeled as a tau-equivalent factor complemented with a time-invariant autoregressive effect.

external efficacy questions ask about people's perceptions of the responsiveness of public officials and government institutions to their demands (Niemi, Craig, and Mattei 1991).

Although there are debates about the concept of political efficacy, these are beyond the scope of this study, and readers can find more details elsewhere (Balch 1974; Morrell 2003; Chamberlain 2012).

Out of the concern that the A/D format would potentially cause acquiescent response bias, the 2012 ANES included four IS items in addition to the A/D items. The IS items were designed to measure the same internal and external efficacy as the A/D items. The two question formats were tested in a between- and within-subject experiment across two waves of interviews. In the pre-election study, the respondents were randomly assigned to one of the two kinds of scales. As a result, there are 3023 respondents for the A/D and 2890 for the IS scale. In the post-election study, the assignment was reversed. The between-subject design allows for examination of the ERS between the two question formats if we treat the pre- and post-surveys as two cross-sectional surveys. The within-subject design permits study of the change in response style between different question formats for the same respondent. The experimental design and sample sizes are illustrated in Figure 3.1. The respondents were interviewed either face-to-face or via the Web, and the response rates were 38% and 2%, respectively. The overall re-interview rates were 94% and 93%, for face-to-face and Web respectively. To rule out the possibility of differential nonresponse bias between the two modes, we examined the demographic variables, including gender, age, race/ethnicity, education level, household income, and marital status. None of them was statistically different between the two modes (Appendix 3.2). Also, we compared the same set of demographic variables between the two rating scale formats within each mode, and they did not differ significantly either.

Pre-election:



Post-election:

Figure 3.1. Experimental design and sample sizes.

### **3.4 Analytical approaches and results**

In this section, we will describe the analytical approaches and report the results for the between-subject and within-subject experiments separately. Similar latent class factor analysis models were used to measure ERS to both the between- and within-subject experiments but the measurement models are different for between- and within-subject experiments.

#### *3.4.1 Between-subject experiment analysis*

We will first examine the distributions of answers to each question between the two question formats using the pre-election interview data. This will show us whether there is any difference between the A/D and IS formats in terms of ERS and whether the difference is driven by any particular question or whether it is a consistent phenomenon across all four questions.

We will then perform latent class factor analysis to examine ERS between the two question formats. In the literature, there is no single accepted statistical procedure for measuring the ERS. We choose this method because it can simultaneously estimate response style and the substantive contents of the scale as different latent class factors. As a result, the estimate of the response style factor is not confounded with the question contents. Billiet and McClendon (2000) initially propose to use confirmatory factor analysis (CFA) to estimate both content factors and the acquiescent response style factor (see also Billiet and Davidov 2008). CFA treats the rating scales as ordinal variables, so it is adequate for measuring acquiescent response style because a monotone relationship is assumed between the response style latent variable and rating scale. In other words, respondents with higher levels of ARS are more likely to choose a category closer to the positive end than the negative end of the scale. When it comes to the ERS, a U-shape relationship is expected between the ERS latent class factor and response items. That is,

respondents with higher levels of ERS are more likely to choose endpoints than the intermediate categories. CFA cannot capture such a relationship.

Moors (2003) extends this approach to the latent class analysis framework and treats the rating scales as nominal variables in order to estimate ERS. Morren, Gelissen, and Vermunt (2011) further extend the model, imposing an ordinal relationship between the rating scales and content latent class factors but for ERS, maintaining a nominal relationship between the same rating scales and ERS latent class factors in order to capture the non-monotone (U-shape) relationship. This is a more parsimonious model; for each response item, only one coefficient is estimated for the content factors. The model was further simplified to impose an equality constrain on the style factor coefficients (Kieruj and Moors 2013; Moors, Kieruj, and Vermunt 2014). In other words, only one set of coefficients is estimated for all items. This is theoretically meaningful since the style factor should affect each item equally regardless of the question content. This constraint is similar to the equal factor loading constraint in the CFA model by Billiet and colleagues (Billiet and McClendon 2000; Billiet and Davidov 2008). The model can be written as follows:

$$\log \frac{P(Y_{ij} = c + 1 | F1_i, F2_i, E_i)}{P(Y_{ij} = c | F1_i, F2_i, E_i)} = (\beta_{0jc+1} - \beta_{0jc}) + \beta_{1j_1} F1_i + \beta_{2j_2} F2_i + (\beta_{3jc+1} - \beta_{3jc}) E_i$$

Where  $Y_{ij}$  denotes the response of respondent  $i$  to item  $j$ ,  $i=1, \dots, I, j=1, \dots, 4$ ;

$F1_i$  denotes the internal efficacy latent class factor;

$F2_i$  denotes the external efficacy latent class factor;

$E_i$  denotes the extreme response style latent class factor;

$\beta_{1j_1}$  denotes the effects on the adjacent category logits for the internal efficacy latent class factor,

$j_1=1, 2$ ;



$\beta_{2j_2}$  denotes the effects on the adjacent category logits for the external efficacy latent class factor,

$j_2=3, 4$ ;

$\beta_{3jc+1} - \beta_{3jc}$  denotes the nonmonotone (U-shape) relationship between the extreme response style latent class factor and the items;

and  $c$  denotes the number of response options,  $c=1, 2, 3$ , or  $4$ .

To study ERS bias for the two types of rating scales, we fit the same measurement model for the four groups (pre-election A/D, pre-election IS, post-election A/D, and post-election IS). As illustrated in Figure 3.2, Items 1-4 refer to either the four A/D items or the four IS items, depending on the experimental groups under analysis. The model has two content latent class factors ( $F1_i, F2_i$ ) and one ERS latent class factor. The first content factor ( $F1_i$ ) measures internal efficacy and the second measures external efficacy ( $F2_i$ ). The items can only load on their corresponding content factors. At the same time, all items load on the same ERS latent factor. Previous studies suggest that, given our data structure, this model should fit the data well (Moors, Kieruj, and Vermunt 2014; Kieruj and Moors 2013). We will also fit several alternative models in each group to test whether other measurement models can better fit the data. Bayesian information criterion (BIC) is used to compare models and determine the empirically best-fitting model.

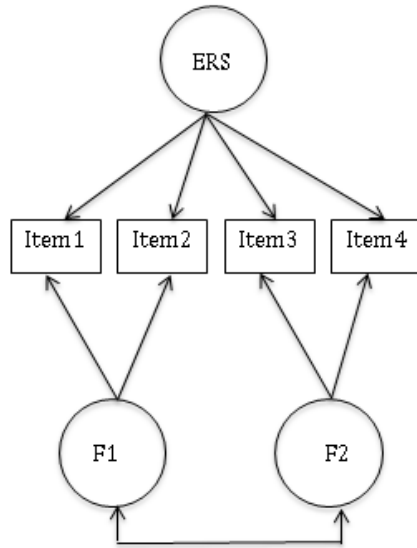


Figure 3.2. Latent class model with one style factor and two content factors. Items 1-4 represent A/D1-A/D4 or IS1-IS4.

### 3.4.2 *Between-subject experiment results*

To examine whether ERS is a consistent phenomenon or confined to one or two questions, we first present the proportion of extreme responses for each question from both question formats in the pre-election and post-election surveys (Table 3.1). In both the pre- and post-election surveys, respondents gave more extreme responses to all four questions when using the A/D format compared to the IS format. The independent t-test shows that differences in pre-election (i.e., columns (1)-(2)) and post-election (i.e., (3)-(4)) are all statistically significant at  $p < .001$ . We also compared scale format effects within subject across two waves. The difference between (1) and (4) represents the difference in extreme responses among the same group of respondents who responded using an A/D scale in the pre-election survey and an IS scale in the post-election survey. The paired t-test reveals significant differences for all questions at  $p < .001$ . The comparison between (3) and (2) shows that, for the same respondents, more extreme answers were given using the A/D scale (post-election) than the IS scale (pre-election). The differences are significant at  $p < .001$  for the “complicated” and “understand” questions,  $p < .05$  for the “official care” question, and marginally significant ( $p < .1$ ) for the “affect government” question. These results offer evidence that the A/D format elicits more extreme answers than the IS format and that the difference is not confined to a single question or the particular set of respondents but is consistent across all four questions for all respondents..

Table 3.1. Proportions and standard errors of extreme responses for Agree/Disagree (A/D) and Item Specific (IS) scales, 2012 ANES.

		A/D		IS				
		(1)	S.E.	(2)	S.E.	(1)-(2)	(1)-(4)	(3)-(2)
Pre-election	Complicated	29.00	0.83	15.43	0.67	13.57***	15.29***	15.42***
	Understand	27.79	0.82	14.23	0.65	13.56***	14.91***	9.42***
	Official care	25.77	0.80	22.04	0.77	3.74***	6.77***	3.59***
	Affect government	27.14	0.81	22.34	0.78	4.80***	6.94***	2.27***
		(3)	S.E.	(4)	S.E.	(3)-(4)		
Post-election	Complicated	30.84	0.88	13.70	0.66	17.14***		
	Understand	23.65	0.81	12.88	0.64	10.78***		
	Official care	25.62	0.83	19.00	0.75	6.62***		
	Affect government	24.61	0.82	20.20	0.77	4.41***		

\*\*\*  $p < .001$ , \*\*  $p < .05$ , \*  $p < .1$ .

Although Table 3.1 illustrates that the A/D scale has more extreme responses than the IS scale, it is possible that this difference is confounded with the true attitudinal differences between these two question formats. In other words, it is possible that the difference is the reflection of the actual attitudinal differences respondent expressed to the two sets of questions, or a combination of the true difference and response style difference. Table 3.1 does not tell us to what extent the pattern reflects response option preference, controlling for respondents' true attitudes toward the content of the questions. In order to control for this potential confounding effect, we performed latent class factor analysis to explicitly control for question content and to separate out the response style effect from the influence of substantive attitudes. Table 3.2 shows all the alternative models that we performed and the corresponding BIC. In principle, a model with a small BIC and relatively fewer parameters is preferred. We fit seven different models within each experimental condition for both pre- and post-election surveys. Models 1 and 2 are used to determine the appropriate latent class factors for the data. Model 1 has two content latent class factors only and Model 2 includes the ERS latent class factor in addition to the content latent class factors. For all four experimental conditions, adding the style factor improves the model fit (a smaller BIC), which suggests that the responses, regardless of the question formats, reflect something other than just the content factors. Next, we need to determine the number of classes for the latent class factors by comparing four alternative models, Model 2, 3a, 4, and 5, with two, three, four, and five classes, respectively. The BICs decrease monotonically when the number of classes increases from two to four for all four conditions, but the changes are very small when moving from three to four classes. When it comes to five-class models, however, the model fits deteriorate in comparison to four-class models for all but A/D in pre-election survey. Based on the BIC and ease of interpretation, we use three-class latent class factor models for all

four conditions. The model can be further simplified by imposing the  $\beta$  coefficients of ERS to be equal across all items. Models 3b and 3c test two alternative constraints. In Model 3b, the equality constraint is imposed on both ERS and the two content latent class factors. In Model 3c, only the ERS coefficients are constrained to be equal, but the coefficients associated with the two content factors are allowed to vary across items. The BIC indicates that Model 3c fits the data best for both A/S and IS formats in both pre- and post-election surveys. This is theoretically meaningful because ERS latent class factor is irrelevant to the question content and it is evident by the identical  $\beta$  coefficients across items.

Table 3.2. Model fit statistics (BIC) for Agree/Disagree (A/D) and Item Specific (IS) scales, 2012 ANES pre-election and post-election.

