8-2011

Is Less More & More Less…? The Effect of Two Types of Interviewer Experience on "Don't Know" Responses in Calendar and Standardized Interviews

Ipek Bilgen
University of Nebraska–Lincoln, ipekbilgen@huskers.unl.edu

IS LESS MORE & MORE LESS…? THE EFFECT OF TWO TYPES OF INTERVIEWER EXPERIENCE ON "DON'T KNOW" RESPONSES IN CALENDAR AND STANDARDIZED INTERVIEWS

by

Ipek Bilgen

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Survey Research and Methodology

Under the Supervision of Professor Robert F. Belli

Lincoln, Nebraska

August, 2011

IS LESS MORE & MORE LESS…? THE EFFECT OF TWO TYPES OF INTERVIEWER EXPERIENCE ON "DON'T KNOW" RESPONSES IN CALENDAR AND STANDARDIZED INTERVIEWS

Ipek Bilgen, Ph.D.

University of Nebraska, 2011

Adviser: Robert F. Belli

Among interviewing context factors, the level of interviewer experience has been observed to be associated with item nonresponse rates in surveys (Singer et al., 1983, Bailar et al., 1977; Pickery & Loosveldt, 1998). The findings regarding the direction of this association, however, are equivocal. This dissertation addresses competing theories behind the relationship between interviewer experience and item nonresponse. The explored experience types are general interviewer experience, *gained via survey administration during a lifetime*, and within-study interviewer exposure, *gained during administration of a particular study fielding period*. Item nonresponse was measured via respondents' *"don't know"* responses.

To date, methodological studies examining the relationship between interviewer experience and data quality focused on standardized interviews. As the interviewing technique—standardized or flexible—relates to data quality, this dissertation discusses the relationship between interviewer experience, exposure, and item nonresponse in both conventional standardized and flexible calendar interviews.

Participants sampled from the 2001 Panel Study of Income Dynamics study were interviewed via telephone. This dissertation used a random sample of these interviews to examine the study-relevant verbal behaviors used by both interviewers and respondents during the question administration process in 165 calendar and 162 standardized interviews. The interviewer and respondent behaviors studied are: *1) Interviewer deviation from conventional ideals*, *2) Interviewer and respondent interpersonal dynamics*, and *3) Interviewer and respondent retrieval strategies.*

Overall, interviewer experience and exposure are positively associated with item nonresponse in both standardized and calendar interviews. The inclusion of the three sets of verbal behaviors moderated this relationship. The association between interviewer and respondent behaviors and item non-response changed depending on when they were used (early versus later interviews), who they were used by (experienced versus inexperienced interviewers), and the interviewing method used (calendar versus standardized interviews). Additionally, the differences in item non-response probabilities, due to the differential use of interviewer behaviors among interviewers with diverse experience levels, were significantly smaller in calendar interviews than in standardized interviews.

ACKNOWLEDGEMENTS

I feel very fortunate to be surrounded by great friends, family, classmates, and professors who were not only very supportive during my dissertation process; they also made the experience more fulfilling. First and foremost, I would like to thank my supervisor Robert F. Belli for his support and guidance on every stage of my dissertation and the PhD process. Over the years, he was not only a mentor but he became like a family member. Our collaborative research inspired some of the main ideas in this dissertation. I also greatly appreciate Bob's contribution to this dissertation by making the Life-Course Health and Economic Measures (HEM) study data and materials available to me.

This dissertation could not have completed without contributions from several institutions and funders. I would like to thank the Charles Cannell Fund in Survey Methodology which funded the verbal behavior data collection for this dissertation research. I also would like to acknowledge the National Institute on Aging and National Institute of Child Health and Human Development for funding the Life-Course Health and Economic Measures (HEM) study used in this dissertation. Also, I specifically would like to thank the Gallup Research Center and Gallup Organization for funding my MS and PhD degrees in the Survey Research and Methodology Program (SRAM) at the University of Nebraska-Lincoln. Their funding made it possible for me to complete my education and training leading to this dissertation. My time in the SRAM program is the stepping stone of my survey research career as I learned a tremendous amount and have grown intellectually during my time in the program.

In addition, I thank all my dissertation supervisory committee members Allan McCutcheon, Janet Harkness, Kristen Olson, and Lesa Hoffman. Each and every one of them helped me in different ways and in different phases of this study and my PhD. I very much appreciate their feedback, ideas, and support during the dissertation planning, data collection and writing.

I especially would like to thank Lesa Hoffman for her guidance, dedication, and support during the dissertation data analyses planning and execution. I very much appreciate her patience and insightful responses during my data analyses "crises." I feel very fortunate to have had Lesa as an instructor as I learned so much from her. This dissertation also benefited from each and every class I took from her.

Furthermore, my collaborative research with Kristen Olson is one of the inspirational sources for this dissertation. The courses I took from her, her recent published work, and the two papers we wrote together all contributed to the key ideas and theories provided and developed in this dissertation. During the dissertation process, she has helped me by providing valuable comments and feedback, but also has challenged me with her ingenious ideas and questions. For all these reasons, I am very thankful.

I also thank the graduate and undergraduate students (Clarissa Steele, Hakan Solak, Jamie Snell, William England, Ayisha Sydnor, and Lindsey Johnson) who were involved in the verbal behavior production process. The student coders not only collected the dissertation data, but also provided insightful comments which improved the coding scheme. Special thanks to Clarissa Steele for her commitment and support during the verbal behavior production. She was very helpful not only at the early stages of the

coding scheme development, but also during the coding production process and with editorial revisions of some of the sections in this dissertation.

My friends and colleagues at SRAM made my PhD experience enjoyable and fun. I especially would like to thank Mathew Stange, Ana Villar, Kathleen Kephart, Rene Bautista, Yue Zoe, Debbie Miller, Tarek Albaghal, Olena Kaminska, Lindsey Witt-Swanson, and Barb Rolfes for their encouragement and emotional support. During my time at SRAM, I learned so much from each and every one of them. In addition, I very much appreciate all the comments and revisions on the drafts of my dissertation provided by friends and colleagues.

I also would like to thank Vincent Ryan for always knowing how to cheer me up when I felt tired or discouraged during my dissertation process. His patience and great cooking skills made the critical moments of the PhD process manageable.

Last but not least, my heartiest gratitude goes to my mother, father, and sister. This dissertation or my PhD could not have happened without their endless love, support, and encouragement. So, thank you for always believing in me and thanks for being my biggest cheerleaders!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION

1. Background, Significance, and Specific Aims

Interviewers are the link between researchers and respondents in interviewer-administered surveys. They are usually asked to complete multiple complicated tasks, which from time-to-time require multi-tasking. Interviewers' duties during sampling include locating addresses and constructing sampling frames[1], contacting sampling units, and screening households to find eligible respondents[2]. Interviewers also play a role in obtaining respondent cooperation, as they are expected to convince sampled individuals to cooperate in the survey and motivate respondents to continue and finish the survey once respondents agree to cooperate. During the questionnaire administration process, interviewers' tasks include introducing the survey and questionnaire to respondents, helping respondents to learn their role in the survey, administering the questionnaire, responding to respondents' questions, confusions, and concerns, and occasionally negotiating with respondents regarding their responses. In addition, interviewers are also expected to troubleshoot computer problems while administering computer-assisted interviews (CAI), record responses, and deliver the data to survey organizations or directly to researchers. Hence, interviewers play a crucial role during multiple stages of the survey lifecycle in interviewer-administered surveys (Groves et al., 2004). The multifaceted role that interviewers play explains why they have been sometimes referred

[1] This is the case in face-to-face surveys, which use area-probability sampling.
[2] This is the case in surveys where the household is the sampling unit.

as "the agents of the researchers" in the survey research literature (Biemer & Lyberg, 2003; Fowler & Mangione, 1990; O'Muircheartaigh, 1997; Kennickell, 2002).

While performing these wide-ranging and complex tasks, interviewers may introduce different types of errors in different stages of the survey life cycle (Biemer & Lyberg, 2003; Groves, 1989; Lessler & Kalsbeek, 1992). These include coverage, non-response, measurement, and processing errors that contribute to total survey error (Groves, 1989). Coverage error[3] due to interviewer variation and bias could occur while locating and constructing sampling frames and enumerating household members (Boyd & Westfall, 1955; Bailar et al., 1977; Groves, 1989; Groves et al., 2004; Lessler & Kalsbeek, 1992; Tourangeau et al., 1997). Interviewer variation and bias can also influence response rates and non-response error, which occurs when there is a discrepancy between respondents and non-respondents on any statistic of interest (Groves, 1989). Interviewers can be one of the main causes of unit non-response during contacting and gaining cooperation of sampled individuals (Boyd & Westfall, 1955; Durrant, Groves, Staetsky, & Steele, 2010; Groves & Couper, 1998; Hox & De Leeuw, 2002; Merkle & Edelman, 2002; O'Muircheartaigh & Campanelli, 1999; Pickery, Loosveldt, & Carton, 2001). Both measurement error[4] and item non-response[5] may occur during the administration of the questionnaire (Beatty & Herrman, 2002; Biemer & Lyberg, 2003; Brick & Kalton, 1996; Groves, 1989; Groves et al., 2004; Lessler & Kalsbeek, 1992). Lastly, processing error occurs while interviewers record the answers received from respondents (data entry) and while editing interviews prior to submitting