	Pre-election		Post-election		NoPar
	A/D	IS	A/D	IS	
Model 1: Content only (2-class)	31410	29185	28146	27203	23
Model 2: Content+ERS (2-class)	30838	28715	27468	26804	40
Model 3a: Content+ERS (3-class)	30483	28643	27119	26757	43
Model 3b: Equality on all latent factors (3-class)	30945	29110	27536	27118	28
Model 3c: Equality on style latent factors (3-class)	<b>30434</b>	<b>28583</b>	<b>27054</b>	<b>26687</b>	<b>31</b>
Model 4: Contents+ERS (4-class)	30398	28639	27049	26762	46
Model 5: Contents+ERS (5-class)	30381	28654	27069	26785	49

We next report the coefficients of the style latent class factors from Model 3c for both pre- and post-election surveys. Figure 3.3 presents the coefficients for A/D and IS format from the pre-election survey. The coefficients are log odds of the style latent class factor effect on response items. Since Model 3c imposes an equality constraint on the style latent class factor, there is only one set of coefficients for each question format. Also because of the response item is treated as a nominal variable, there is one coefficient for each response category. A positive coefficient in Figure 3.3 indicates a positive association between use of ERS and the likelihood of choosing the corresponding response category. As can be seen from Figure 3.3, a common feature shared by the two question formats is that the two endpoints have positive coefficients while the three middle response options have negative coefficients. Recall that the latent class factors are ordinal discrete variables with equidistant category scores. This suggests that a higher level of the style latent class variable represents tendencies toward ERS and a lower level represents a tendency to avoid ERS. However, there are two distinct patterns of ERS associated with these two question formats. For the A/D format, the “agree somewhat” and “disagree somewhat” categories fall between the two endpoints and the middle point, suggesting that the ERS latent class factor reflects the contrast between endpoints and midpoint. For the IS format, the three intermediate categories are close to each other while the endpoints are farther away. It indicates that the ERS coefficient for the IS format reflects the contrast between a preference for extreme responses and a preference for non-extreme responses. Regardless of these question-format specific ERS effects on response categories, the overall effect of ERS on response items is clearly observed from both types of questions. Figure 3.4 shows the ERS coefficients of A/D and IS formats from the post-election survey. We find exactly the same pattern as that from pre-



election survey. The consistent findings across two waves of surveys lend strong support that there are two different ERS patterns associated with A/D and IS formats.

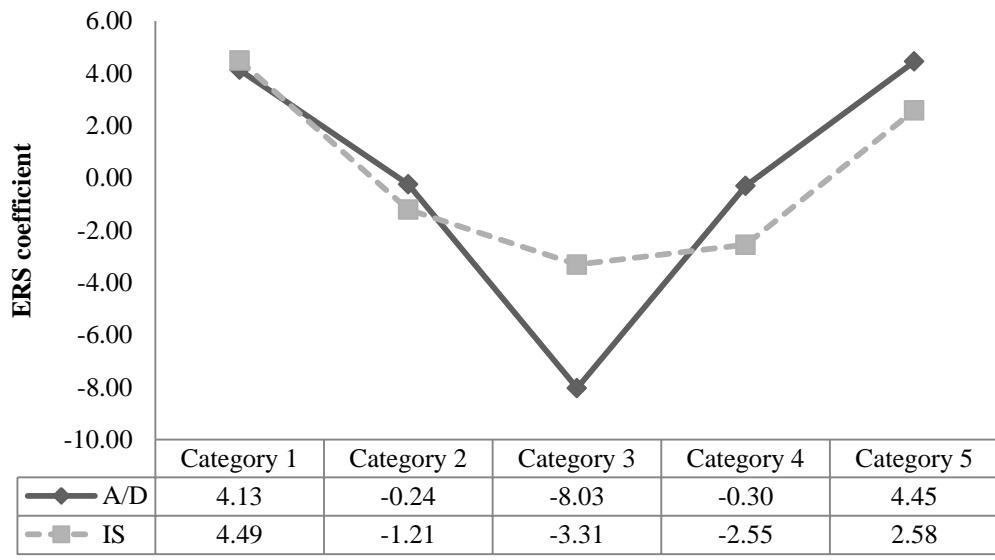


Figure 3.3. Extreme response style coefficients on agree/disagree (A/D) and item-specific (IS) scales, 2012 ANES pre-election survey. See Appendix for response categories 1-5.

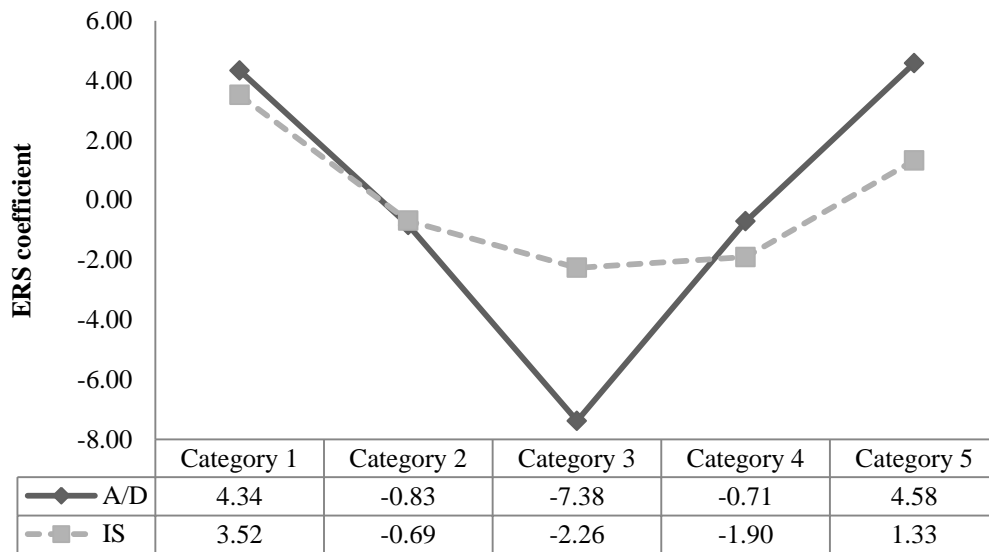


Figure 3.4. Extreme response style coefficients on agree/disagree (A/D) and item-specific (IS) scales, 2012 ANES post-election survey. See Appendix for response categories 1-5.

### 3.4.3 *Within-subject experiment analysis*

The experimental design embedded in a longitudinal survey not only allows us to test the ERS of two question formats between subjects, but it also permits us to examine within-subject differences between the question formats. Unlike the between-subject analysis, the within-subject analysis must take within-subject correlation into consideration when building the analysis model since the same respondents answer each question twice: once in A/D format and once in IS format. Figures 5(a) and (b) illustrate two alternative models for examining ERS across two waves. There are two groups of respondents for the within-subject analysis. The first group answered A/D formatted questions in the pre-election survey and IS formatted questions in the post-election survey, and the order of question formats is reversed for the second group. We fit latent class factor models the two groups separately in order to measure ERS. We are interested in comparing the ERS latent class factors between the two question formats within each group rather than between the two groups. As can be seen from Figure 3.5(a), the A/D questions were asked in the pre-election (or post-election) and IS were asked in the post-election (or pre-election) survey.  $F_1$  and  $F_2$  are the two content latent class factors for the A/D format and  $F_3$  and  $F_4$  are the two content latent class factors for the IS format. We allow correlation between the four latent class factors because in theory the within-subjects political efficacy should be correlated across question formats and across pre- and post-election surveys. In contrast, there is only one style latent class factor (ERS) for both question formats. This model suggests that within subjects, ERS is stable even though the question formats changed and the questions were administered at two points in time. The model in Figure 3.5(b) contains the same response items and content latent class factors. The only difference is that there are two style latent class factors ( $ERS_1$  and  $ERS_2$ ) for A/D and IS formats, respectively. The two style factors are correlated with

each other. This model suggests that the two question formats elicit two types of ERS within-subjects across two time points.

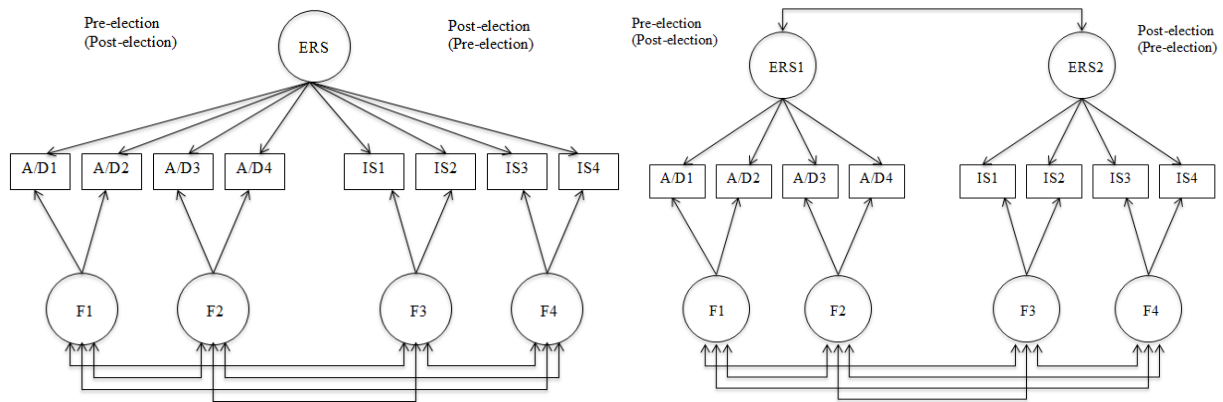


Figure 3.5. (a) Latent class model with one style factor and four content factors;  
 (b) latent class model with two style factors and four content factors.

#### 3.4.4 *Within-subject experiment results*

Figure 3.5 presents two theoretical measurement models measuring ERS across time using the panel data. However, we also test alternative models that are slight modifications of the ones in Figure 3.5. The BIC and number of parameters in these models are presented in Table 3.3. Model 1 contains four content latent class factors only, Model 2 contains four content factors and one style factor (Figure 3.5a), and Model 3 contains four content factors and two style factors (Figure 3.5b). For Group 1 members who were asked questions using the A/D format in the pre-election and IS format in the post-election, Model 2 has the smallest BIC. We then proceed to test two alternatives of this model by imposing equality constraints on the all latent class factors (Model 2a) or on the style latent class factor only (Model 2b). The second approach results in the smallest BIC, which suggests that for Group 1, the model in Figure 3.5a fits the data best. Similarly, for Group 2 members, who received IS questions in the pre-election and A/D questions in the post-election survey, the smallest BIC comes from Model 2b as well. This means that the model in Figure 3.5a consistently outperforms Figure 3.5b, regardless of the question administration order. In other words, the single ERS latent class factor for Group 1 indicates that once the within-subject correlation is built in to model, the question formats do not affect ERS.

Table 3.3. Model fit statistics (BIC) for within-subject analysis, 2012 ANES pre-election and post-election<sup>c</sup>.

	Group 1	Group 2	NoPar
	A/D→IS <sup>a</sup>	IS→A/D <sup>b</sup>	
Model 1: Content only	19334	18386	54
Model 2: Content+1 ERS	19201	18214	88
Model 2a: Equality on all latent factors	19619	18570	56
Model 2b: Equality on style latent factor	<b>19144</b>	<b>18091</b>	<b>60</b>
Model 3: Content+2 ERS	19207	18192	91
Model 3a: Equality on all latent factors	19637	18629	71
Model 3b: Equality on style latent factors	19151	18105	67

<sup>a</sup> Agree/disagree in pre-election and item specific in post-election survey.

<sup>b</sup> Item specific in pre-election and agree/disagree in post-election survey.

<sup>c</sup> All latent class factors have 3 classes.



The ERS coefficients for Groups 1 and 2 are plotted in Figures 3.6 and 3.7. The general conclusion from these two figures is that ERS represents a contrast between two extreme categories and the middle three categories. This is evident by the negative and relatively similar coefficients for the three non-extreme categories and the two positive coefficients at the endpoints. The ERS coefficients from the two figures are very similar to each other, which indicate the single ERS model is quite reliable.

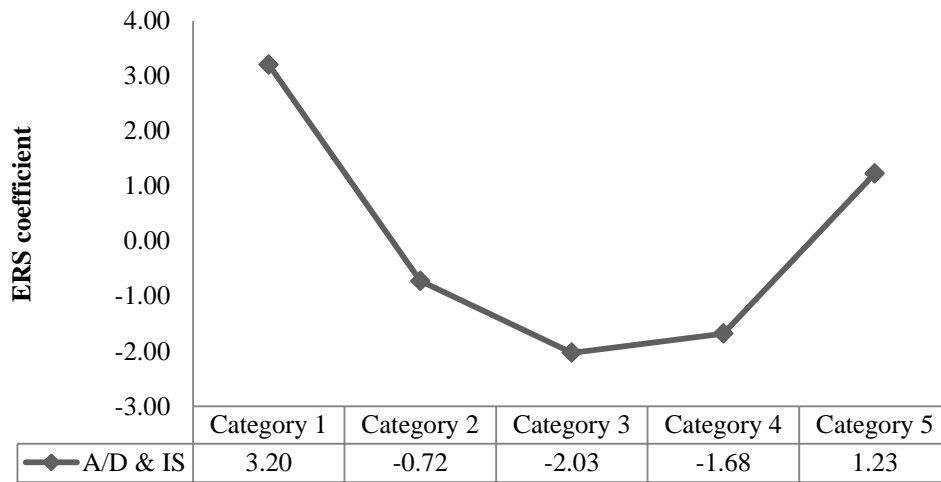


Figure 3.6. Extreme response style coefficients for Group 1 (A/D in the pre-election and IS in the post-election), Model 2b (Table 3.3), 2012 ANES. See Appendix for response categories 1-5.

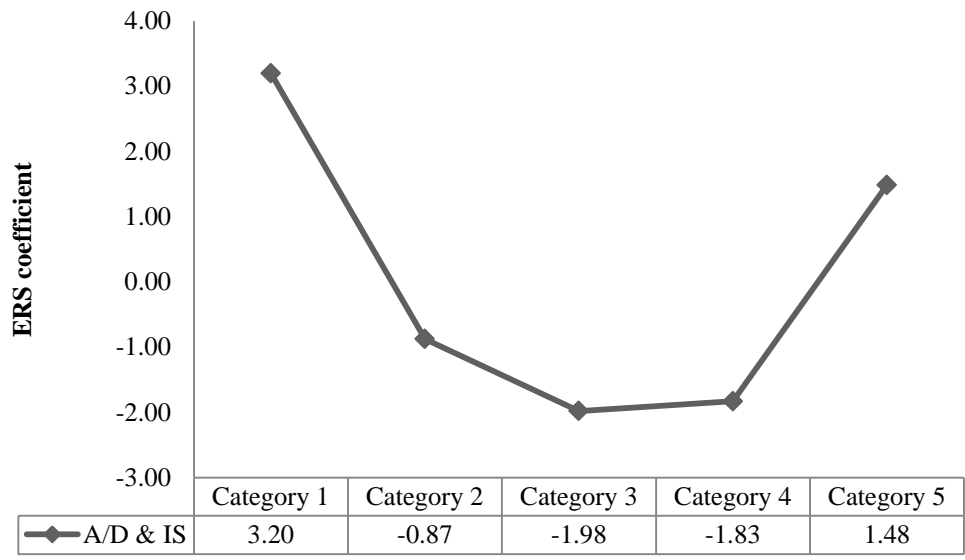


Figure 3.7. Extreme response style coefficients for Group 2 (IS in the pre-election and A/D in the post-election), Model 2b (Table3), 2012 ANES. See Appendix for response categories 1-5.

### 3.5 Discussion

In this study, we examined extreme response styles from two types of rating scales, namely agree-disagree (A/D) scales and item-specific (IS) scales, using data from between- and within-subject experiments. The four items we examined come from 2012 American National Election Studies and measure internal and external political efficacy. The experiments were carried out in two waves of data collection, with respondents randomly assigned to one of the two question formats in wave 1, and then assigned to the remaining format in wave 2. We conducted a latent class factor analysis and discovered three main findings: first, ERS exists in both A/D and IS scale formats. Ideally, we would have identified a scale format that is response style-free. Unfortunately, this study shows that we cannot achieve this by using item-specific scales. Second, the ERS from the two formats show slightly different shapes. Although respondents with higher level of ERS are more likely to choose the two endpoints in both formats, those who have lower level of ERS show different patterns in these two formats. In A/D format, ERS reflects a contrast between the two endpoints and the middle category while the two intermediate categories (“agree somewhat” and “disagree somewhat”) are somewhere in between. In IS format, ERS is the contrast between the two endpoints and the three non-extreme categories. This finding is consistent in both pre-election and post-election studies. Third, when analyzing ERS within subjects across two waves, there is only one single ERS for both question formats, regardless of which question type was asked first, after controlling for the correlation within respondents.

Why do the ERSs differ between A/D and IS, and why do they differ in this particular pattern? A closer look at the response options of the two scale formats may shed some light on this question. The verbal labels of these two scale formats are different. The A/D format is a

bipolar scale while the verbal labels for IS seem to be more unipolar. In a bipolar scale, the two endpoints represent the ends of an underlying continuum while the midpoint represents the conceptual center of this continuum. The conceptual distance from “agree strongly” to “neither agree nor disagree” and the distance from “disagree strongly” to “neither agree nor disagree” are equal. Since the respondents can easily pinpoint the conceptual center of an A/D scale, those who tend to avoid extreme points will move away from the endpoints toward the midpoint, the one that is the furthest from the extreme points. Thus, we observe a contrast between the extreme points and midpoints for the A/D scale.

For the IS scale, however, the conceptual center is more difficult to establish. For example, consider the following IS scale:

*Extremely well, very well, moderately well, slightly well, or not well at all*

It is clear that “moderately well” is the closest to the conceptual center of the scale but is not necessarily the center. Also, the conceptual distance between “extremely well” and “moderately well” and the distance between “not well at all” and “moderately well” are not necessarily equal. Consequently, when respondents want to avoid extreme options, they will pick one of the middle three categories with relatively equal likelihood, since the verbal labels of response categories on an IS scale do not necessarily contain a conceptual midpoint. These findings are consistent with the results from Moors, Kieruj, and Vermunt (2014). In their study, the balanced bipolar A/D scale with numerical labels from -3 to +3 elicits an ERS similar to the one for the A/D scale in the present study, while a unipolar A/D scale with numerical labels from 1 to 7 has an ERS similar to the one from the IS scale in the present study.

Another important finding is the similarity of the ERS between two question formats within subjects. Each respondent answered both types of scales across two waves of interviews.

Our analysis shows that once the within-person correlation is incorporated into the latent class factor analysis, the ERS differences between question formats disappear for both the group of respondents who answered A/D in pre-election and IS in post-election interview, and the group of respondents who answered IS in pre-election and A/D in post-election interview. As a result, there is only one ERS latent class factor for both question formats in both groups. This suggests that the response style is stable across two question formats and across two time points (pre- and post-election). This finding lends strong support to the argument that response styles have more to do with individual respondents than with survey design or question content. This finding is in line with the results reported in previous studies (Kieruj and Moors 2013; Weijters, Geuens, and Schillewaert 2010a; Weijters, Geuens, and Schillewaert 2010b). Those studies report stable response style for the A/D format in longitudinal surveys or demonstrate an association between response style and personality measures. In the 2012 ANES, the two waves are about four months apart, and the question formats are different in the two interviews. The identical and similar ERS we observed in the data provide further support for the argument that response style is a stable behavior.

One of the aims of this study is to identify a rating scale format that is ERS-free. As we already know, A/D format suffers from both acquiescent and extreme response style while IS format reduces the acquiescent response bias. This study shows that, like A/D format, IS format is still susceptible to ERS. Thus, the goal of eliminating ERS cannot be achieved by using IS format. However, given the advantage of IS format in reducing acquiescent response style, we would still recommend IS over A/D rating scales, in general. On the bright side, although we fail to remove ERS through scale design, we can still control for ERS, as well as other response styles, in the statistical analysis stage. Latent class factor analysis is one analytic approach that

measures substantive latent class factors and response style latent class factors at the same time. The estimates for the substantive latent construct should be free of the response style bias. Of course, this analytic approach imposes requirement on the rating scales. At the very least, the latent construct needs to be measured by multi-item scales rather than by one single question. Also, reverse-worded questions must be included for measuring acquiescent response style.

As this study shows, more work is required to examine and ultimately eliminate response styles from rating scales. In the present study, we used latent class factor analysis to estimate the response style. As we mentioned above, this is certainly not the only analytical approach one can take. There are other methods in the literature that have been used to examine response such as item response theory (Jong et al. 2008), multidimensional unfolding model (Javaras and Ripley 2007), and representative indicators response style means and covariance structure (Thomas, Abts, and Weyden 2014). However, we are not aware of any studies comparing these different analytical tools. Future research should evaluate and compare the existing analytic methods and resolve any differences in the conclusions in order to better adjust for response styles and produce more accurate survey estimates.

Furthermore, although the IS format fails to eliminate ERS, it is possible other designs of the IS format may elicit different levels of ERS. Future research should explore the scale length and the use of verbal and numeric labeling in IS format. Some research has examined variations of A/D format and their impacts on ERS (Kieruj and Moors 2010; Kieruj and Moors 2013; Moors, Kieruj, and Vermunt 2014), but this is lacking for IS format. Finally, would these findings change if the study had been conducted in another culture? Numerous studies have shown response-style differences between racial/ethnic groups within country and across different countries (for a comprehensive review, see Yang et al. 2010). As international

comparative studies become more prominent, it is critical to identify a scale format that can reduce response-style bias across different countries and cultural groups. So far, there is very little knowledge on this topic.



## References

- Alwin, D. F., and J. A. Krosnick. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods & Research* 20: 139–81.
- Balch, George I. 1974. "Multiple Indicators in Survey Research: The Concept 'Sense of Political Efficacy.'" *Political Methodology* 1: 1–43.
- Baumgartner, Hans, and Jan-Benedict EM Steenkamp. 2001. "Response Styles in Marketing Research: A Cross-National Investigation." *Journal of Marketing Research* 38 (2): 143–56.
- Bearden, William O., and Richard G. Netemeyer. 1999. *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. Thousand Oaks, CA: Sage.
- Berkowitz, Norman H., and George H. Wolkon. 1964. "A Forced Choice Form of the F Scale-Free of Acquiescent Response Set." *Sociometry* 27: 54–65.
- Billiet, Jaak B., and Eldad Davidov. 2008. "Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design." *Sociological Methods & Research* 36: 542–62.
- Billiet, Jaak B., and McKee J. McClendon. 2000. "Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items." *Structural Equation Modeling: A Multidisciplinary Journal* 7: 608–28.
- Bradburn, Norman M., Seymour Sudman, and Brian Wansink. 2004. *Asking Questions: The Definitive Guide to Questionnaire Design—for Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco, CA: Jossey-Bass.

- Bruner, Gordon C., Paul J. Hensel, and Karen E. James. 2001. *Marketing Scales Handbook*, Vol 5. Chicago, IL: American Marketing Association.
- Campbell, Angus, Gerald Gurin, and Warren E. Miller. 1954. *The Voter Decides*. Evanston, IL: Row, Peterson.
- Carpenter, Patricia A., and Marcel A. Just. 1975. "Sentence Comprehension: A Psycholinguistic Processing Model of Verification." *Psychological Review* 82: 45–73.
- Chamberlain, A. 2012. "A Time-Series Analysis of External Efficacy." *Public Opinion Quarterly* 76: 117–30.
- Cibelli, Kristen L., and Mario Callegaro. 2011. "Assessing the Measurement Quality of Agree/Disagree Items Versus Item Specific Answer Scales." Paper presented at the annual meeting for the Midwest Association for Public Opinion Research, Chicago, Illinois, November 18-19, 2011.
- Clark, Herbert H., and Eve V. Clark. 1977. *Psychology and Language*. San Diego, CA: Harcourt Brace Jovanovich.
- Converse, Jean M., and Stanley Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire*. Vol. 63. Thousand Oaks, CA: Sage.
- Counte, Michael A. 1979. "An Examination of the Convergent Validity of Three Measures of Patient Satisfaction in an Outpatient Treatment Center." *Journal of Chronic Diseases* 32: 583–88.
- DeVellis, Robert F. 2011. *Scale Development: Theory and Applications*. Vol. 26. Thousand Oaks, CA: Sage.
- Fowler, F. J., and Carol Cosenza. 2008. "Writing Effective Questions." *International Handbook of Survey Methodology* 8: 136–59.

- Fowler, Floyd J. 1995. *Improving Survey Questions: Design and Evaluation*. Vol. 38. Thousand Oaks, CA: Sage.
- . 2013. *Survey Research Methods*. Vol. 1. Thousand Oaks, CA: Sage.
- Greenleaf, Eric A. 1992. “Measuring Extreme Response Style.” *Public Opinion Quarterly* 56: 328–51.
- Javaras, Kristin N., and Brian D. Ripley. 2007. “An ‘unfolding’ Latent Variable Model for Likert Attitude Data: Drawing Inferences Adjusted for Response Style.” *Journal of the American Statistical Association* 102: 454–63.
- Jong, Martijn G. de, Jan-Benedict E. M. Steenkamp, Jean-Paul Fox, and Hans Baumgartner. 2008. “Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation.” *Journal of Marketing Research* 45: 104–15.
- Kieruj, Natalia D., and Guy Moors. 2010. “Variations in Response Style Behavior by Response Scale Format in Attitude Research.” *International Journal of Public Opinion Research* 22: 320–42.
- . 2013. “Response Style Behavior: Question Format Dependent or Personal Style?” *Quality & Quantity* 47: 193–211.
- Krosnick, Jon A., and Robert P. Abelson. 1992. “The Case for Measuring Attitude Strength in Surveys.” In *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*, edited by Judith M. Tanur, 177–203, New York, New York: Russell Sage Foundation.
- Krosnick, Jon A., and Matthew K. Berent. 1993. “Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format.” *American Journal of Political Science* 37: 941–64.