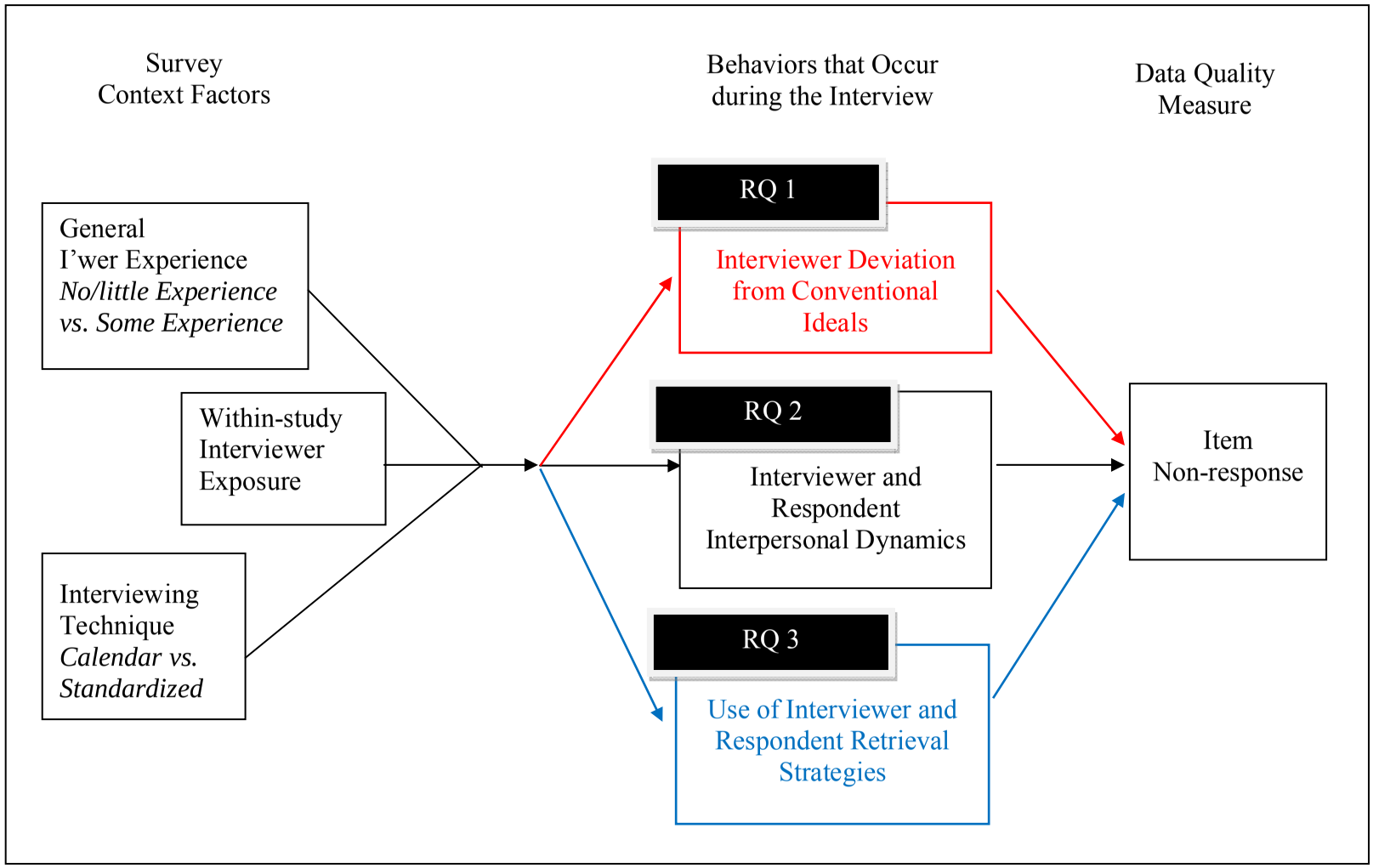
[3] Coverage error: The discrepancy between the "target population" and the "sample frame population" (Groves, 1989; Lessler & Kalsbeek, 1992).
[4] Measurement Error: The discrepancy between respondents' behaviors and their survey responses (Groves, 1989; Biemer & Lyberg, 2003).
[5] Item Nonresponse: Failure to obtain responses from part of the questionnaire (Biemer & Lyberg, 2003)

them for further processing (Boyd & Westfall, 1955). In my dissertation, I will focus on the interviewer-related errors that occur during the administration of the questionnaire in the collection of factual information. Hence, the scope of my dissertation is to concentrate on interviewer-related item non-response as a measure of data quality, particularly in autobiographical respondent retrospective reports.

Figure 1.1 outlines a proposed model that illustrates the interactive effect of two sets of factors—three interviewer-related context factors and relevant verbal and non-verbal behaviors that occur during the survey interview administration—on item nonresponse. The purpose of my dissertation will be to examine all factors of this model and how all of the factors work together in a survey interview while impacting item non-response. A set of methodological literature has explored some of the pieces of this model, but my aim is to explore the model in its entirety and bring diverse types of methodological literature together.

The following section in this chapter introduces the terminology I will use throughout my dissertation. The literature review that follows illustrates what aspects of this model have already been explored and which aspects still need to be explored. Whereas **Chapter I** focuses on a summary of what is studied in this dissertation and how this all fits into the survey research literature; the detailed literature review in **Chapter II** provides a more comprehensive picture of the competing theories, the empirical findings, and the missing pieces in the earlier studies. For the purposes of clarity, previous to explaining the theoretical framework, **Chapter III** introduces a description of the data sources, verbal behavior data collection, and the details regarding the construction of the proposed measures that I use in my study. Next, **Chapter IV** illustrates the theoretical

Figure 1.1 The Big Picture – Research Questions



framework and, accordingly, provides hypotheses regarding the associations in the model illustrated in Figure 1.1. **Chapter V** provides and discusses the results regarding the relationship between the survey context factors and the behaviors of interest that occur during the interviews. **Chapter VI** illustrates and examines the results regarding the whole picture illustrated in Figure 1.1. Lastly, **Chapter VII** illustrates the contribution of this dissertation to the scientific community, provides a summary discussion regarding the findings illustrated in earlier chapters and future work related to the study, and illustrates the limitations of my dissertation.

1.1. Background: Introduction to the Dissertation Terminology

As early as 1957, Kahn and Cannell argued that sources of data quality due to interviewers are a function of 1) *Interviewers' fixed characteristics* such as race, gender, age, education, socio-economic status, and experience, 2) *Psychological factors* including interviewers' expectations, perceptions, attitudes, and motives, and 3) *Behavioral factors,* by which they meant interviewer's behavioral reflections of his/her background and psychological factors via communication. Later on, interviewers' fixed characteristics were illustrated to relate to data quality during administration of the survey because of the survey actors' (i.e., interviewers and respondents) expectations and perceptions, their social distance – i.e., social status differences between the survey actors – (Hyman et al., 1954; Weiss, 1968-69; Sudman et al., 1977; Schuman & Presser, 1981), and social desirability (Dohrenwend et al., 1968; Hughes et al., 2002; Chromy et al., 2005). In addition, it has been shown that fixed interviewer characteristics such as race, sex, age, education, and experience have a significant impact on the differences in interviewer behaviors and how both interviewer and respondent interactions are shaped during the

survey administration process (Bradburn, Sudman, & Associates, 1979; Lepkowski, Siu, & Fisher, 2000; Olson & Peytchev, 2007; van der Zouwen, Dijkstra, & Smit, 1991).

Among all the interviewer-related factors that impact data quality, one specific fixed interviewer characteristic—interviewer experience—specifically stands out in the literature as a research gap. It is quite clear that interviewer experience is a potential predictor of data quality (Cleary, Mechanic, & Weiss 1981; Singer, Frankel, & Glassman, 1983, Bailar, Bailey, & Stevens 1977; O'Muircheartaigh & Campanelli 1998; Cannell, Marquis, & Laurent 1977; Hughes et al., 2002; Chromy et al., 2005). However, the findings on the impact of interviewer experience on data quality are equivocal and the mechanisms that drive this relationship are not very well understood (Groves et al., 2004; Olson & Bilgen, 2011). Accordingly, I aim to disentangle and empirically test the theories that may explain the relationship between interviewer experience and item non-response (an indicator of data quality).

Three types of interviewer experience have been introduced in the literature. The oldest measure is *general interviewer experience,* which is interviewers' survey interviewing experience over their lifetime (Berk & Bernstein, 1988; Bradburn, Sudman, & Associates, 1979; Cleary, Mechanic, & Weiss, 1981; Goudy & Potter, 1975-76; Hill & Hall, 1963; Kennickell, 2007; Singer, Frankel, & Glassman, 1983). The second measure is *survey organization-specific interviewer experience*, which has been described as interviewing experience gained during working in one survey research organization (Bailar, Bailey, & Stevens, 1977; Chromy et al., 2005; Hughes et al., 2002; Kennickell, 2007; Olson & Bilgen, 2011; Olson & Peytchev, 2007). The *survey organization-specific experience* description also includes the experience that is gained by interviewers via

conducting the same questionnaire or study in different rounds and years of the study (i.e., while conducting the same survey in a longitudinal panel study). The most recently explored type of experience is *within-study interviewer experience,* which has been identified as the interviewer experience gained during the administration of a study in one particular survey fielding period (Hughes et al., 2002; Olson & Peytchev, 2007; Pickery & Loosveldt, 2001).

In my dissertation, I focus on the "general interviewer experience" and "within-study interviewer experience," and for purposes of clarity, I call these variables of interest *general interviewer* ***experience***[6] and *within-study interviewer* ***exposure***, respectively. Even though I do not examine *survey organization-specific interviewer experience,* I point out the previous findings and theories regarding this type of experience and refer it as *interviewer* ***experience within organization*** throughout my dissertation.

1.2. How does the explored model relate to the findings in the literature?

Some authors mention that researchers used to have a "common belief" that experienced interviewers would perform better than inexperienced interviewers because they thought "practice makes perfect, if not better" (Bailar, Bailey, & Stevens, 1977; Cannell, Marquis, & Laurent, 1977; O'Muircheartaigh, 1977; Olson & Bilgen, 2011; Singer, Frankel, & Glassman, 1983). However, research has shown that the effect of any type of interviewer experience on survey error was not as simple as believed, such that empirical evidence is inconclusive. The survey research literature contains numerous studies on the impact of both *general* and *survey organization-specific interviewer experience* on data

[6] Please note that throughout my dissertation ***general interviewer experience*** refers to interviewer standardized interviewing experience gained during an interviewer's lifetime period.

quality; however, the empirical findings in these studies are mixed (Bailar, Bailey, & Stevens, 1977; Berk & Bernstein, 1988; Booker & David, 1952; Cleary, Mechanic, & Weiss, 1981; Chromy et al., 2005; Eyerman, Odom, Wu, & Butler, 2002; Gales & Kendall, 1957; Kennickell, 2002; Pickery & Loosveldt, 2001; Singer, Frankel, & Glassman, 1983; Hughes et al., 2001; 2002; O'Muircheartaigh & Campanelli, 1998; Olson & Bilgen, 2011; Peeks, 2005; Stember & Hyman, 1949; Tu & Liao, 2007). The inconsistencies in these findings regarding the association between interviewer experience and data quality may be due to competing mechanisms interfering with each other, thus cancelling out or decreasing the effects of one another. I aim to disentangle the different mechanisms that may play a role in this relationship via behavior coding in my dissertation.