- Krosnick, Jon A., and Stanley Presser. 2010. "Question and Questionnaire Design." *Handbook of Survey Research 2*: 263–314.
- Moors, Guy. 2003. "Diagnosing Response Style Behavior by Means of a Latent-Class Factor Approach. Socio-Demographic Correlates of Gender Role Attitudes and Perceptions of Ethnic Discrimination Reexamined." *Quality and Quantity 37*: 277–302.
- Moors, Guy, Natalia D. Kieruj, and Jeroen K. Vermunt. 2014. "The Effect of Labeling and Numbering of Response Scales on the Likelihood of Response Bias." *Sociological Methodology*, accessed July 29, 2014, doi:10.1177/0081175013516114.
- Morrell, Michael E. 2003. "Survey and Experimental Evidence for a Reliable and Valid Measure of Internal Political Efficacy." *Public Opinion Quarterly 67*: 589–602.
- Morren, Meike, John P.T.M. Gelissen, and Jeroen K. Vermunt. 2011. "Dealing With Extreme Response Style In Cross-Cultural Research: A Restricted Latent Class Factor Analysis Approach: Extreme Response Style In Cross-Cultural Research." *Sociological Methodology 41*: 13–47.
- Niemi, Richard G., Stephen C. Craig, and Franco Mattei. 1991. "Measuring Internal Political Efficacy in the 1988 National Election Study." *The American Political Science Review 85* (4): 1407–13.
- O'Muircheartaigh, Colm, Jon Krosnick, and Armen Helic. 2000. *Middle Alternatives, Acquiescence, and the Quality of Questionnaire Data*. Harris School of Public Policy Studies, University of Chicago.
- Paulhus, Delroy L. 1991. "Measurement and Control of Response Bias." In *Measures of Personality and Social Psychological Attitudes. Measures of Social Psychological*

- Attitudes, Vol. 1*, edited by John P. Robinson, Phillip R. Shaver, Lawrence S. Wrightsman, 17-59, San Diego, CA: Academic Press.
- Ray, John J. 1979. "A Quick Measure of Achievement Motivation—validated in Australia and Reliable in Britain and South Africa." *Australian Psychologist* 14: 337–44.
- Revilla, M. A., W. E. Saris, and J. A. Krosnick. 2013. "Choosing the Number of Categories in Agree-Disagree Scales." *Sociological Methods & Research* 43: 73–97.
- Ross, Caroline K., Colette A. Steward, and James M. Sinacore. 1995. "A Comparative Study of Seven Measures of Patient Satisfaction." *Medical Care* 33: 392–406.
- Saris, Willem E., Melanie Revilla, Jon A. Krosnick, and Eric M. Shaeffer. 2010. "Comparing Questions with Agree/disagree Response Options to Questions with Item-Specific Response Options." *Survey Research Methods* 4: 61–79.
- Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA: Sage.
- Thomas, Troy D., Koen Abts, and Patrick Vander Weyden. 2014. "Response Styles and the Rural–Urban Divide." *Educational and Psychological Measurement* 74: 97–115.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. UK: Cambridge University Press.
- Trabasso, Tom, Howard Rollins, and Edward Shaughnessy. 1971. "Storage and Verification Stages in Processing Concepts." *Cognitive Psychology* 2: 239–89.
- Vaerenbergh, Yves Van, and Troy D. Thomas. 2012. "Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies." *International Journal of Public Opinion Research* 25: 195–217.

- Weijters, Bert, Maggie Geuens, and Niels Schillewaert. 2010a. "The Stability of Individual Response Styles." *Psychological Methods* 15: 96–110.
- . 2010b. "The Individual Consistency of Acquiescence and Extreme Response Style in Self-Report Questionnaires." *Applied Psychological Measurement* 34: 105–21.
- Weijters, Bert, Niels Schillewaert, and Maggie Geuens. 2008. "Assessing Response Styles across Modes of Data Collection." *Journal of the Academy of Marketing Science* 36: 409–22.
- Yang, Yongwei, Janet A. Harkness, Tzu-Yun Chin, and Ana Villar. 2010. "Response Styles and Culture." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell and Tom W. Smith, 203–223, Hoboken, New Jersey: Wiley.

Appendix 3.1 Question wording and response options for agree-disagree and item specific questions.

Likert Scale:

A/D1: Sometimes, politics and government seem so complicated that a person like me can't really understand what's going on.

A/D2: I feel that I have a pretty good understanding of the important political issues facing our country.

A/D3: Public officials don't care much what people like me think.

A/D4: People like me don't have any say about what the government does.

Response options are agree strongly (category 1), agree somewhat (category 2), neither agree nor disagree (category 3), disagree somewhat (category 4), disagree strongly (category 5).

Item specific items:

IS1: How often do politics and governments seem so complicated that you can't really understand what's going on?

Always (category 1), most of the time (category 2), about half the time (category 3), some of the time (category 4), or never (category 5)

IS2: How well do you understand the important political issues facing our country?

Extremely well (category 1), very well (category 2), moderately well (category 3), slightly well (category 4), or not well at all (category 5)

IS3: How much do public officials care what people like you think?

A great deal (category 1), a lot (category 2), a moderate amount (category 3), a little (category 4), or not at all (category 5)

IS4: How much can people like you affect what the government does?

A great deal (category 1), a lot (category 2), a moderate amount (category 3), a little (category 4), or not at all (category 5)



Appendix 3.2 Means and standard errors of demographic variables for Agree/Disagree (A/D) and Item Specific (IS) scales by mode of data collection, 2012 ANES.

	Face-to-face				Web				All			
	A/D	IS	$\chi^2$	p	A/D	IS	$\chi^2$	p	FTF	Web	$\chi^2$	p
Female	51.6	52.5	0.09	0.77	51.9	52.3	0.04	0.85	52.0	52.1	0	0.96
Age												
<30	21.4	20.9	2.59	0.76	20.1	22.0	4.93	0.42	21.2	21.0	4.81	0.44
30-39	16.0	15.1			15.3	14.6			15.6	15.0		
40-49	18.1	16.4			17.8	15.6			17.3	16.7		
50-59	17.7	20.0			20.7	19.5			18.9	20.1		
60-69	14.9	13.7			15.8	16.7			14.3	16.3		
>=70	11.7	13.9			10.2	11.7			12.8	10.9		
Race/ethnicity												
White	71.4	70.3	2.13	0.55	70.7	71.1	0.93	0.82	70.9	70.9	0.34	0.95
Black	12.6	11.3			11.8	12.1			11.9	11.9		
Hispanic	10.4	11.4			11.8	10.6			10.9	11.3		
Other	5.6	7.0			5.6	6.2			6.3	5.9		
Education												
High school or less	41.1	40.1	0.70	0.70	38.3	42.0	3.26	0.20	40.6	40.1	0.07	0.96
Some post-high school	29.0	31.3			31.2	29.6			30.1	30.4		
Bachelor's and above	29.9	28.6			30.5	28.4			29.3	29.5		
Household income (US Dollar)												
0-49999	51.4	48.8	0.67	0.88	48.3	50.7	2.60	0.46	50.1	49.5	2.59	0.46
50000-99999	28.8	31.1			32.9	30.9			30.0	31.9		
100000-149999	11.6	11.7			11.4	12.0			11.7	11.7		
150000+	8.2	8.4			7.4	6.3			8.3	6.9		
Marital status												
Married	51.6	54.9	7.36	0.12	55.3	51.2	6.05	0.20	53.2	53.3	0.62	0.96
Widowed	6.2	5.9			4.9	6.5			6.1	5.7		
Divorced	11.9	13.3			12.8	13.3			12.6	13.1		
Separated	3.6	1.6			2.0	2.6			2.6	2.3		
Never married	26.7	24.4			25.1	26.3			25.5	25.7		

## **Chapter 4**

### **The Effect of Acculturation on Extreme Response Style: A Mediation Analysis among a Sample of Arab American Adults**

#### **4.1 Introduction**

Since Cronbach (1942; 1946; 1950) developed and popularized the concept of response style more than 70 years ago, it has been the focus of much research. Three review papers reflecting on the early (Rorer 1965) and latest (Vaerenbergh and Thomas 2012; Yang et al. 2010) developments in this field show that researchers have always focused a great deal of attention on cross-cultural comparisons of response style. Given that cultural attributes are strongly related to response style differences, it is logical to expect that acculturation should also have an impact on response styles. That is, when respondents from one culture become acculturated to another culture, their response style should become more similar to that of respondents native to the new host culture. In contrast, less acculturated respondents should reveal a response style that reflects the native culture of their home country and is more similar to the response style of respondents who still reside in that country.

I am aware of only two studies that examine the relationship between acculturation and response style. Both studies focus on Hispanic respondents in the U.S. (Davis, Resnicow, and Couper 2010; Marin, Gamba, and Marin 1992). Since culture and response styles are closely related, the impact of acculturation may exhibit a different pattern for a different racial/ethnic

group. This study will expand this line of research by examining another ethnic minority group, Arab Americans, using the 2003 Detroit Arab American Study (DAAS). According to the 2006-2010 American Community Survey 5-year estimates, the number of people with Arab ancestry living in the U.S. has increased by 76% to over one and a half million since 1990. Surprisingly, little survey methodology work has been conducted among this fast-growing group. Like many immigrant groups, Arab Americans have a distinct culture that is different from mainstream American culture. Because response style is related to culture, it is important to understand the response style associated with Arabic culture in order to improve the quality of survey measurements used with this ethnic group. More importantly, in cross-cultural comparative studies, a thorough understanding of each cultural group's response style is critical for adjusting for such biases in the analysis process. In addition, as I will review in the next section, immigrants typically carry two sets of cultural norms, those of their home culture as well as those of their host culture. This characteristic, unique to immigrants, results in a response style that is more changable than stable. Respondents can favor various response styles depending on specific culture-related communication and interaction conditions. Consequently, examining the relationship between acculturation and response style can improve our knowledge of the mechanisms behind the impact of culture on response style.

## **4.2 Literature Review**

### *4.2.1 Extreme response style*

Extreme response style (ERS) refers to the response behavior in which, regardless of the content of the survey question, the respondent shows a higher tendency to choose the two endpoints (e.g., always/never, strongly agree/strongly disagree) of a response scale than to

choose the other categories on the scale (Paulhus 1991). Studies conducted in the U.S. have focused on comparisons of ERS between Caucasian and racial minority respondents. Findings from this line of study have shown that African Americans and Hispanics tend to give more extreme responses than European Americans (Bachman and O'Malley 1984; Marin, Gamba, and Marin 1992; Aday, Chiu, and Andersen 1980; Hui and Triandis 1989; Johnson et al. 1997). Another line of study compares ERS between North American respondents and East Asian respondents (Chen, Lee, and Stevenson 1995; Zax and Takahashi 1967; Shiomi and Loo 1999; Chun, Campbell, and Yoo 1974; Stening and Everett 1984; Culpepper and Lowery 2002). Generally speaking, these studies have found that East Asian respondents tend to give more midpoint and fewer extreme responses than North American respondents. Although earlier studies compare just a few societies, some recent studies involve samples from several countries and provide a more comprehensive understanding of country differences in response style (Harzing 2006).

#### *4.2.2 Culture and response styles*

Researchers have proposed some cultural explanations for the response style differences between racial/ethnic groups and across countries (e.g., Bachman and O'Malley 1984; Javeline 1999; Shiomi and Loo 1999). Generally, researchers believe that the various cultural backgrounds of respondents are reflected in their response behaviors and that response style is one of the culturally influenced behaviors. The norm for responding is influenced by the dominant social norms of the culture (Hui and Triandis 1989). An intensified contrast between cultures will result in more salient response style differences.

A popular post hoc explanation involves the individualism versus collectivism cultural contrast. East Asian culture is heavily influenced by Confucian-based collectivism. This interdependent self-construal emphasizes modesty, humility, and fitting in rather than standing out. Western societies, by contrast, are based on individualism, which emphasizes uniqueness and difference (Markus and Kitayama 1991). This cultural contrast can affect the cognitive process and hence the survey responses (Schwarz, Oyserman, and Peytcheva 2010). When the individualism vs. collectivism cultural difference is reflected in the respondents' survey scale usage, it would follow that more acquiescent responses will be reported from respondents from collectivist societies, which promote the achievement of harmonious interpersonal relationships, and more extreme answers from respondents from individualistic societies, which promote the freedom to express personal opinions without worrying about undue societal constraints.

Some recent studies quantitatively connect cultural orientation with response style. For example, Chen, Lee, and Stevenson (1995) conduct an ad hoc analysis of the relationship between individualism and response style at the respondent level and find that individualism is negatively correlated with midpoint response use and positively correlated with extreme response use. Other studies examine the cultural impact at the country level using Hofstede's cultural dimensions framework (Hofstede 2001; Hofstede, Hofstede, and Minkov 2010), which includes individualism-collectivism, masculinity-femininity, power distance, and uncertainty avoidance, and find mixed results (Van Hemert et al. 2002; Smith 2004; Johnson et al. 2005; Harzing 2006).

#### *4.2.3 Arabic culture, acculturation, and their impacts on response style*

The Middle East, including Arabic society, is dominated by honor-based collectivism, a deep-rooted cultural orientation that is derived from a history of warriorhood (Gregg 2005; Uskul, Oyserman, and Schwarz 2010). Arabic culture highly encourages and values actions that demonstrate self-worth, positive self-presentation, and self-reputation (Stewart 1994; Cohen et al. 1996; Nisbett and Cohen 1996). In an honor-based society, people are more likely to refer to honor-related values and to react to insult with more intense anger and shame than people in non-honor-oriented societies in order to maintain and protect their reputations (Cohen and Nisbett 1994; Cohen and Nisbett 1997; Mosquera, Manstead, and Fischer 2000; Mosquera, Manstead, and Fischer 2002; Rodriguez Mosquera et al. 2008). In a collectivist culture of honor, people not only maintain their own reputation but also those of their relatives and close friends. The self-image and reputation of a person in such a culture can be boosted by the reputational gains of his close associates or ruined by their loss of reputation (Stewart 1994; Gregg 2005). In other words, Arabic culture emphasizes a clear in-group versus out-group distinction, and people maintain their honor and reputation through a positive presentation of their in-group and a negative presentation of the out-group.

When honor-based culture is reflected in response style, it would follow that more extreme responses will be reported from respondents from honor-oriented societies than those from non-honor-oriented societies. This is because that I expect that a decisive and assertive answer, such as the extreme response, is preferred as a way of showing one's unambiguous attitude and standing, an important quality in an honor-based society. Some empirical findings support the argument that Arabic culture promotes ERS than Jew (Baron-Epel et al. 2010).

The pride of manhood, as demonstrated by masculine courage, warrior virtue, and physical strength, can also predict greater use of ERS in Arab countries, which are categorized as masculine societies (Cohen et al. 1996). In a masculine society, people are expected to be “assertive, competitive, and tough” (Hofstede 2001, 280). Hofstede also shows that both men and women in a masculine society show more of these qualities as compared to men and women in a feminine society, although there is gender variation within each culture. The masculine quality of an honor culture would encourage respondents to choose more extreme answers over other options as a demonstration of assertiveness (Johnson et al. 2005).

Logically, the cultural contrast between response styles should be reflected in the acculturation process among immigrants. Acculturation was first described as “those phenomena which result when groups of individuals having different cultures come into continuous first-hand contact, with subsequent changes in the original cultural patterns of either or both groups” (Redfield, Linton, and Herskovits 1936, 149). As immigrants acculturate into a new culture, their response style will reflect the predominant cultural norms of the host country. Less acculturated respondents, such as new immigrants, are more likely to reveal response styles consistent with the cultural norms of their home societies. Two previous studies using a sample of U.S. Hispanic respondents demonstrate that response style is partially explained by the respondent’s acculturation level (Marin, Gamba, and Marin 1992; Davis, Resnicow, and Couper 2010).

When it comes to Arab Americans, I hypothesize that more acculturated respondents will use a less extreme response style. In contrast, less acculturated Arab Americans are more likely to choose extreme responses, which are promoted by the honor cultures of their native societies. In this study, the language of the interview (i.e., Arabic or English) is treated as a mediation factor between acculturation and response style. Figure 4.1 represents the theoretical model of

acculturation, language, and ERS. The *a* path refers to the relationship between acculturation and interview language. More acculturated respondents have higher English proficiency and are therefore more likely to choose to be interviewed in English, whereas less acculturated respondents are more likely to choose Arabic. The *b* path of the mediation model refers to the path from language of interview to ERS.



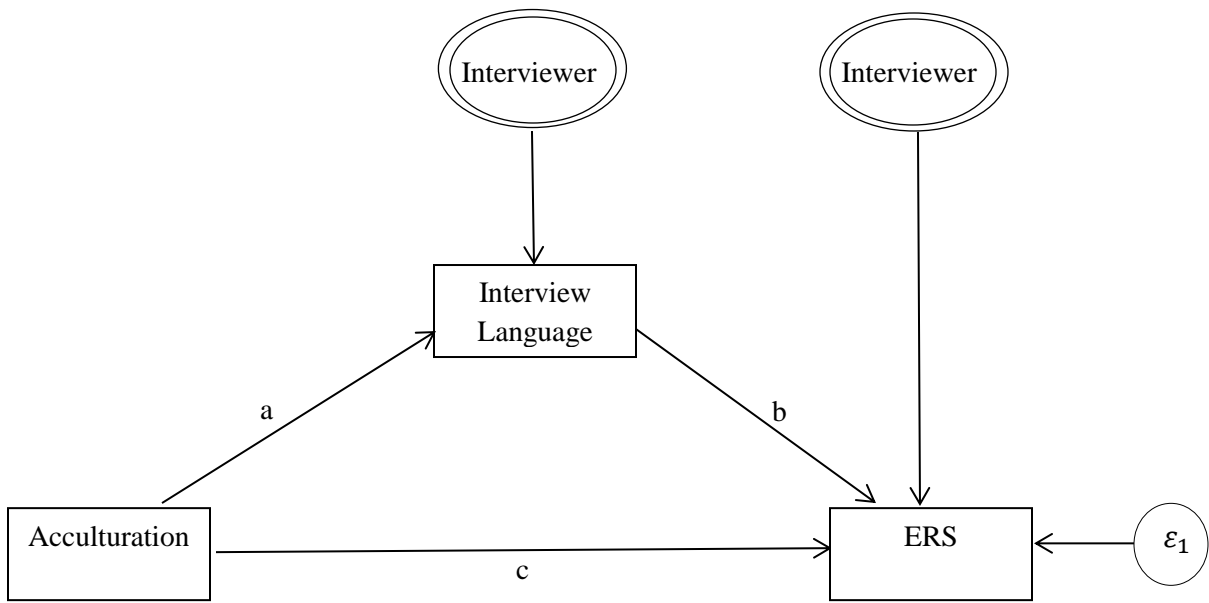


Figure 4.1 Theoretical Multilevel Generalized Structural Equation Modeling of Acculturation, Interview Language, and Extreme Response Style (ERS). Double-circle refers to level-2 variable.

Previous research has documented the complex inter connections between language, culture, and cognition and showed that they are intercorrelated. (See Semin 2009 for a review.) The attention of both the speaker and the listener can be channeled and directed by the language used for the conversation (Semin 2000; Tomasello 2003). In other words, language has a priming effect, and domain-specific psychological constructs can be activated by the language (Kashima 2009). This is especially true for bi-cultural people such as Arab Americans. In addition, the bilinear model of acculturation suggests that, instead of discarding the original culture, immigrants can maintain both original and host cultures simultaneously and change their behaviors based on the specific context in order to fit into different cultures (LaFromboise, Coleman, and Gerton 1993; Ryder, Alden, and Paulhus 2000). Specifically, the bilinear model suggests that both American culture and Arabic culture coexist in the minds of Arab Americans. However, a specific interaction feature, such as interview language, can prime one culture, and the activated culture will not only influence the respondent's communication style but also shape the cognitive construct. When interviewed in Arabic, the honor-based collectivist culture is likely to be activated and, as a result, respondents' response style will reflect the norms of that culture. Consequently, respondents will give more extreme responses. The *c* path of the mediation model is the direct effect of acculturation on response style, after adjusting for the mediator.

### **4.3 Data, Measures, and Models**

#### *4.3.1 Data source*

The 2003 Detroit Arab American Study (DAAS) is a face-to-face survey conducted as a companion to the 2003 Detroit Area Study. The DAAS uses a dual-frame sample design with an area probability representative sample from the tri-county Detroit metropolitan area and a list

sample drawing from lists provided by Arab American organizations in the greater Detroit metropolitan area. The target population is adults aged 18 or older who live in the greater Detroit metropolitan area and self-identify as Arab or Chaldean. The final sample size is 1016 with a response rate of 73.1% (AAPOR RR3).

#### *4.3.2 Dependent variables*

The ERS is the percentage of the items for which respondents chose the extreme answers of ordinal scales. Adapting from the criteria of Bachman and O'Malley (1984), I selected questions that use ordinal response scales, do not deal with demographics, and are not strictly factual. I also excluded questions that were not asked of the whole sample. The resulting analysis examines 13 questions with 5-point scales and 11 questions with 3-point scales (listed in the Appendix). Two ERS indicators—one for 5-point scales and one for 3-point scales—were created. They are the dependent variables. To put it differently, ERS indicators are the percentages of the two extreme responses provided by each respondent  $i$  out of the number of questions answered, that is,

$$Y_i = 100 \times \frac{\text{extreme response}_i}{\text{questions answered}_i}$$

As a result, the dependent variables are two continuous variables ranging from 0 to 100. This coding convention has been used in several previous studies (Davis et al., 2010; Greenleaf, 1992; Reynolds & Smith, 2010).

#### *4.3.3 Independent variables*

Five acculturation indicators are the key predictors in the analysis:

*Years in the U.S.:* The length of time a respondent has lived in the U.S. is calculated using either the year he or she moved to the U.S. for those who were not born in the U.S. or the respondent's age if he or she was born in the U.S.

*Parent born in U.S.:* 1 = both parents were born outside the U.S., 0 = at least one parent was born in the U.S.

*Home English only:* 1 = English only at home, 0 = speak some language other than English at home.

*Not watch Arabic TV:* 1 = does not watch TV news broadcast in Arabic in a typical week, 0 = does watch.

*Interaction with other race:* number of times respondent has been in the home of a friend of a different race or had a friend of a different race in the respondent's own home in the previous 12 months.

#### *4.3.4 Mediation variable*

The language of interview serves as a mediation variable in the analysis. Respondents are provided with the choice of Arabic or English. This variable is measured as Arabic only, mostly Arabic, equal division between Arabic and English, mostly English, and English only. In the analysis, it is coded as Arabic only (reference group), both languages, and English only. All interviews were conducted by bilingual interviewers.

#### *4.3.5 Control variables*

Demographic variables related to response style are controlled for in the analysis.

*Sex:* 1 = female, 0 = male.

*Age*: a continuous variable measured in years.

*Education level*: 1 = high school or less, 2 = some college/college, 3 = graduate school, 4 = other. It is used as nominal variable.

*Religion*: 1 = Christian, 2 = Muslim, 3 = Other<sup>1</sup>. It is treated as nominal variable.

Descriptive data are presented in Table 4.1.

---

<sup>1</sup> Other includes Jewish, other religion, atheist, or none.

Table 4.1. Descriptive Data for Variables in the Analysis

	Mean	Std. Dev.	Min	Max
<i><u>Dependent variable</u></i>				
Extreme response style 5-point (%)	30.94	21.91	0	100
Extreme response style 3-point (%)	51.34	22.79	0	100
<i><u>Mediation variable</u></i>				
Arabic interview (%)	37	48	0	100
Both language (%)	16	37	0	100
English interview (%)	45	50	0	100
<i><u>Acculturation variables</u></i>				
Years in the U.S.	25.58	18.33	0	88
Parent born in U.S. (%)	13	34	0	100
Home English only (%)	20	40	0	100
Interaction with other race	9.77	31.24	0	365
Not watch Arabic TV (%)	52	50	0	100
<i><u>Control variables</u></i>				
Female (%)	54	50	0	100
Age	43.63	16.48	18	88
High school or less (%)	48	50	0	100
College/Some college (%)	47	50	0	100
Graduate school (%)	04	19	0	100
Other education (%)	1	9	0	100
Christian (%)	57	50	0	100
Muslim (%)	42	49	0	100
Other religion (%)	1	11	0	100

#### 4.3.6 Analytical strategy

I fitted a two-level mediation model using a multilevel generalized structural equation modeling (GSEM) approach. A mediation model is one that seeks to identify the relationship between an independent variable and a dependent variable via the inclusion of a third variable, known as a mediator variable (David P. MacKinnon, Fairchild, and Fritz 2007; David Peter MacKinnon 2008). Ideally, a mediation model should be applied in a longitudinal design with the predictor, mediator, and outcome measured at different time points sequentially. However, considering the nature of the retrospective information in my model, it is reasonable to apply a mediation model to this cross-sectional data set. More specifically, the acculturation variables are all factual questions asking about events or behaviors that happened prior to the survey. The same responses are expected even if the acculturation measures were administered to the respondents at an earlier time point. The level of acculturation can influence the choice of language for the interview. A reverse causal relationship is unlikely. That is, the choice of interview language is unlikely to exert any influence on respondents' acculturation.