Only a few studies have examined the relationship between interviewer exposure and data quality. Even fewer explored the effects of exposure on item non-response. These studies provide more consistent results and found that interviewers with higher exposure obtain lower quality of reports (Cannell, Marquis, & Laurent, 1977; Hughes et al., 2002), have lower interview length (Olson & Peytchev, 2007), and differ in their use of behaviors (van der Zouwen, Dijkstra, & Smit, 1991). The implication from these findings is that there may be a negative relationship between within-study interviewer exposure and item non-response due to a decrease in interviewer performance with exposure. Moreover, the behaviors of interviewers change as they gain exposure; therefore, the mechanisms behind interviewer behavior and data quality relationship also change during the course of a survey fielding period of a particular study (Cannell, Marquis, & Laurent, 1977; Hughes et al., 2002; Olson & Peytchev, 2007).

General interviewer experience may also be associated with the changes in interviewing behaviors, as the interviewers conduct additional interviews during a survey fielding period. Thus, it is important to take into account the interaction between interviewer experience and exposure while studying the relationship between interviewer experience and data quality. However, only a handful of studies have explored how the experience and exposure interaction shapes interviewer behavior. These studies illustrated that the effect of exposure on interviewer behavior was significantly different for experienced and inexperienced interviewers and made inferences about how these behavioral differences may potentially affect data quality (Bilgen, Belli, & Olson, 2009; Cannell, Marquis, & Laurent, 1977; Olson & Peytchev, 2007). However, to my knowledge, the exposure and experience interaction effect on data quality, let alone item non-response, has not been yet empirically tested. My dissertation is intended to fill this research gap.

From the earlier findings in the literature, it is quite clear that all three types of interviewer experience may introduce systematic biases in the interviewing situation and may affect data quality during administration of both attitudinal and behavioral questions (Bailar, Bailey, & Stevens, 1977; Berk & Bernstein, 1988; Chromy et al., 2005; Cleary, Mechanic, & Weiss, 1981; Eyerman, Odom, Wu, & Butler, 2002; Hughes et al., 2001, 2002; Kennickell, 2002; O'Muircheartaigh & Campanelli, 1998; Olson & Bilgen, 2011; Singer, Frankel, & Glassman, 1983; Stember & Hyman, 1949; Tu & Liao, 2007; Turner, Lessler, & Devore, 1992). The reasons for this systematic error are not as well understood (Groves et al., 2004; Olson & Bilgen, 2011). As response errors are governed by the interviewer-respondent interaction, the missing link in studies exploring any type

of interviewer experience in relation to data quality is the examination of the set of verbal and non-verbal behaviors exercised and utilized differently among experienced and inexperienced interviewers (Cannell et al., 1977; Olson & Peytchev, 2007; Olson & Bilgen, 2011).

Some researchers have argued that interviewer-related bias occurs due to the behaviors interviewers use during data administration, which sometimes may play a mediator role between interviewer characteristics and data quality or between psychological factors of the interviewer and data quality (Bradburn, Sudman, & Associates, 1979; Hill & Hall, 1963; Kahn & Cannell 1957; Olson & Peytchev, 2007; Olson & Bilgen, 2011). Interviewers are trained to employ certain behaviors during the administration of the interview (e.g., providing neutral feedback in standardized interviews). But, given that interviewers are not mechanical and cannot be programmed by the researchers, they may be likely to adapt to the interviewing situation and respondent reactions. Thus, they may deviate from the training and originate new behaviors as they conduct additional interviews (Bradburn, Sudman, & Associates, 1979; Cannell et al., 1977; Cannell & Oksenberg, 1988; Dykema, Lepkowski, & Blixt, 1997; Schaeffer, 1991).

Interviewers with higher levels of experience and exposure have been observed to have lower interview length (in minutes), provide higher rates of improper feedback[7] and lower rates of probing after "don't know" responses, and deviate more from the script (such as omitting a part of the question or the whole question or adding words or phrases

[7] Improper feedback is a verbal behavior which signals interviewers' "approval or disapproval" of a response (Bilgen & Belli, 2010b). While in theory improper feedback has positive and negative connotations; in practice interviewers provide an insignificant amount of improper negative feedback (Belli et al., 2004, Bilgen and Belli, 2010b). Therefore, I will refer this behavior as improper positive feedback in the methods, analyses, and conclusion sections of my dissertation.

to questions) in comparison to interviewers with no or little general experience and exposure (Bradburn, Sudman, & Associates, 1979; Gfroerer, Eyerman, & Chromy, 2002; Groves et al., 2004; Olson & Peytchev, 2007; Pickery & Loosveldt, 2001; van der Zouwen, Dijkstra, & Smit, 1991). These studies concluded that interviewers tend to modify their behaviors as they gain experience and exposure; thus, they are more likely to deviate from the conventional ideals and standardized interviewing protocols in comparison to interviewers with less experience and exposure. One theory regarding the greater likelihood to deviate is that interviewers with higher levels of experience and exposure may be more careless and faster, and therefore obtain lower quality of responses. A different theory is that interviewers learn from previous interviews; thus, they are more efficient in resolving difficulties and obtaining higher quality responses including low item non-response rates (Cannell et al., 1977, 1981; Fowler & Mangione, 1990; Fowler, 1991; Gfroerer, Eyerman, & Chromy, 2002; Groves et al., 2004; Olson & Peytchev, 2007; Pickery & Loosveldt, 2001). This study tests these competing theories by exploring interviewers' *deviation from conventional ideals* as one of the mechanisms that potentially interfere the relationship between interviewer experience, exposure, and item non-response (see Figure 1.1).

Another mechanism that has been explored in the literature is the association between interviewer experience and rapport. Some researchers have found that interviewers with higher levels of general experience have reported engaging in higher levels of respondent rapport in comparison to interviewers with little or no experience (Goudy & Potter, 1975-76), whereas others have found that interviewers with high experience and exposure levels report engaging in lower respondent rapport (Hill & Hall,

1963) and respondent interest (Olson & Peytchev, 2007) in comparison to interviewers with low or little experience and exposure.

There are two competing theories on the role of experience and exposure on rapport. Some suggest that as interviewers gain experience and exposure, they observe and learn in earlier interviews and develop different communicative strategies to establish rapport (i.e., "harmonious relationship" and "friendliness") and facilitate a "productive interpersonal atmosphere" with different types of respondents and become more comfortable while communicating with respondents in later interviews (Collins et al., 2002; Cleary et al., 1981; Olson & Bilgen, 2011; Olson & Peytchev, 2007). Others indicate that as interviewers gain experience and exposure, they start putting more emphases on pace and efficiency in order to complete more interviews in a shorter amount of time (Olson & Bilgen, 2011; Olson & Peytchev, 2007). As the importance of efficiency increases, building rapport while administering the questionnaire may become a low priority (Groves et al., 2004; Olson & Bilgen, 2011).

To my knowledge, none of these studies test the relationship between interviewer rapport behaviors and respondent engagement and interest[8] or loss of engagement and motivation[9], which may potentially have an interactive effect on data quality. As the interviewer and respondent rapport increases, respondents may feel more comfortable to ask for clarifications and express any cognitive difficulties they experience during the questionnaire administration, which may also potentially provide indications regarding to

[8] Respondent behaviors which are measures of respondents' engagement and interest with the study (i.e. willingness to help the interviewer) will be referred as respondent cooperative behaviors throughout my dissertation.

[9] Respondent behaviors that are measures of respondents' loss of interest and engagement (i.e. their willingness to help the interviewer) will be referred as respondent non-cooperative behaviors throughout my dissertation.

data quality, specifically item non-response rates. Thus, this study explores the theories regarding interviewer experience, exposure, and rapport. In addition, the analyses in this dissertation assess the relationship between interviewer rapport, respondent engagement and interest, and respondent expressions of cognitive difficulty and disinterest behaviors via examining *interviewer and respondent interpersonal dynamics* as the second mechanism that may affect the interviewer experience, exposure, and item non-response association (see Figure 1.1).

Over the last few decades, behavior coding studies have explored verbal behaviors that measure respondent cognition and interviewer-respondent communication occurring during the administration of interviewer-administered surveys and their association with data quality indicators (Belli, Lee, Stanford, & Chou, 2004; Belli, Lepkowski, & Kabeto, 2001; Dijkstra, 1987; Fowler & Cannell, 1996; Henson, Cannell, & Lawson, 1976). Studies regarding interviewer experience and exposure have explored interviewer communicative strategies. However, they have not taken into account the possible changes in interviewer behaviors that may affect respondent cognition due to differential levels of experience and exposure. This dissertation also intends to explore this relationship.

Traditionally, researchers have used behavior coding to monitor interviewer performance and examine whether interviewers were following the conventional rules provided during standard basic training, such as using non-directive probing and neutral feedback, appropriate probing to "don't know" responses, reading questions as written, and not failing to ask required questions or any parts of questions (Cannell & Oksenberg, 1988; Mathiowetz & Cannell, 1980). Researchers have also used this technique to detect

respondent cognitive and communicative difficulties during the administration of a survey, such as respondent expressions of difficulty and uncertainty, seeking clarifications, guessed or estimated responses, and interruptions (Fowler, 1992; Fowler & Cannell, 1996; Morton-Williams & Sykes, 1984; Oksenberg, Cannell, & Kalton, 1991; Presser & Blair, 1994). The inspection of cognition and communication in surveys comes into play particularly in the questionnaire testing and assessment stage of a survey (Conrad, Schober, & Dijkstra, 2008; Fowler, 1995; Presser et al., 2004; Willis, 2005). Using verbal behavior coding techniques, researchers can detect problematic questions that may potentially lead to both cognitive and communicative difficulties during the questionnaire administration process (Conrad & Blair, 2004; Drennan, 2003; Oksenberg, Cannell, & Kalton, 1991). In essence, behavior coding has been used as a tool to assess data administration quality and problems with the questionnaire in interviewer-administered surveys.