The mediation variable, or the language of the interview, is the first question in the survey, preceding all other questions. The scale questions are in the middle or toward the end of the questionnaire. By the time interviewers asked those questions, the language should have already activated the related culture and hence influenced the response style. A reverse causal relation is very unlikely. That is to say, the choice of the language is due to the acculturation level rather than the extremity revealed in the responses.

In order to test the language-mediated relationship between acculturation and ERS, I fitted three models simultaneously to predict the language of interview and the ERS. More specifically, in the first model, I regressed the language of interview on each acculturation

variable and control variables using multinomial logistic regression with the Arabic interview as the reference group. The coefficients from this model correspond to the  $a$  path in Figure 4.1, in log odds scale. In the second model, I regressed the ERS on the language of interview, acculturation, and control variables using OLS regression. The coefficients from this model correspond to the  $b$  path in Figure 4.1, in linear scale. In the third mode, the ERS is regressed on the acculturation variable and control variables using OLS regression. The coefficients from this model correspond to the  $c$  path in Figure 4.1, known as the *direct effect* after controlling for the mediator variable. The extent to which the relationship between acculturation and ERS is mediated by the mediation variable is evaluated by a product of the two component paths,  $a$  and  $b$ . This is called indirect effect. I compute the variance of the indirect effect by taking the nonlinear transformation of the estimated parameter vector from the fitted model and applying the delta method. This is due to the fact that the  $a$  path is estimated on a log odds scale and the  $b$  path is estimated on a linear scale.

Analyzing each acculturation variable one at a time is a more conservative method than analyzing all acculturation variables in one model. If the acculturation variables were all included in one model, I would be testing multiple indirect effects at once and would more likely to observe a significant indirect effect just by chance. A random intercept for the interviewer (level 2) is introduced in both models in order to account for the clustering of respondents within each interviewer. All other variables in the model, including independent, mediation, and outcome variables, are observed at the respondent level (level 1). As Krull and MacKinnon (2001) suggest, for a data structure like the DAAS in which respondents are nested within interviewers, a multilevel mediation model produces accurate standard errors of the mediated effects while the single-level underestimates the standard errors by 20% or more. Also, as



compared to the traditional multilevel mediation analysis in which direct and indirect effects are modeled separately, the multilevel SEM approach produces non-conflated or unbiased indirect effects (Preacher, Zyphur, and Zhang 2010). Besides, the generalized SEM permits me to use the generalized linear model with various family and link functions to estimate direct or indirect effects, or both, which is the case in our model.

I estimated the generalized multilevel SEM using the `gsem` function and computed the indirect effect and its variance using the `nlcom` function in Stata 13. The `gsem` procedure only produces AIC, BIC, and log likelihood as model fit statistics. Although some traditional model fit statistics, such as the root mean square error of approximation (RMSEA), are not available, they are not crucial because the analysis models are theoretically driven, and we will not revise them purely based on empirical data.

#### **4.4 Results**

Before reporting the results from the mediation analysis, I want to briefly describe two aspects of the sample. First, as we know, the Arab American population is diversified, comprised of several different nationalities. The survey asks “Is there any other term like ‘Arab American’ that better describes you?” Among all the responses, 36 mentioned Chaldean, 20 mentioned Lebanese, 8 mentioned Yemeni, 6 mentioned Iraqi, 5 mentioned Egyptian, 5 mentioned Palestinian, 2 mentioned Jordan, and 2 mentioned Syrian. The relatively small subgroup sample size prevents us from doing statistical modeling within in each group. As a result, we combined all the respondents for the mediation analysis. Second, the language capability of the respondents is not necessarily equal. One reason that respondents responded in English could be that they do not speak Arabic. The survey asks whether the respondents speak Arabic at home, whether they

watch TV news broadcast in Arabic, whether they listen to radio news in Arabic, whether they read newspaper in Arabic, and whether they have satellite dish receiving Arabic programs. If someone answered no to all these questions and was interviewed in English, it suggests that he/she does not speak Arabic. I found five such respondents. I removed these five cases from further analysis to make the analysis more conservative.

The mediating effects of the acculturation on ERS are presented in Tables 2 and 3, for 5-point scales and 3-point scales respectively<sup>2</sup>. As expected, more acculturated respondents were less likely to choose the extreme options on the Likert scales, and this effect was mediated by the language of interview. In this analysis, higher acculturation means having lived longer in the U.S., having at least one parent born in the U.S., using only English at home, not watching TV in Arabic regularly, or interacting frequently with non-Arab people. Specifically, more acculturated respondents were more likely to choose either English only or both languages to conduct the interview compared to the less acculturated respondents (*a* path). Also, interviews conducted in English only or both languages resulted in fewer extreme responses, compared to Arabic only interviews (*b* path). Consequently, there was a negative association between acculturation and use of extreme response style.

---

<sup>2</sup> After fitting the model in Stata 13, we found that the model did not converge after over 100 iterations. After forcing the model to produce estimates, I found that the variance of the random intercept in the model predicting language of interview by acculturation variables (*a* path) was close to zero. To resolve that, I followed the suggestion of Acock (2013) and constrained this variance to 0.00001 to allow the model to converge. Also, theoretically, it is reasonable to assume that the choice of interview language does not vary substantially across interviewers. Rather, it is mainly driven by the preference or language ability of the respondents.

Table 4.2. Mediation Effects of Acculturation on Extreme Response Style, 5-point scales<sup>1</sup>.

Independent variable (IV)	Years in U.S.	Parent born in U.S.	Home English only	Not watch Arabic TV	Interaction with other race
<i>a path</i>					
IV→Both language	<b>0.07</b>	0.48	<b>0.97</b>	<b>0.52</b>	<b>0.01</b>
IV→English	<b>0.19</b>	<b>3.95</b>	<b>3.53</b>	<b>2.32</b>	<b>0.02</b>
<i>b path</i>					
Both language→ERS <sup>2</sup>	<b>-5.73</b>	<b>-6.38</b>	<b>-6.21</b>	<b>-6.05</b>	<b>-6.55</b>
English→ERS	<b>-4.93</b>	<b>-5.47</b>	<b>-5.49</b>	<b>-4.85</b>	<b>-6.98</b>
<i>Indirect effect (a×b path)</i>					
IV-->Both language-->ERS	<b>-0.38</b>	-3.06	-6.07	-3.13	-0.09
IV-->English-->ERS	<b>-0.91</b>	<b>-21.64</b>	<b>-19.39</b>	<b>-11.26</b>	<b>-0.12</b>
Total indirect effect	<b>-1.29</b>	<b>-24.69</b>	<b>-25.46</b>	<b>-14.39</b>	<b>-0.21</b>
<i>Direct effect (c path)</i>					
IV-->ERS	-0.08	-3.65	-2.85	<b>-3.54</b>	<b>0.06</b>
Total effect	<b>-1.37</b>	<b>-28.31</b>	<b>-28.31</b>	<b>-17.93</b>	-0.14
AIC	10421.48	10647.66	10647.66	10647.29	10779.27
BIC	10583.79	10809.98	10809.98	10809.61	10941.59
Log Likelihood	-5177.74	-5290.83	-5290.83	-5290.64	-5356.64
n	1011	1011	1011	1011	1011

Note:  $p < .05$  are in bold and  $p < .1$  are in italics.

<sup>1</sup> All models control for sex, age, education level, and household annual income.

<sup>2</sup> ERS=extreme response style

I will use the *years in the U.S.* variable in Table 2 as an example to explain how to interpret the coefficients of the mediation analysis. Every additional year the respondents spent in the U.S., the log odds of using both English and Arabic for the interview increased by 0.07 ( $p < .05$ ) and the log odds of using English only increased by 0.19 ( $p < .05$ ), in comparison to Arabic only interviews. Compared to Arabic interviews, people who used both languages provided 5.73 percentage points ( $p < .05$ ) fewer extreme responses, and people who were interviewed in English provided, on average, 4.93 percentage points ( $p < .05$ ) fewer extreme responses. As a result, there was a negative indirect effect from *years in the U.S.* to extreme response style through interview in both languages ( $\hat{\beta} = -0.38, p < .05$ ), and through interview in English ( $\hat{\beta} = -.91, p < .05$ ). The total indirect effect ( $\hat{\beta} = -1.29$ ) was also statistically significant ( $p < .05$ ). After taking account of the indirect effect, the direct effect from *years in the U.S.* to extreme response usage became statistically insignificant ( $\hat{\beta} = -0.08, p = .19$ ). Adding up the total indirect effect and direct effect gives a total negative effect of *years in the U.S.* on extreme response style ( $\hat{\beta} = -1.37, p < .05$ ).

For the *parent born in U.S.* variable, the results show that, compared to people whose parents were both born outside the U.S., people with at least one parent born in the U.S. were less likely to use extreme responses, and the effect was mediated through English interview ( $\hat{\beta} = -21.64, p < .05$ ). For the *home English only* variable, the indirect effect through English interview was statistically significant ( $\hat{\beta} = -19.39, p < .05$ ), and the indirect effect through interview in both languages was marginally significant ( $\hat{\beta} = -6.07, p = .07$ ). This suggests that, if the respondent used English only at home, they were less likely to reveal ERS than those who used a language other than English as well. For respondents who did not watch TV in Arabic on a regular basis, there were negative indirect effects on ERS, one through interview in both languages ( $\hat{\beta} = -3.13, p = .07$ )

and one through English interview ( $\hat{\beta}=-11.26, p<.05$ ). Finally, the effect of *interaction with other race* was also mediated by the interview language. The indirect effect through interview in both languages was marginally significant ( $\hat{\beta}=-.09, p=.07$ ) and the indirect effect through interview in English was statistically significant ( $\hat{\beta}=-.12, p<.05$ ).

For both the *not watch Arabic TV* and *interaction with other race* variables, the direct effects on ERS were still statistically significant ( $\hat{\beta}=-3.54, p<.05$ ;  $\hat{\beta}=0.06, p<.05$ ), and the *parent born in U.S.* variable had a marginally significant direct effect ( $\hat{\beta}=-3.65, p=.09$ ), after adjusting for the mediator. In general, the acculturation effects on ERS were partially mediated by the language of interview. More specifically, there was a negative association between acculturation and ERS for all variables but *interaction with other race*, where a positive direct relationship was discovered.

When it comes to 3-point scales, many fewer acculturation effects were found in comparison to 5-point scales. As shown in Table 3, similarly to the results of 5-point scales, acculturation affected the language respondents chose to use. More acculturated respondents were more likely to use English or both English and Arabic than Arabic only. When it comes to the impact of acculturation on ERS, however, I did not find any indirect effects. I did detect two direct effects of acculturation on ERS, after adjusting for the mediator. As Arab Americans spent more time in the U.S. ( $\hat{\beta}=0.18, p<.05$ ) and as they interacted more frequently with non-Arabians ( $\hat{\beta}=.10, p<.05$ ), they were more likely to engage in ERS in the survey. This suggests that the relationship between the acculturation and ERS is not fully mediated by the interview language. We call it complementary mediation model (Zhao, Lynch Jr., and Chen 2010).

Table 4.3. Mediation Effects of Acculturation on Extreme Response Style, 3-point scales<sup>1</sup>.

Independent variable (IV)	Years in U.S.	Parent born in U.S.	Home English only	Not watch Arabic TV	Interaction with other race
<i>a path</i>					
IV→Both language	<b>0.07</b>	0.48	<b>0.98</b>	<b>0.52</b>	<b>0.01</b>
IV→English	<b>0.19</b>	<b>3.95</b>	<b>3.53</b>	<b>2.32</b>	<b>0.02</b>
<i>b path</i>					
Both language→ERS <sup>2</sup>	0.54	1.96	1.85	1.99	1.58
English→ERS	-2.95	1.13	-0.49	0.66	-0.35
<i>Indirect effect (a×b path)</i>					
IV-->Both language-->ERS	0.04	0.94	1.81	1.03	0.02
IV-->English-->ERS	-0.55	4.48	-1.73	1.54	-0.01
Total indirect effect	-0.51	5.42	0.07	2.57	0.01
<i>Direct effect (c path)</i>					
IV-->ERS	<b>0.18</b>	-2.26	3.23	-0.25	<b>0.10</b>
Total effect	-0.33	3.16	3.31	2.32	<b>0.11</b>
AIC	10454.41	10695.35	10687.20	10692.68	10809.31
BIC	10616.73	10857.66	10849.52	10855.00	10971.63
Log Likelihood	-5194.21	-5314.67	-5310.60	-5313.34	-5371.66
n	1011	1011	1011	1011	1011

Note:  $p < .05$  are in bold and  $p < .1$  are in italics.

<sup>1</sup> All models control for sex, age, education level, and household annual income.

<sup>2</sup> ERS=extreme response style

## 4.5 Discussion

The findings of this study lend support to the hypothesis that as Arab Americans become more acculturated, they are less likely to use extreme responses. Also, the relationship between acculturation and ERS is mediated by the language of interview. Respondents with higher levels of acculturation are more likely to choose to participate in the interview using English or both English and Arabic whereas respondents with lower level of acculturation are more likely to be interviewed in Arabic only. Also, as long as English is used during the interview, no matter fully or partially, the likelihood of providing extreme responses decreases. In addition, after accounting for the indirect effects, some significant direct effects still exist depending on the type of acculturation indicators, suggesting that the language of interview partially mediated the relationship between acculturation and ERS. Finally, acculturation exerts a larger effect on questions with a 5-point answer scale than those with a 3-point scale, as all five acculturation indicators are significantly associated with ERS when a 5-point scale was used, while only two acculturation indicators have significant impact when a 3-point scale was used.

These findings are consistent with the theoretical expectation that, in an honor-based collectivist society, such as the Arabic world, respondents are more likely to use ERS when expressing opinions in a survey in order to demonstrate assertiveness and decisiveness, maintain the reputation of the respondent and close associates, and earn respect, all of which are mainstream values in such cultural contexts (Uskul, Oyserman, and Schwarz 2010). As Arab Americans become acculturated in the U.S., they are likely to carry both cultural norms with them and selectively choose one to follow in a specific context. The language of the interview creates a specific communication context in which certain aspects of the cultural norm are activated by the language used by the interviewer and respondent. This priming effect of

interview language can shape the cognitive processes of answering survey questions and hence influence the response behaviors. Consequently, the responses given by the respondents will reflect the cultural norm associated with the language. For example, if the interview is conducted in Arabic, the honor-based culture will be primed and it will influence the response behaviors, such as ERS.

One contradictory finding is the positive direct effect between *interaction with other race* and ERS, which suggests that acculturation results in more extremity. However, it is entirely possible that social activity such as visiting others' home or having others at one's own home is determined by personality. Those who interact frequently with people of other races may interact even more with Arab Americans due to their extroverted personalities. It is also entirely possible that the non-Arabians are also from cultures that promote ERS. However, the data at hand do not allow me to test these hypotheses. The absence of some acculturation effects on the 3-point scale is not entirely surprising. In a study across four countries, Clarke (2001) finds that the ERS differences from 3-point scales are smallest, in comparison with longer scales. The finding is consistent with this literature in that the ERS differences across respondents with various acculturation levels are smaller or absent in 3-point scale questions versus 5-point scale questions.

The DAAS was not designed to assess response style and, consequently, has several limitations. First, the questions in the analysis address a few different topics, which suggests that the result is not driven by the content of certain questions. However, as Greenleaf (1992) suggested, I would ideally prefer questions that cover a wider range of subjects. Second, in order to study the scale length in a more rigorous manner, future study needs to experimentally manipulate the scale length, while maintaining the same question content. Third, the survey does



not contain any acculturation scales and hence I am not able to compare the predictability of ERS between scales and the subjective acculturation measures such as those used in the analysis.

Fourth, the findings suggest that the effect from language of interview on ERS is due to the priming effect of the language. However, it is also possible that some structural aspects of Arabic itself can contribute to Arabic language interview respondents' more frequent use of ERS than those interviewed in English. In order to test this, it would be necessary to randomly assign bilingual respondents a language of interview so that one could separate the culture from the language itself. Fifth, I acknowledge that generalization of the findings from DAAS can be limited since the survey was conducted in an area with a high density of Arab Americans. More work needs to be done to explore the response style among Arab Americans who reside in areas with a lower density of Arab Americans since the acculturation process can be different in such areas.

## References

- Acock, Alan C. 2013. *Discovering Structural Equation Modeling Using Stata*. College Station, TX: Stata Press.
- Aday, Lu Ann, Grace Y. Chiu, and Ronald Andersen. 1980. "Methodological Issues in Health Care Surveys of the Spanish Heritage Population." *American Journal of Public Health* 70 (4): 367–74.
- Bachman, Jerald G., and Patrick M. O'Malley. 1984. "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles." *Public Opinion Quarterly* 48 (2): 491–509.
- Baron-Epel, O., G. Kaplan, R. Weinstein, and M. S. Green. 2010. "Extreme and Acquiescence Bias in a Bi-Ethnic Population." *The European Journal of Public Health* 20 (5): 543–48.
- Chen, Chuansheng, Shin-Ying Lee, and Harold W. Stevenson. 1995. "Response Style and Cross-Cultural Comparisons of Rating Scales among East Asian and North American Students." *Psychological Science* 6 (3): 170–75. doi:10.2307/40063010.
- Chun, Ki-Taek, John B. Campbell, and Jong Hae Yoo. 1974. "Extreme Response Style in Cross-Cultural Research A Reminder." *Journal of Cross-Cultural Psychology* 5 (4): 465–80.
- Clarke III, Irvine. 2001. "Extreme Response Style in Cross-Cultural Research." *International Marketing Review* 18 (3): 301–24.
- Cohen, Dov, and Richard E. Nisbett. 1994. "Self-Protection and the Culture of Honor: Explaining Southern Violence." *Personality and Social Psychology Bulletin* 20 (5): 551–67.

- . 1997. “Field Experiments Examining the Culture of Honor: The Role of Institutions in Perpetuating Norms about Violence.” *Personality and Social Psychology Bulletin* 23 (11): 1188–99.
- Cohen, Dov, Richard E. Nisbett, Brian F. Bowdle, and Norbert Schwarz. 1996. “Insult, Aggression, and the Southern Culture of Honor: An Experimental Ethnography.” *Journal of Personality and Social Psychology* 70 (5): 945–60.
- Cronbach, Lee J. 1942. “Studies of Acquiescence as a Factor in the True-False Test.” *Journal of Educational Psychology* 33 (6): 401–15.
- . 1946. “Response Sets and Test Validity.” *Educational and Psychological Measurement* 6 (4): 475–94.
- . 1950. “Further Evidence on Response Sets and Test Design.” *Educational and Psychological Measurement* 10: 3–31.
- Culpepper, Robert A., and Clyde Lowery. 2002. “Survey Response Bias Among Chinese Managers.” In *Academy of Management Proceedings*, 2002:J1–J6.
- Davis, R. E., K. Resnicow, and M. P. Couper. 2010. “Survey Response Styles, Acculturation, and Culture Among a Sample of Mexican American Adults.” *Journal of Cross-Cultural Psychology* 42 (7): 1219–36.
- Greenleaf, Eric A. 1992. “Measuring Extreme Response Style.” *Public Opinion Quarterly* 56 (3): 328–51.
- Gregg, Gary S. 2005. *The Middle East: A Cultural Psychology*. New York, NY: Oxford University Press.
- Harzing, Anne-Wil. 2006. “Response Styles in Cross-National Survey Research A 26-Country Study.” *International Journal of Cross Cultural Management* 6 (2): 243–66.

- Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*. Thousand Oaks, CA: Sage.
- Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov. 2010. *Cultures And Organizations: Software For The Mind*. New York, NY: McGraw-Hill.
- Hui, C. Harry, and Harry C. Triandis. 1989. "Effects of Culture and Response Format on Extreme Response Style." *Journal of Cross-Cultural Psychology* 20 (3): 296–309.
- Javeline, Debra. 1999. "Response Effects in Polite Cultures: A Test of Acquiescence in Kazakhstan." *Public Opinion Quarterly* 63 (1): 1–28.
- Johnson, Timothy, Patrick Kulesa, Isr Llc, Young Ik Cho, and Sharon Shavitt. 2005. "The Relation Between Culture and Response Styles: Evidence From 19 Countries." *Journal of Cross-Cultural Psychology* 36 (2): 264–77.
- Johnson, Timothy, Diane O'Rourke, Noel Chavez, Seymour Sudman, Richard Warnecke, Loretta Lacey, and John Horm. 1997. "Social Cognition and Response to Surveys Questions among Culturally Diverse Populations." *Survey Measurement and Process Quality*, 87–113.
- Kashima, Yoshihisa. 2009. Culture Comparison and Culture Priming. In Robert S. Wyer, Chi-yue Chiu, and Ying-yi Hong (Eds.), *Understanding Culture: Theory, Research, and Application*: 53–77. New York, NY: Psychology Press.
- Krull, Jennifer L., and David P. MacKinnon. 2001. "Multilevel Modeling of Individual and Group Level Mediated Effects." *Multivariate Behavioral Research* 36 (2): 249–77.
- LaFromboise, Teresa, Hardin L. Coleman, and Jennifer Gerton. 1993. "Psychological Impact of Biculturalism: Evidence and Theory." *Psychological Bulletin* 114 (3): 395–412.

- Lee, Chol, and Robert T. Green. 1991. "Cross-Cultural Examination of the Fishbein Behavioral Intentions Model." *Journal of International Business Studies* 22 (2): 289–305.
- Marin, Gerardo, Raymond J. Gamba, and Barbara V. Marin. 1992. "Extreme Response Style and Acquiescence among Hispanics The Role of Acculturation and Education." *Journal of Cross-Cultural Psychology* 23 (4): 498–509.
- Markus, Hazel R., and Shinobu Kitayama. 1991. "Culture and the Self: Implications for Cognition, Emotion, and Motivation." *Psychological Review* 98 (2): 224–53.
- Mosquera, Patricia M. Rodriguez, Antony S.R. Manstead, and Agneta H. Fischer. 2002. "The Role of Honour Concerns in Emotional Reactions to Offences." *Cognition & Emotion* 16 (1): 143–63.
- Mosquera, Patricia M. Rodriguez, Antony SR Manstead, and Agneta H. Fischer. 2000. "The Role of Honor-Related Values in the Elicitation, Experience, and Communication of Pride, Shame, and Anger: Spain and the Netherlands Compared." *Personality and Social Psychology Bulletin* 26 (7): 833–44.
- Nisbett, Richard E., and Dov Cohen. 1996. *Culture of Honor: The Psychology of Violence in the South*. Boulder, CO: Westview Press.
- Paulhus, D. L. (1991), Measurement And Control Of Response Bias. In John P. Robinson, Phillip R. Shaver, and Lawrence S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes. Measures of Social Psychological Attitudes, Vol. 1*: 17-59. California: Academic Press.
- Preacher, Kristopher J., Michael J. Zyphur, and Zhen Zhang. 2010. "A General Multilevel SEM Framework for Assessing Multilevel Mediation." *Psychological Methods* 15 (3): 209–33.

- Redfield, Robert, Ralph Linton, and Melville J. Herskovits. 1936. "Memorandum for the Study of Acculturation." *American Anthropologist* 38 (1): 149–52.
- Reynolds, Nina, and Anne Smith. 2010. "Assessing the Impact of Response Styles on Cross-Cultural Service Quality Evaluation: A Simplified Approach to Eliminating the Problem." *Journal of Service Research* 13 (2): 230–43.
- Rodriguez Mosquera, Patricia M., Agneta H. Fischer, Antony SR Manstead, and Ruud Zaalberg. 2008. "Attack, Disapproval, or Withdrawal? The Role of Honour in Anger and Shame Responses to Being Insulted." *Cognition and Emotion* 22 (8): 1471–98.
- Rorer, Leonard G. 1965. "The Great Response-Style Myth." *Psychological Bulletin* 63 (3): 129–56.
- Ryder, Andrew G., Lynn E. Alden, and Delroy L. Paulhus. 2000. "Is Acculturation Unidimensional or Bidimensional? A Head-to-Head Comparison in the Prediction of Personality, Self-Identity, and Adjustment." *Journal of Personality and Social Psychology* 79 (1): 49–65.
- Schwarz, Norbert, Daphna Oyserman, and Emilia Peytcheva. 2010. Cognition, Communication, and Culture: Implications for the Survey Response Process. In Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith (Eds.), *Survey methods in multinational, multiregional and multicultural context*: 177–190. Hoboken, NJ: Wiley
- Semin, Gün R. 2000. "Agenda 2000—Communication: Language as an Implementational Device for Cognition." *European Journal of Social Psychology* 30 (5): 595–612.