Examination of interviewer and respondent interaction has also demonstrated that the behaviors that occur during the survey questionnaire administration may significantly impact data quality (Belli & Lepkowski, 1996; Belli, Lepkowski, & Kabeto, 2001; Cannell, Miller, & Oksenberg, 1981; Dykema, Lepkowski, & Blixt, 1997; Schaeffer & Maynard, 1996; van der Zouwen, 2002). Moreover, researchers have also applied behavior coding of interviewer and respondent interactions to examine the standardized versus flexible interviewing controversy (Belli, Lee, Stafford, & Chou, 2004) and to investigate the problems regarding standardization in survey interviews (Houtkoop-Steenstra, 2000; Maynard & Shaffer, 2002; van der Zouwen, 2002). Studies examining

standardized interviewing and how it relates to data quality, specifically to item non-response, have provided mixed results and will be explained further in Chapter II.

Some researchers have studied the conversational versus standardized controversy from a cognitive psychology perspective and have discussed a more specific use of conversational interviewing that extensively utilizes a calendar during the questionnaire administration to facilitate respondents' access to autobiographical memory. This technique has been designated as "Life History Calendar," "Calendar," or "Event History Calendar" interviewing (Axinn, Pearce, & Ghimire, 1999; Belli, 1998; 2000; Freedman et al., 1988; Lepkowski, Sui, & Fisher, 2000; Means et al., 1991). I refer to this technique as *calendar interviewing* throughout this dissertation. Researchers have also compared the use of verbal behaviors and interviewer-respondent interaction in calendar and standardized conventional interviewing and examined the relationship between certain behaviors and data quality to understand which method leads to higher data quality (Belli, Lee, Stafford, & Chou, 2004; Belli, Lepkowski, & Kabeto, 2001; Belli, Shay, & Stafford, 2001; Belli, Smith, Andreski, & Agrawal, 2007; Bilgen & Belli, 2010b; Yoshihama et al., 2005). Because of the importance that interviewing technique—whether standardized or calendar—has an impact on data quality, interviewing technique is a third survey factor that will be explored in my dissertation (see Figure 1.1).

Calendar interviewing has "emerged" in the last decade and has been used in different fields of research to obtain retrospective behavior reports from respondents (Belli & Callegaro, 2009). Researchers have also assessed the quality of retrospective reports when collecting responses using calendar interviews within specific populations such as older (Axinn, Pearce, & Ghimire, 1999; Hurd & Rohwedder, 2009) or younger

individuals (Roberts & Mulvey, 2009), women with low socio-economic status (Yoshihama et al., 2005; Yoshihama, 2009), respondents from different cultures and racial backgrounds (Callegaro, Belli, Serrano, & Palmer, 2007), and different socio-economic status (Hurd & Rohwedder, 2009).

In essence, researchers have investigated the quality of calendar reports across different respondent characteristics (Axinn, Pearce, & Ghimire, 1999; Hurd & Rohwedder, 2009; Yoshihama et al., 2005; Yoshihama, 2009). However, the effect of interviewer characteristics such as race, gender, education, age, socio-economic status, and experience on quality of reports in calendar interviews is an under-researched area. Specifically, studies regarding the research on item non-response are even scarcer. Only one study has examined how interviewer race and racial differences between interviewer and respondent relate to data quality in calendar interviews (Callegaro, Belli, Serrano, & Palmer, 2007). To my knowledge, the association among interviewer experience, exposure, and data quality in calendar interviews has yet to be explored. Also, no studies have examined how interviewer experience and exposure relate to the item non-response differences between calendar and standardized interviews. This dissertation also intends to fill this research gap.

During the collection of behavioral reports, the main behaviors that are examined as data quality indicators are the dimensions of behaviors related to communication and cognition (Belli, Lee, Stafford, & Chou, 2004; Belli & Lepkowski, 1996; Belli, Lepkowski, & Kabeto, 2001; Bilgen & Belli, 2010b; Dykema, Lepkowski, & Blixt, 1997; van der Zouwen, 2002). Interviewer communication behaviors are mainly referred to interviewers' conversational rapport behaviors (Belli, Lepkowski, & Kabeto, 2001;

Dykema, Lepkowski, & Blixt, 1997; Fowler & Cannell, 1996; Houtkoop-Steenstra, 2000; Maynard & Schaeffer, 2002), whereas interviewer cognition behaviors are classified as the use of retrieval probes[10]—the strategies that use the knowledge regarding the structure of autobiographical memory to help respondents recall events more efficiently in behavioral questions (Belli, 1998; Belli, Lee, Stafford, & Chou, 2004; Bilgen & Belli, 2010b). For instance, retrieval probes include strategies that link a contemporaneous event during recall of another event, such as residential change due to graduation, or link similar events that occur earlier or later in time, such as recalling the order of schools a respondent has attended (Belli, 1998).

The studies that have examined differences in interviewer behaviors among interviewers with different experience and exposure levels in administration of factual reports have focused on interviewer communicative behaviors. However, to my knowledge, only Bilgen, Belli, and Olson (2009) have explored the association between interviewer experience, exposure, and interviewer's use of retrieval behaviors. They found that general interviewing experience does not come into play in the majority of retrieval behaviors explored. In addition, they found that the use of interviewer retrieval probes increases during the survey fielding period. One possibility is that interviewers practice and learn these behaviors with each interview they administer; therefore, they gradually use retrieval behaviors more during the fielding period (Bilgen, Belli, & Olson, 2009). It has also been hypothesized that retrieval behaviors aid respondents to recall events more productively (Belli, 1998; Belli, Lee, Stafford, & Chou, 2004; Bilgen & Belli, 2010a). Both interviewer and respondent retrieval strategies have been found to

[10] Probing is the act of an interviewer asking the respondent for different kinds of study-related information.

increase the accuracy of retrospective reports in calendar interviews (Belli, Lee, Stafford, & Chou, 2004), especially while obtaining respondent reports with more difficult histories (Bilgen & Belli, 2010a). The usage of both interviewer and respondent retrieval cues were found to be more trivial, and thus, less beneficial in standardized interviews (Belli et al., 2004).

As far as I know, there are no studies that have explored the relationship between interviewer exposure, retrieval strategies, and item non-response. Therefore, I examine whether or not interviewers with more exposure and experience use retrieval behaviors more productively than interviewers with no or little experience and exposure. Moreover, I assess whether *interviewer and respondent retrieval behavior interaction* is the third mechanism that mediates experience, exposure, and item non-response association in different interviewing methods (see Figure 1.1).

In sum, the studies exploring the relationship among interviewer experience, exposure, and behaviors have indicated that the differential uses of interviewer behaviors among interviewers with differential experience and exposure may affect data quality (Bradburn, Sudman, & Associates, 1979; Cannell et al, 1977; Olson & Peytchev, 2007; Olson & Bilgen, 2011). However, only two studies have tested this theory empirically (Hill & Hall, 1963; Olson & Bilgen, 2011). Hill and Hall (1963) studied interviewer reports of respondent rapport and found interviewers with high levels of general experience report lower scores for respondent rapport and obtain higher levels of item non-response in comparison to interviewers with no or little general experience. However, there are several limitations of this study. First, they used interviewer reports of perceived respondent rapport, and, second, they did not control for the potential

confounding effects of other interviewer and respondent characteristics such as race, sex, education, and socio-economic status.

Olson and Bilgen (2011) also found a significant relationship between interviewer experience and acquiescent responses. They hypothesized that this relationship may be due to pace and rapport differences among interviewers with difference experience levels, though they only had the data to empirically test the pace hypothesis. Even after controlling for differences in pace among interviewers with different experience levels, the relationship between experience and acquiescence rates remained. This relationship indicates that one or more other mechanisms explain the interviewer experience and data quality association. The common limitation in these studies is the lack of data that enables researchers to empirically test possible mediating mechanisms (i.e., measured via different interviewer and respondent behaviors) behind the interviewer experience, exposure, and data quality relationship. This dissertation intends to fill in this research gap. I will use behavior coding as a tool to test mediating mechanisms behind the interviewer experience and item non-response relationship in different stages of the fielding period during the administration of calendar and standardized interviews.

1.3. Specific Aims: Research Questions

The main purpose of my dissertation is to disentangle the mechanisms behind interviewer experience and exposure effects on item non-response in two interviewing techniques: standardized interviewing and calendar interviewing (see Figure 1.1). I will measure these mechanisms via both verbal and non-verbal behaviors used by the interviewers and respondents during the questionnaire administration in a telephone survey experiment. Each research question in my study deals with how each of these sets of behaviors are

differentially used by interviewers with different experience and exposure levels and how these may relate to respondent behaviors and item non-response.

Research Questions #1: Interviewer Deviation from Conventional Ideals

- Overall, do interviewers *with some or several years of general experience* differ in their levels of deviation from conventional ideals in comparison to interviewers *with little or no general experience*?
- Does the effect of *general interviewer experience* on interviewer deviation from conventional ideals change when interviewers gain *exposure* during the survey fielding period?
    - (If yes,) Do these interviewer behavioral changes due to the diversity in *experience* and *exposure* levels cause systematic changes in item non-response?
- Does the effect of experience and/or exposure on deviation behaviors significantly differ by interviewing method (i.e., standardized versus calendar-based interviewing)?
    - (If yes,) Do these interviewer behavioral changes attributable to the diversity in *experience* or *exposure* levels cause systematic differences in item non-response levels in standardized interviews OR in calendar interviews?

Research Questions #2: Interviewer-Respondent Interpersonal Dynamics

- Overall, do interviewers *with some or several years of general experience* differ in providing and obtaining communicative behaviors in comparison to interviewers *with little or no general experience*?

- Does the effect of *general interviewer experience* on interviewer and respondent communicative behaviors change when interviewers gain *exposure* during the survey fielding period?
  - (If yes,) Do these interviewer and respondent behavioral differences due to diversity in interviewer *experience* and *exposure* levels cause systematic changes in item non-response?
- Does the effect of experience and/or exposure on interviewer and respondent communicative behaviors significantly differ by interviewing method (i.e., standardized versus calendar-based interviewing)?
  - (If yes,) Do these behavioral changes (attributable to the diversity in interviewer *experience* or *exposure* levels) have an impact on item non-response levels in standardized OR calendar interviews?

Research Questions #3: Interviewer-Respondent Retrieval Strategies

- Overall, do interviewers *with some or several years of general experience* differ in providing and obtaining retrieval behaviors than interviewers *with little or no general experience*?
- Does the effect of *general interviewer experience* on interviewer and respondent retrieval behaviors change when interviewers gain *exposure* during the survey fielding period?
  - (If yes,) Do these interviewer and respondent behavioral changes attributable to the differences in *experience* and *exposure* levels cause systematic changes in item non-response?

- Does the effect of experience and/or exposure on interviewer and respondent retrieval behaviors significantly differ by interviewing method (i.e., standardized versus calendar-based interviewing)?
    - (If yes,) Do these behavioral differences attributable to the diversity in interviewer *experience* or *exposure* levels affect item non-response in standardized OR calendar interviews?

CHAPTER II: LITERATURE REVIEW

1. Item Nonresponse as a Measure of Survey Data Quality

Item nonresponse arises in surveys when survey participants (i.e., respondents) do not respond to one or more questions during the administration of a questionnaire. The incompleteness due to item non-response in the product data may contribute to error in survey estimates when the respondents who do not answer one or more questions significantly differ from the respondents who provide an answer to these questions (Beatty & Herrmann, 2002; De Leeuw, Hox, & Huisman, 2003; Groves, 1989; Groves et al., 2004; Pickery & Loosveldt, 2001). The error in survey estimates due to item non-response becomes an important concern for the survey researchers as the inferences obtained from these survey estimates will be erroneous. Hence, item non-response error has been acknowledged as one of the components of total survey error (Groves, 1989; Groves et al., 2004) and item non-response measures has been commonly used as data quality indicators in the literature (De Leeuw, Hox, & Huisman, 2003).

De Leeuw, et al. (2003, p. 154) adapted the "reducers" and "measurers" analogy, which is first introduced by Groves (1989, p. 311)[11], and summarized the division of labor among survey researchers who deal with item non-response as *"reducers"* and *"adjusters"*. The authors mentioned that, "reducers" investigate the reasons of item non-response and try to find ways to "reduce" the effects of item non-response on survey inferences before it occurs, whereas the "adjusters" investigates solutions regarding how

[11] Groves (1989, p. 311) uses the analogy of "reducers" and "measurers" of total survey error to provide a division between the researchers who aim to prevent, if not "reduce", the reasons of survey error and who "measure" the components of total survey error (i.e. mean square error).

to deal with item non-response (such as single and multiple imputation methods) after it occurs in surveys (Groves & Couper, 1998; Huisman, 2000; Huisman & Van Der Zouwen, 1999; Little & Rubin, 1987; Marker et al., 2002; Pickery & Loosveldt, 2001; Rubin, 1996). It is not the scope of this dissertation to investigate the latter (i.e., the adjustment methods), nor to find ways to eliminate or reduce item non-response; rather, this dissertation explores potential determinants of item non-response related to general interviewer experience, within-study interviewer exposure, and interviewing technique during the collection of autobiographical reports.

Reasons and types of item nonresponse differ in attitude and behavioral questions (Beatty & Herrmann, 2002; Krosnick, 2002; Krosnick et al., 2002; Pickery & Loosvelt, 1998, 2001; Groves et al., 2004). This dissertation focuses on item nonresponse in behavioral questions, specifically retrospective autobiographical reports. Hence, it is not the scope of this dissertation to examine the "no opinion" responses, as they occur in attitudinal questions. Two different types of item nonresponse – "don't know" responses and explicit refusals–are likely to occur in interviews that seek information on respondents' retrospective autobiographical reports (i.e., seek to collect information regarding respondents' past behaviors). This dissertation specifically focuses on "don't know" responses[12].

It has been hypothesized that in behavioral questions item nonresponse is an outcome of two different psychological routes: 1) cognition, related to the response process concerning respondent's retrieval strategies and cognitive difficulty and 2) communication, related to interpersonal dynamics such as interviewer and respondent

[12] Due to low occurrences in the explored interviews, respondent explicit refusals are not included as a measure of item non-response.

rapport, and respondent cooperation (Beatty & Herrmann, 2002; Schwarz, 1996). The studies that examined item non-response from a cognitive point of view generally focused on respondents' interaction with the questionnaire. These studies have assessed the effect of respondent characteristics (Converse, 1977; Ferber, 1966; Krosnick, 1991; Schuman and Presser, 1981), questionnaire characteristics such as question wording, sensitivity, or position (Shoemaker, Eichholz, & Skewes, 2002; Sudman, Bradburn, & Schwarz, 1996; Tourangeau, Rips, & Rasinski, 2000), and questionnaire branching difficulty (Messmer and Seymour, 1982) on item non-response.

Other researchers looked at item non-response from a communicative point of view and examined the effect of interviewer characteristics, question and questionnaire characteristics such as question sensitivity or questionnaire type, the interactive effect of respondent and interviewer characteristics, and interviewer and respondent interpersonal dynamics on item non-response in interviewer-administered surveys (Pickery & Loosveldt, 2001; Singer, Frankel, & Glassman, 1983; Sudman, Bradburn, & Schwarz, 1996; Tu & Liau, 2007). Lastly, the possible cause of item non-response occurs when interviewers interact with the questionnaire (Chesnut, 2005; De Leeuw et al., 2003). This happens when interviewers fail to ask one or more questions or fail to record the answer as a result of technical difficulty or carelessness (Brick & Kalton, 1996; Beatty & Herrman, 2002). This dissertation focuses on the differential effects of interviewer experience and exposure on item non-response in calendar and standardized interviews. The aim of this dissertation is also to assess whether the interviewer and respondent behaviors, relating cognition and communication, have a mediating/intervening effect on the interviewer experience, exposure and item non-response relationship (see Figure 1.1).

The next sections provide a literature review regarding the different components of the research plan illustrated in Chapter I and Figure 1.1. Section II reviews studies that examine the association between interviewer experience, exposure and item non-response. In order to provide a more comprehensive picture, Section II initially reports the interviewer experience and exposure studies that examine several dimensions of data quality (such as response accuracy, rates of sensitive responses, acquiescence, refusals, don't knows, and no-opinion responses). Section III reviews the literature on behavioral differences among interviewers with different experience and exposure levels. Section IV assesses studies regarding the association between interviewer and respondent behaviors with data quality, whereas the last section (Section V) reviews studies that compare this relationship in calendar and standardized interviews. To my knowledge, the studies on interviewer and respondent behaviors and item nonresponse relationship are relatively sparse. The literature examining this relationship in different interviewing techniques is even sparser. Hence, the last two sections of this chapter focus on several other data quality measures (such as response accuracy and inadequate responses), as well as item non-response measures (such as number of "don't know" responses).

2. Interviewer Experience, Exposure and Survey Data Quality

Interviewers play a vital role in influencing the quality of respondents' answers while they are administering survey questionnaires (Chromy et al., 2005; O'Muircheartaigh & Campanelli, 1998; Singer et al., 1983; Cleary et al., 1981; Borland, 1975). However, previous findings on the relationship between *general interviewer experience* and data

quality[13] provide mixed results. Some researchers found that *general interviewer experience* (i.e., experience over a lifetime) increases data quality by decreasing the number of missing items (Kennickell, 2002) and by obtaining higher rates of behavior reports to sensitive questions such as mental health symptoms (Cleary, Mechanic, & Weiss, 1981) and drinking habits (Singer, Frankel, & Glassman, 1983). Others demonstrated that *general interviewer experience* decreases data quality by obtaining higher rates of acquiescent responses in attitudinal reports (Olson & Bilgen, 2011) and by providing higher rates of "yes" responses to socially desirable questions regarding future behavioral reports (O'Muircheartaigh & Campanelli, 1998). Moreover, Berk and Bernstein (1988) found that *general interviewer experience* increases item non-response during reports of factual questions. One interesting idea is that the relationship between experience and data quality is curvilinear. Specifically, some experience increases data quality more than no experience, whereas greater levels of experience hurt data quality (Fowler & Mangione, 1990; Mathiowetz & Cannell, 1980). However, it is also important to note that some authors have found no relationship between the general interviewer experience and item non-response (Singer, Frankel, & Glassman, 1983).

Similarly, no consensus exists among the findings of the several studies examining the relationship between *interviewer experience within an organization*[14] and data quality. Some studies have found that higher *interviewer experience within an organization* increases data quality, given that experienced interviewers are less likely to

---

[13] The data quality measures in behavioral and attitudinal research differ from each other. Studies exploring interviewer experience in behavioral research mainly focus on response accuracy, rates of sensitive responses, and "don't know" responses. Studies exploring experience in attitudinal questions mainly focus on acquiescence (respondents' tendency to agree) and no-opinion rates. This dissertation only focuses on item non-response data quality measure, specifically "don't know" responses in behavioral questions.

[14] **Interviewer experience within an organization:** Interviewing experience gained while working in one survey research organization

bias responses due to their prior expectations (Stember & Hyman, 1949), and that interviewers who are more experienced conducting the same survey are also more likely to obtain lower acquiescent reports in later waves in comparison to interviewers with no experience with prior waves of the study (O'Muircheartaigh & Campanelli, 1998). In contrast, *interviewer experience within an organization* has been shown to decrease data quality such that experienced interviewers obtained higher levels of item non-response during the administration of sensitive items such as income (Bailar, Bailey, & Stevens, 1977) and sexual attitudes and behaviors (Tu & Liao, 2007), lower rates of behaviors such as numbers of times hospitalized (Cannell, Marquis, & Laurent, 1977), and lower rates of sensitive behaviors such as illegal substance abuse (Chromy et al., 2005; Eyerman, Odom, Wu, & Butler, 2002; Turner, Lessler, & Devore, 1992). Moreover, respondents' depression scores were no different between experienced and inexperienced interviewers in a recent mental health survey (Peeks, 2005).