- . 2009. “Language, Culture, Cognition: How Do They Intersect?” In Robert S. Wyer, Chi-yue Chiu, and Ying-yi Hong (Eds.), *Understanding Culture: Theory, Research, and Application*: 259–70. New York, NY: Psychology Press.
- Shiomi, Kunio, and Robert Loo. 1999. “Cross-Cultural Response Styles on the Kirton Adaption-Innovation Inventory.” *Social Behavior and Personality: An International Journal* 27 (4): 413–19.
- Smith, Peter B. 2004. “Acquiescent Response Bias as an Aspect of Cultural Communication Style.” *Journal of Cross-Cultural Psychology* 35 (1): 50–61.
- Stening, Bruce W., and James E. Everett. 1984. “Response Styles in a Cross-Cultural Managerial Study.” *The Journal of Social Psychology* 122 (2): 151–56.
- Stewart, Frank Henderson. 1994. *Honor*. Chicago, IL: University of Chicago Press.
- Tomasello, Michael. 2003. The Key Is Social Cognition. In Dedre Gentner and Susan Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought*: 47–57. Cambridge, MA: The MIT Press.
- Uskul, Ayse K., Daphna Oyserman, and Norbert Schwarz. 2010. Cultural Emphasis on Honor, Modesty, or Self-Enhancement: Implications for the Survey Response Process. In Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith (Eds.), *Survey methods in multinational, multiregional and multicultural context*: 191-202. Hoboken, NJ: Wiley
- Vaerenbergh, Yves Van, and Troy D. Thomas. 2012. “Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies.” *International Journal of Public Opinion Research* 25 (2): 195–217.

- Van Hemert, Dianne A., Fons JR Van de Vijver, Ype H. Poortinga, and James Georgas. 2002. "Structural and Functional Equivalence of the Eysenck Personality Questionnaire within and between Countries." *Personality and Individual Differences* 33 (8): 1229–49.
- Warnecke, Richard B., Timothy P. Johnson, Noel Chávez, Seymour Sudman, Diane P. O’rourke, Loretta Lacey, and John Horm. 1997. "Improving Question Wording in Surveys of Culturally Diverse Populations." *Annals of Epidemiology* 7 (5): 334–42.
- Yang, Yongwei, Janet A. Harkness, Tzu-Yun Chin, and Ana Villar. 2010. "Response Styles and Culture." In Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith (Eds.), *Survey methods in multinational, multiregional and multicultural context*: 203–23. Hoboken, NJ: Wiley
- Zax, Melvin, and Shigeo Takahashi. 1967. "Cultural Influences on Response Style: Comparisons of Japanese and American College Students." *The Journal of Social Psychology* 71 (1): 3–10.
- Zhao, Xinshu, John G. Lynch Jr., and Qimei Chen. 2010. "Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis." *Journal of Consumer Research* 37 (2):



Appendix 4.1 Question wordings and response options.

5-point scale questions

D16a. For others to know me as I really am, it is important for them to know that I am (religion/ have no religion/ am atheist/ other).

D16b. Being (religion/ have no religion/ am atheist/ other) is a major factor in my social relationships.

D16c. I identify with other (religion/ have no religion/ am atheist/ other)s.

D16d. (religion/ have no religion/ am atheist/ other)s are not respected by the broader American society.

F5. The United States is a land of equal opportunity.

F6. Income differences in the United States are too large.

F7. Immigrants are generally good for the U.S. economy.

F8. Immigrants increase crime rates.

F9. Immigrants make America more open to new ideas and cultures.

F10. And last in this set, the government should give a special tax break to parents who send their children to religious or private schools instead of the public schools.

H8. Since 9-11, as an Arab (Chaldean) I worry more about the future facing me and my family here in the US.

H9. To what extent do you agree or disagree with this statement: I feel at home in America.

H16. The Arab American community is doing all it can to assist in the fight against terror.

Response scale for questions D16a-H16 is strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree.

3-point scale questions

C35A. How well different organizations serve community needs...Churches

C35B. Mosques

C35C. Village or town clubs, such as Ramallah or Bint Jebail?

C35D. Business and professional associations

C35E. Civil liberties and anti-discrimination groups

C35F. Arab and Chaldean social service agencies

C35G. Local Arab media

Response scale for questions C35a-C35f is very effective, somewhat effective, or not effective.

C36A. First, think about American news coverage of religion and religious people.

C36B. Next, think about American news coverage of Islam and Muslims.

C36C. Next, American news coverage of Arab/Chaldean Americans.

C36D. Next, American news coverage of Israelis.

C36E. Next, American news coverage of Palestinians.

Response scale for question C36A-C36E is biased in favor, balanced, or biased against.

Appendix 4.2 Proportions of Extreme Responses for Each Question by Acculturation Level.

	Years in the U.S.		Parent born in U.S.		Home English only		Watch Arabic TV		Interaction with other race	
	0-23 years	23+ years	Both outside U.S.	At least one in U.S.	Other language	English only	Not watch	Watch	Never	At least once
<u>5-point scale</u>										
D16a	0.38	0.31	0.36	0.24	0.36	0.26	0.29	0.39	0.34	0.37
D16b	0.26	0.21	0.25	0.16	0.26	0.16	0.20	0.27	0.24	0.25
D16c	0.27	0.22	0.24	0.22	0.26	0.17	0.20	0.28	0.25	0.25
D16d	0.24	0.18	0.21	0.22	0.21	0.20	0.19	0.23	0.19	0.23
F5	0.48	0.47	0.48	0.44	0.47	0.49	0.49	0.45	0.48	0.44
F6	0.33	0.27	0.29	0.32	0.30	0.29	0.27	0.32	0.30	0.29
F7	0.39	0.30	0.36	0.22	0.37	0.24	0.31	0.38	0.35	0.34
F8	0.21	0.19	0.20	0.19	0.20	0.17	0.20	0.19	0.21	0.18
F9	0.39	0.35	0.38	0.31	0.38	0.32	0.32	0.42	0.36	0.38
F10	0.25	0.22	0.23	0.26	0.23	0.23	0.20	0.26	0.24	0.24
H8	0.29	0.24	0.27	0.22	0.28	0.24	0.23	0.30	0.22	0.32
H9	0.36	0.59	0.45	0.69	0.43	0.69	0.60	0.37	0.46	0.48
H16	0.35	0.30	0.35	0.16	0.35	0.21	0.25	0.39	0.34	0.30
<u>3-point scale</u>										
C35a	0.57	0.49	0.53	0.54	0.51	0.60	0.58	0.49	0.53	0.51
C35b	0.42	0.43	0.44	0.36	0.43	0.41	0.43	0.44	0.44	0.41
C35c	0.36	0.38	0.38	0.31	0.38	0.33	0.36	0.38	0.39	0.36
C35d	0.35	0.35	0.35	0.31	0.33	0.40	0.39	0.32	0.34	0.36
C35e	0.38	0.37	0.38	0.37	0.38	0.38	0.36	0.41	0.34	0.42
C35f	0.43	0.50	0.49	0.35	0.48	0.41	0.47	0.47	0.47	0.47
C35g	0.38	0.45	0.42	0.37	0.44	0.32	0.39	0.44	0.40	0.43
C36a	0.41	0.38	0.39	0.43	0.38	0.42	0.39	0.39	0.31	0.52
C36b	0.54	0.49	0.51	0.54	0.51	0.56	0.49	0.54	0.43	0.65
C36c	0.47	0.35	0.40	0.44	0.39	0.46	0.39	0.43	0.29	0.56
C36d	0.77	0.76	0.76	0.81	0.75	0.80	0.74	0.79	0.70	0.83
C36e	0.71	0.70	0.69	0.78	0.69	0.76	0.71	0.70	0.64	0.76

## **Chapter 5**

### **Conclusion**

The purpose of this dissertation was to explore some sources of response style bias, a unique measurement error associated with rating scales. This dissertation had three primary objectives. The first objective was to evaluate the data collection mode effect on response style, and the interaction effect between mode and race/ethnicity on response style (Chapter 2). The second objective was to investigate the response style bias resulting from the use of an item specific scale and compare it to the response style bias for an agree/disagree Likert scale (Chapter 3). The third objective was to explore the relationship between acculturation and response style among an ethnic minority group (Chapter 4). The ultimate hope, of course, was to provide some practical applications as well as theoretical contributions to survey research.

Chapter 2 examines the mode effect on acquiescent response style (ARS) and extreme response style (ERS). More specifically, it compares the differences in response style between face-to-face and Web surveys using the 2012 American National Election Studies (ANES). The findings from this study imply that the Web survey elicits less ARS and ERS than the face-to-face survey. These findings conform to the theoretical expectation of both types of response styles. The impression management concerns motivate respondents to provide acquiescent responses since these are seen as more polite. In addition, giving extreme answers is seen as more sincere because they are easier to interpret than intermediate responses. The presence of interviewer in the face-to-face mode makes the respondents more likely to present a more polite and sincere image and hence use more ARS and ERS.

Chapter 2 also finds that the mode effect is larger for black than white respondents. Some previous studies have found that racial and ethnic minorities show more ARS and ERS than do white respondents (Bachman and O'Malley 1984a; Bachman and O'Malley 1984b). Chapter 2 shows that, although black respondents still reveal more ERS than white respondents in the face-to-face survey, the difference in incidence of ERS between race groups disappears in the Web survey. This is an important finding in that our previous knowledge of the relationship between race/ethnicity and response style may only apply to the face-to-face survey. Whether or not the same relationship holds in a different mode may be a fruitful line of future research.

Chapter 3 expands the research on ERS to consider differences between the agree/disagree (A/D) scale and the item-specific (IS) scale. The IS scale was developed mainly out of the concern about ARS bias in surveys using A/D scales. Although previous research has shown that the ARS is reduced when the IS format is used, not much is known about ERS in the context of IS scales. This study examines a within-subject and between-subject experiment embedded in the 2012 ANES. The result shows that ERS exists in the context of both A/D and IS scales, although the patterns are different. The ERS for the A/D scale reveals a contrast between the two endpoints and the midpoint of the rating scale, while the ERS for the IS scale shows a contrast between the two endpoints and the three options in between. The longitudinal design of the experiment allows for the examination of the stability of ERS within respondents over time. The findings indicate that, although the scale format changes across the two waves of interview, the ERS remains stable. This suggests that ERS, or response style in general, is primarily determined by some respondent-dependent characteristics rather than question formats.

Chapter 4 studies the relationship between acculturation and ERS using a sample of Arab Americans. Researchers believe that the various cultural backgrounds of respondents are

reflected in their response behaviors and that response style is one of the culturally influenced behaviors. Arabic culture is an honor-based collectivist culture (Uskul, Oyserman, and Schwarz 2010). Such a culture promotes a decisive and assertive answer, such as the extreme response, as a way of showing one's unambiguous attitude and standing, an important quality in an honor-based society. The results confirm these hypotheses. They show that, as the level of acculturation increases, respondents become less likely to select the extreme responses. In contrast, less acculturated Arab American respondents are more likely to endorse the extreme responses. This study also finds that the language of interview mediates the relationship between acculturation and ERS.

This dissertation was a first attempt to investigate several previously unexplored issues related to response style. Findings from this study will inform the design and implementation of surveys using rating scales. The complete elimination of response style is a daunting task. It may require that survey researchers reconfigure survey design and data collection paradigms in ways that exceed our current imagination. I deeply hope that this dissertation will pave the road for future research on response style, and survey measurement more broadly.

## References

- Bachman, Jerald G., and Patrick M. O'Malley. 1984a. "Black-White Differences in Self-Esteem: Are They Affected by Response Styles?" *American Journal of Sociology* 90 (3): 624–39.
- . 1984b. "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles." *Public Opinion Quarterly* 48 (2): 491–509.
- Uskul, Ayse K., Daphna Oyserman, and Norbert Schwarz. 2010. "Cultural Emphasis on Honor, Modesty, or Self-Enhancement: Implications for the Survey-Response Process." *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, 191–201.