The relationship between within-study interviewer exposure[15] and data quality has only been examined in a few studies. Cannell, Marquis, and Laurent (1977) conducted one of the pioneer studies that examined data quality change during the survey fielding period. They observed an increase in response errors in later interviews due to interviewer exposure during the study fielding period regardless of the interviewers' experience levels. Similarly, Hughes et al. (2002) found that, in early interviews during the fielding period, interviewers with lower levels of within-study exposure were obtaining higher reports of illegal drug use than interviewers with higher levels of within-study exposure. On the contrary, Pickery and Loosveldt (1998) found no relationship

[15] **Within-study interviewer exposure:** Interviewer experience gained during the administration of the same study in one particular survey fielding period.

between within-study interviewer exposure and "no-opinion" responses. In addition, interviewers with higher within-study exposure were found to obtain shorter interviews[16] (Olson & Peytchev, 2007), provide higher numbers of directive probes[17], which may potentially bias the respondents' answer (van der Zouwen, Dijkstra, & Smit, 1991), were fatigued, acted more careless, and conducted interviews faster in comparison to interviewers with no or little exposure (Cannell & Kahn, 1968; Cannell & Oksenberg, 1988; Fowler, 1991; Olson & Peytchev, 2007). These results indicated a clear behavioral change among interviewers during the survey fielding period.

2.1. Interviewer Experience and Exposure Association with Item Non-Response

To sum up the results above, there is no consistent pattern among studies that examine the relationship between different types of experience and item non-response. The findings regarding *interviewer general experience* and item non-response association are mixed in studies that explore behavioral questions. For instance, Berk and Bernstein (1988) illustrated that experienced interviewers increased item non-response rates by obtaining higher numbers of missing data at a survey regarding health expenses than inexperienced interviewers. However, Kennickell (2002) found that experienced interviewers decreased item non-response rates by obtaining lower "don't know" responses and refusals than inexperienced interviewers. On the contrary to these findings, Singer, Frankel, and Glassman (1983) found no relationship between the general interviewer experience and refusals or "don't know" response rates.

---

[16] In this study, interview length is used as a proxy measure of interviewer pace.

[17] Example: "*Do you remember if it was winter, or…? (In this case, a nondirective way of asking the question can be 'do you remember which season this incident occurred?')" – taken from Bilgen and Belli (2010b), pg.28.*

In addition, studies that examined item non-response rates among sensitive questions (such as income and sexual attitudes and behaviors) found a positive correlation between *interviewer experience within an organization* and item non-response. In other words, experienced interviewers were more likely to obtain refusals and "don't know" responses in questions regarding respondents' sexual attitudes and behaviors (Tu & Liao, 2007) and income levels (Bailar, Bailey, & Stevens, 1977). In addition, Sanchez (1992) explored the skipped/missed question rate among interviewers, which also contributes to item non-response rates in surveys. This study examined interviewers with differential survey-specific organization experience levels among two different survey organizations. For either of these organizations, this study found no difference between experienced and inexperienced interviewers in question skip rates.

Lastly, to my knowledge the studies that explore the relationship between within-study interviewer exposure and item nonresponse is very sparse. I am only aware of one study that explored this relationship in attitudinal questions (Pickery & Loosveldt, 1998) and the authors found no relationship between within-study interviewer exposure and "no-opinion" responses. However, as it has been mentioned earlier, several studies found a negative relationship between interviewer performance (measured via interviewer pace, directive probing, etc.) and within-study interviewer exposure. These findings may suggest that interviewers with higher exposure may potentially increase item non-response rates. However, more research is needed on interviewer exposure and item non-response relationship and whether interviewer performance and deviation from conventional ideals mediate this relationship. This dissertation aims to fill in this gap.

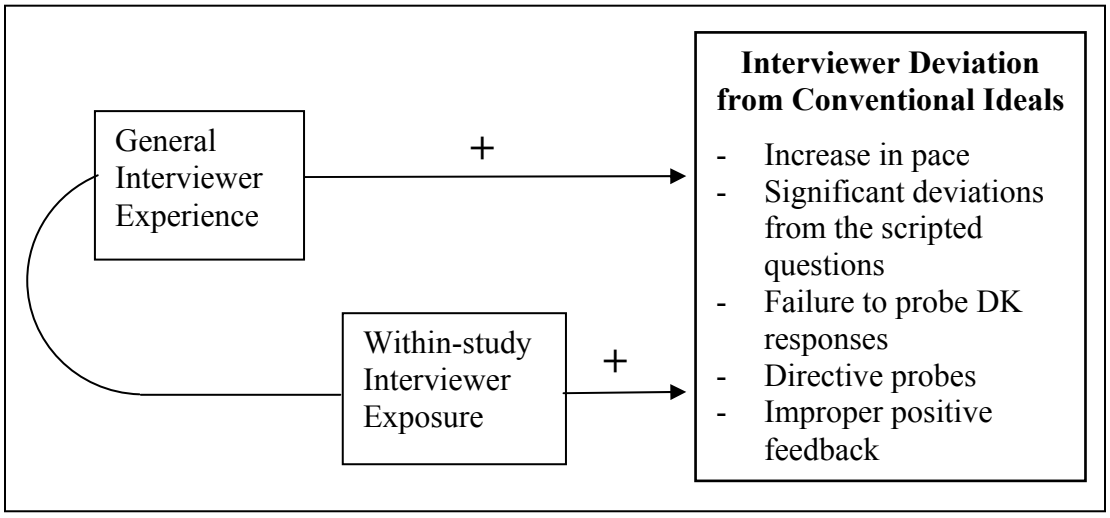
3. Behavioral Differences among Interviewers with Differential Experience and Exposure Levels

The common finding regarding interviewer experience and exposure is that even though in standardized interviews researchers train interviewers to avoid different questionnaire administration, feedback, and probing strategies, variation among interviewers' experiences leads to deviation from intended standardization due to higher rates of directive probing, improper feedback, reading errors, and speech variations (Bradburn et al., 1979; Cleary et al., 1981; Fowler & Mangione, 1990; O'Muircheartaigh & Campanelli, 1998).

Studies exploring the interviewer experience, exposure, and behavior relationship are centered on a broader topic of interviewer variation in administering questionnaires and standardized interviewing. Two sets of behaviors are explored in this line of research: 1- deviations from standardization, training, and protocols and 2- differential rapport behaviors among interviewers with different levels of experience and exposure levels. Studies that are focused on deviation from standardized protocols (i.e., conventional ideals) found that interviewers with higher levels of general experience have higher rates of reading errors, speech variations, improper feedback, and directive probes (Bradburn, Sudman, & Associates, 1979; van der Zouwen, Dijkstra, & Smit, 1991) and are less likely to follow conventional training protocols such as not reading wording exactly as scripted, omitting a part of the question/the whole question, or adding words/phrases to questions (Gfroerer, Eyerman, & Chromy, 2002; Groves et al., 2004) in comparison to interviewers with no or little general experience. Also, interviewers with higher levels of exposure have been observed to obtain shorter interviews (Olson & Peytchev, 2007) and

have higher rates of directive probes (van der Zouwen, Dijkstra, & Smit, 1991) than interviewers with little or no exposure.

Figure 2.1 Relationship between Experience, Exposure and Deviation Behaviors



Both "+" signs in Figure 2.1 illustrates the overall findings in the literature[18] regarding interviewers with more experience and exposure levels deviating more from standardized ideals in comparison to interviewers with little or no experience or exposure levels (Bradburn et al., 1979; Gfroerer et al., 2002; Olson & Peytchev, 2007; van der Zouwen et al., 1991); however, whether and how this deviation influences data quality is unresolved. One theory is that the deviation from conventional ideals can be an indication of a decrease in data quality because interviewers become more careless and the increase in pace does not permit respondents to think through their answers and complete all of the response process steps (Cannell et al., 1981; Fowler, 1991; Gfroerer et al., 2002; Groves et al., 2004; Groves & Magilavy, 1986; Olson & Bilgen, 2011; Olson & Peytchev, 2007; Pickery & Loosveldt, 2001). In contrast, another theory suggests that increases in experience might mean that interviewers become more familiar with what

[18] Please note that Figures 2 through 7 are aimed to illustrate the empirical findings in the literature.

works best for them during the interviews; thus, they are more proficient and efficient at resolving difficulties and obtaining higher quality responses (Cannell et al., 1977; Olson & Bilgen, 2011; Olson & Peytchev, 2007; Pickery & Loosveldt, 2001).

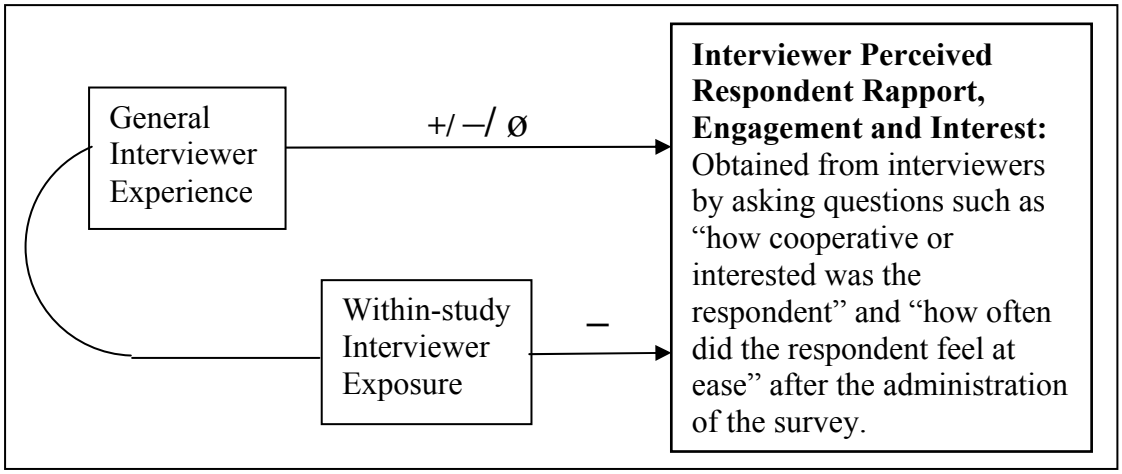
Only two studies (Hughes et al., 2002; Olson & Peytchev, 2007) take into account the interactive effect of both interviewer experience within an organization and within-study interviewer exposure. Hughes et al. (2002) explores the interactive effect of experience within an organization and within-study interviewer exposure on data quality; however, they do not examine how this interaction relates to interviewer behaviors. Olson and Peytchev (2007) take into account the interactive effect of experience and exposure playing a role together on interviewer pace; however, they did not explore how this interaction affects data quality. Olson and Peytchev (2007) also did not have any other interviewer behavior measures such as deviations from the script or failure to probe don't know responses, improper feedback, etc. Therefore, this dissertation explores the interviewer experience and exposure interaction on all of the behaviors illustrated in Figure 2.1 that are measures of interviewer deviation from conventional ideals.

There are equivocal results in the literature regarding the impact of interviewer experience and exposure on respondent rapport, engagement, and interest. Interviewers with higher levels of general experience have been observed to report higher levels of perceived rapport and respondent engagement in comparison to interviewers with little or no experience (Goudy & Potter, 1975-76, indicated via "+" sign in Figure 2.2). However, Hill and Hall (1963) found that interviewers with high general experience[19] have reported lower perceived respondent rapport in comparison to interviewers with low or no general

[19] One limitation of the Hill and Hall (1963) study is that the authors do not specify what they mean by "experienced interviewers" and "interviewers with little or no experience" in their paper.

experience ("—" sign in Figure 2.2). In addition, Olson and Peytchev (2007) found that as interviewers conduct additional interviews during the administration of the same survey (i.e., as the within-study interviewer exposure increases), the interviewer observation of respondent interest decreases (illustrated via "—" sign in Figure 2.2 between interviewer exposure and respondent rapport). Contradictory to previously mentioned findings, they also found that *general interviewer experience* (interviewer experience over a lifetime) has no significant impact on the interviewer reports of respondent interest (illustrated via "Ø" sign in Figure 2.2).

Figure 2.2 Experience, Exposure and Rapport Relationship



One major limitation in all of these studies is that they are not measuring respondent behaviors directly. Rather, they use perceived respondent behaviors that are obtained from interviewers after administering the questionnaire. In these cases, it is impossible to disentangle whether interviewer ratings are due to the "true" respondent behavior differences or to a systematic error due to interviewer experience such that experienced interviewers might be more aware of importance of obtaining respondent

rapport and might adjust their responses accordingly. Also, in these studies the causality chain is broken –In interviewing situations rapport may be the cause of data quality. On the contrary, in these studies data quality may impact the perceived rapport. For instance, if interviewers perceive respondents, who provide higher rates of item-nonresponse, as having lower levels of rapport, then researchers will find a negative relationship between item nonresponse and rapport (the classic 'chicken and egg' story).

In addition, contradicting theories exist on why interviewer experience and exposure impact rapport and respondent motivation. Some authors suggest that as interviewers gain experience and exposure, they start learning their role in the survey game, become more comfortable, and start developing communicative strategies to create a friendly and productive survey interaction environment that facilitates mutual understanding and approval between interviewers and different types of respondents (Collins et al., 2002; Cleary et al., 1981; Olson & Bilgen, 2011; Olson & Peytchev, 2007). Others indicate that as interviewers gain experience and exposure, they become faster (Olson & Bilgen, 2011; Olson & Peytchev, 2007). As the importance of efficiency—i.e., completing more interviews in a shorter amount of time—increases, building rapport while administering the questionnaire may become a low priority for interviewers with higher levels of experience and exposure (Groves et al., 2004; Olson & Bilgen, 2011).

However, none of these studies takes into account that both general experience and within-study interviewer exposure play a role together on how interviewer rapport behaviors change. Also, to my knowledge, no study explores the relationship between experience, exposure, and interviewer rapport and how this affects respondent

cooperation and engagement. Therefore, I also examine the interviewer rapport behaviors while assessing the interviewer experience, exposure, and interviewer-respondent interaction association. Moreover, the more important question for researchers is how the differential interviewer and respondent rapport behaviors (due to different interviewer experience and exposure levels) relate to data quality. The relationship between rapport and data quality is complex and the findings regarding this relationship are not consistent in the literature. The inconsistencies in these findings may be due to inconsistencies in measuring interviewer and respondent rapport. The next section (Section IV) deals with findings regarding the relationship between interviewer and respondent behaviors, which occur during the interview, and data quality, so that it is easier to understand the different behavioral mechanisms among interviewer experience, exposure, and data quality.

4. Interviewer and Respondent Behaviors' Association with Data Quality

In the early ages of survey research, the interaction between the interviewers and respondents was a "black box" for researchers (Cannell & Oksenberg, 1988; Dykema, Lepkowski, & Blixt, 1997; van der Zouwen, 1974, 2006). Van der Zouwen's "black box" metaphor described researchers as providing the input (the questionnaire) to the interviewer and receiving the output (responses) after the administration of the survey; however, what happened during the administration of the survey and whether this related to data quality was not fully understood. Therefore, starting in the 1960s, Cannell and colleagues became pioneers in survey research by using behavior coding (i.e., interaction coding) to gain insight on the interaction between interviewers and respondents (Cannell, Fowler, & Marquis, 1968; Cannell, Lawson, & Hausser, 1975; Cannell, Marquis, &

Laurent, 1977; Cannell, Miller, & Oksenberg, 1981; Cannell & Oksenberg, 1988; Mathiowetz & Cannell, 1980; Fowler & Cannell, 1996).

In the context of standardized interviewing, many behavior coding studies focused on whether interviewers deviated from the conventional ideals such as non-directive probing, neutral feedback, appropriate probing to DK responses, reading questions as written, and not failing to ask a question or a part of a question, as these deviations were believed to lead to response errors (Brenner, 1982; Cannell, Fowler, & Marquis, 1968; Cannell, Lawson, & Hausser, 1975; Cannell, Miller, & Oksenberg, 1981; Cannell & Oksenberg, 1988; Dijkstra & van der Zouwen, 1988; Fowler & Mangione, 1986; Houtkoop-Steenstra, 1996; Schaeffer & Maynard, 1996). These studies discovered that standardization can never be entirely achieved because the interaction between interviewers and respondents during the survey interview is not a mechanical process. They illustrated that interviewers significantly deviated from wording, failed to ask some of the questions, provided improper feedback, and used directive probes (Bradburn et al., 1979; Brenner, 1982; Cannell et al., 1975; Mangione, Fowler, & Louis, 1992; Marquis, 1971; Morton-Williams, 1979). Another interesting finding was that interviewers were providing improper positive feedback (such as "that is good," "you are okay," and "that is interesting") to respondents who were providing inadequate responses, don't know responses, and especially towards refusals to answer a question in order to maintain a friendly communication and a harmonious relationship (Cannell et al., 1981).

Advocates of standardization believed that deviation from conventional ideals would harm data quality (Fowler and Mangione, 1990); however, the few studies that have explored the association between interviewers' deviation from the conventional

ideals and data quality have provided conflicting results. For instance, some studies did not find any relationship between significant deviations from wording and accuracy (Belli & Lepkowski, 1996) or interviewer-related error (Fowler & Mangione, 1990; Groves & Magilavy, 1986; Mangione et al., 1992), whereas Dykema, Lepkowski and Blixt (1997) found that interviewers who significantly changed the question wording were more likely to obtain accurate responses in complicated questions than interviewers who did not deviate from the script as much. One theory is that interviewers may be detecting problematic questions and altering these questions to decrease misunderstandings (Houtkoop-Steenstra, 1995; Houtkoop-Steenstra, 2000; Schaeffer & Maynard, 2001; van der Zouwen & Dijkstra, 1995), and, in doing so, increased data quality (Dykema, Lepkowski, & Blixt, 1997). However, Belli, Lee, Stafford, and Chou (2004) found that "violations of standardization" such as significant deviations from the script that alter question meaning or failure to probe a question were associated with poorer data quality regardless of the interviewing technique used (standardized or calendar interviewing).

Moreover, researchers have used behavior coding to explore behaviors that are measures of respondent *cognitive difficulty* and *interviewer-respondent conversational rapport* and their association with data quality indicators (Belli, Lee, Stanford, & Chou, 2004; Belli, Lepkowski, & Kabeto, 2001; Dijkstra, 1987; Dykema, Lepkowski, & Blixt, 1997; Fowler & Cannell, 1996; Henson, Cannell, & Lawson, 1976). The respondent cognitive difficulty behaviors include respondents' expressions of difficulty and uncertainty, asking for clarifications, guessed or estimated responses, corrections, and interruptions (Belli & Lepkowski, 1996; Conrad & Schober, 2000; Fowler & Cannell,

1996; Mangione, Fowler, & Louis, 1992; Oksenberg, Cannell, & Kalton, 1991; Presser & Blair, 1994). Some studies illustrated that these respondent behaviors occur with specific problematic questions and are indications of poor questionnaire design such as vague respondent tasks, ambiguous question meanings, difficult response tasks, and response and questionnaire order effects (Fowler & Cannell, 1996; Morton-Williams & Sykes, 1984; Oksenberg, Cannell, & Kalton, 1991; Presser & Blair, 1994). In the context of standardized interviewing, studies also found that respondent cognitive difficulty behaviors are indications of lower data accuracy (Belli & Lepkowski, 1996; Belli, Lepkowski, & Kabeto, 2001; Dykema, Lepkowski, & Blixt, 1997). Some of these studies illustrated that when respondents demonstrated cognitive difficulty behaviors, interviewers were more likely to deviate from ideal behaviors by changing questionnaire wording, providing improper feedback, or probing for an adequate response (Belli & Lepkowski, 1996; Mangione, Fowler, & Louis, 1992).

Some of these deviations from conventional ideals have been identified as rapport behaviors (a.k.a. conversational behaviors) such as interviewer improper positive feedback, directive probing, interviewer and respondent digressions, and laughter in the behavior coding literature (Belli, Lepkowski, & Kabeto, 2001; Belli, Lee, Stanford, & Chou, 2004; Bilgen & Belli, 2010b; Cannell, Miller, & Oksenberg, 1981; Lavin & Maynard, 2001). Studies identified rapport as personalized interviewing behavior that is intended to positively affect interviewing by creating a friendly and relaxed environment (Borland, 1975; Collins et al., 2002; Goudy & Potter, 1975-76; Weiss, 1968-69). Even though the identification of the rapport concept is quite similar in these studies, how it

has been measured and how it affects data quality vary greatly throughout the literature (Cannell, Miller, & Oksenberg, 1981; Goudy & Potter, 1975-76; Weiss, 1968-69).

Other studies measure rapport via "personal" interviewing style and compare it to a more "formal" (i.e., conventional) interviewing style (Dijkstra, 1987; Henson, Cannell, & Lawson, 1976; van der Zouwen, Dijkstra, & Smit, 1991). In addition, a third approach to measuring respondent rapport is by obtaining interviewers' ratings of respondent attitudes via interviewer evaluation questionnaires. For instance, Weiss (1968-69) used a five-point scale for items regarding respondent's positive attitudes such as trustfulness and sincerity or negative attitudes such as vagueness, guardedness, and hostility. Both Hill and Hall (1963) and Goudy and Potter (1975-76) measure rapport using items such as "how often respondents and interviewers felt at ease during the interview" and "how favorable the respondent seemed". Moreover, Davis and Silver (2003) compiled their rapport measure from four-point scale items regarding respondent cooperation and interest. However, as mentioned earlier in Chapter II in detail, interviewer perceived rapport is not a precise measure of data quality.

In the context of standardized interviewing, the inconsistent findings regarding rapport and accuracy may be due to the inconsistent measures of rapport. Some researchers found a positive association between rapport and respondent motivation (Dijkstra, 1987; Goudy & Potter, 1975-76) and, therefore, data quality (Dijkstra, 1987). Hill and Hall (1963) illustrated that higher levels of rapport were related to lower levels of item non-response. In addition, Davis and Silver (2003) found a positive association between rapport and correct responses to political knowledge questions. However, Weiss (1968-69) illustrated that higher levels of rapport were detrimental to response quality.

Belli, Lepkowski, and Kabeto (2001) did not find any association between rapport behaviors used during the interview and response accuracy. Henson, Cannell, and Lawson (1976) also did not find any differences in the accuracy of reports between the "personal" interviewing style and the "formal" interviewing style.

There are several theories on why rapport is important in an interviewing situation. One theory suggests that rapport increases respondent motivation so that respondents are willing to provide sincere responses to potentially sensitive and embarrassing questions (Borland, 1975; Collins et al., 2002; Cleary et al., 1981; Williams, 1968) and try harder to help interviewers to meet the research goals such as trying harder to remember for cognitively challenging questions (Collins et al, 2002; Dijkstra, 1987; Henson, Cannell, & Lawson, 1976). In addition, studies also point out that there is a curvilinear relationship between rapport and data quality such that too little or too much rapport may harm data quality (Belli, Lepkowski, & Kabeto, 2001; Dijkstra, 1987). With too much rapport, respondents may adapt their responses for interviewer approval (Hyman et al., 1954).

In summary, studies on the relationship between response accuracy with both interviewer-respondent rapport and interviewer deviation from conventional ideals provide mixed results (Belli & Lepkowski, 1996; Dykema, Lepkowski, & Blixt, 1997; Fowler & Mangione, 1990; Groves & Magilavy, 1986; Mangione et al., 1992). Perhaps these differences are due to the complexity of the interviewer's role in achieving a balance between trying to resolve problematic and difficult questions, which potentially cause respondent cognitive difficulty, and building rapport to create a motivating survey environment while trying not to digress from the researcher's protocols and to be

efficient. Nevertheless, there is still an ongoing debate, referred as the "standardization controversy," on whether rapport and interviewer deviation from conventional ideals are detrimental or beneficial to data quality and how researchers can make sure respondents interpret the questions in exactly the way researchers intend to decrease measurement error. However, my main aim is not to investigate this controversy; rather, my main goal is to explore the role of interviewer experience in this controversy by disentangling different kinds of interviewer and respondent behaviors using three different mechanisms.

5. The Relationship between Interviewer and Respondent Verbal Behaviors and Data Quality in Different Interviewing Techniques

Even though results are mixed, the examination of interviewer and respondent interaction illustrates that the interviewer and respondent interaction and the behaviors that occur during this interaction may have an important effect on data quality (Belli et al., 2004; Bradburn et al., 1979; Brenner, 1982; Cannell et al., 1981; Dykema, Lepkowski, & Blixt, 1997; Houtkoop-Steenstra, 1995; Morton-Williams, 1979; van der Zouwen & Dijkstra, 1995). Therefore, studies related to the standardized versus flexible interviewing controversy literature have studied interviewer and respondent behaviors that occur during the interviewing process (Belli, Lee, Stafford, & Chou, 2004; Houtkoop-Steenstra, 2000; Maynard & Shaffer, 2002; van der Zouwen, 2002).

One flexible interviewing technique that has been consistently provided as an alternative method to standardized interviewing is ***calendar interviewing*** (Belli, Lee, Stafford, & Chou, 2004; Belli, Shay, & Stafford, 2001; Belli, Smith, Andreski, & Agrawal, 2007; Bilgen & Belli, 2010b; Yoshihama et al., 2005). In the next section, I

introduce this technique, provide background, and compare it with standardized interviewing. Also in the following section, I review the studies that compare the use of interviewer and respondent behaviors in calendar and standardized interviews and explore how these behaviors relate to data quality. Lastly, I explain how all of these relate to my research questions.

5.1. Event History Calendar and Conventional Standardized Interviewing

In surveys standardized interviews are accepted as an adequate practice in survey interviewing, given that it is believed to reduce interviewer variance by standardizing the wording of the questions and having respondents interpret the questions (Fowler & Mangione, 1990). The aim of the conventional standardized interview is that each respondent gets the same message, so all questions are asked in the same way. Thus, the interaction between any particular interviewer (regardless of their fixed characteristics such as interviewer experience) and respondent is consistent with all other interviewer-respondent interactions. This technique aims to reduce measurement error due to the interviewer (Schober & Conrad, 1997; Houtkoop-Steenstra, 2000). Another rationale for this technique is that it aims to reduce cost by minimizing the interview length, interviewer training time, and coding time (Fowler & Mangione, 1990).

In standardized interviewing, the interviewers are trained to read questions exactly as written, so the ordinary conversation process is controlled by the researchers (Conrad & Schober, 2000; Schober & Conrad, 1997; Suchman & Jordan, 1990). Probing is expected to be non-directive and guidelines are provided to interviewers to use when probing is needed and how to provide it. Examples of nondirective and neutral feedback are provided to interviewers so that they can motivate respondents to try harder without

biasing them while answering the questions. When there are problems and respondents ask for clarification, the follow-ups are also standardized with phrases such as *"whatever it means to you"* or *"it is up to you."* The biggest criticism to the standardized interviewing technique is that respondents can interpret the same question in different ways because important and memorable events differ for each respondent. Therefore, there is no consistent standardized meaning for a question from the respondents' point of view. As a result, whether or not standardization increases validity of respondent reports (especially retrospective reports) becomes debatable (Belli et al., 2007).

An alternative to the standardized interviewing is the *calendar interviewing* technique, which is designed to collect retrospective reports using different timelines for different domains (such as residence, health, and employment histories) in order to better reflect the structure of autobiographical memory (Belli, 1998; Belli, Alwin, & Stafford, 2009; Belli et al., 2001). One of the rationales for the use of the calendar interviewing technique is that it allows the use of more effective approaches to remembering, encouraging respondents to remember via retrieval cues. Therefore, it is hypothesized to promote productive retrieval and accurate reporting. Another advantage of calendar interviewing regarding data quality is that it allows a flexible conversational interviewing style, which encourages respondents to retrieve events with the help of retrieval strategies that are based on the structure of autobiographical memory (Belli, 1998, 2000; Belli et al., 2004; Bilgen & Belli, 2010b). These cues include 1) top-down retrieval processes in which the specific details are cued with more general information, 2) sequential retrieval processes in which events or spells[20] within the same domain are recalled in their order of

[20] A spell is a continuous or ongoing activity. A spell refers to a period between two points of time. For example, an employment spell is the period between the beginning and end of a particular job.

occurrence, and 3) parallel retrieval processes in which simultaneous events from more than one domain are used in order to provide accuracy during the recall of timing (Belli, 1998; Belli et al., 2001, 2004).

The need for obtaining more accurate and valid autobiographical behavioral responses has encouraged the *"emergence of calendar interviewing"* to collect retrospective reports (Belli & Callegaro, 2009). The flexible nature of this technique presumes that meaning is interpreted through communication between the interviewer and the respondent. Interpretation and meaning are created during the interview as in any flexible interview (Conrad & Schober, 2000; Suchman & Jordan, 1990; Schober & Conrad, 1997). Less standardized methods such as calendar interviewing are believed to provide conversional flexibility and employ retrieval cues, which not only repair misunderstandings during the interview but also allow the use of different life domains (e.g., residence, cohabitation, work, marriage, etc...) to aid recall of lifetime events that belong to other life domains and increase the interest of respondents (Freedman et al., 1988, Means & Loftus, 1991; Belli, 1998; Belli et al., 2004, 2007).

According to Belli et al. (2004), standardized conventional questionnaires (CQ) are designed to efficiently utilize top-down and sequential retrieval cues. Calendar interviewing introduces parallel cues and several types of sequential cues that help respondent recall events more productively, which are not commonly used in standardized interviews (Belli, 1998). The flexible, more conversational style of calendar interviewing also might resolve uncertainties that are a part of conversations (Schober & Conrad, 1997; Belli et al., 2001). Belli (1998) points out that the conventional standardized interviewing technique is likely to disconnect related aspects of

autobiographical events from one another. Therefore, survey questions that are used in standardized interviewing technique do not reflect the associations between events as it is indicated in the autobiographical memory research. However, the calendar method promotes sequential and parallel retrieval cues and uses the literature regarding the structure of autobiographical memory. By utilizing the knowledge of the structure of autobiographical memory, the calendar interviewing technique assists respondents to reconstruct their past events more completely and accurately and, thus, are found to improve the quality of retrospective reports (Belli, Shay, & Stafford, 2001; Yoshihama et al., 2005).

The main goal of calendar interviewing is to increase the quality of retrospective reports, whereas the primary aim of standardized interviewing is to decrease interviewer-related measurement variation. A recent study regarding cost-benefit analyses of retrospective reports discovered that the benefit of calendar interviewing is a small but reliable increase in data quality. However, this benefit results in a slight increase in interviewer variance in comparison to standardized interviewing (Sayles, Belli, & Serrano, 2010). Although the authors illustrated that the cost-benefit analyses slightly favor calendar interviewing, it is apparent that both techniques have their respective strengths and weaknesses.

5.2. Interviewer Experience and Exposure Role in Interviewer-Respondent Behavioral Differences among Calendar and Standardized Interviews

Verbal behavior coding is useful during disentangling some of the respective strengths and weaknesses of each interviewing method and their impact on data quality. As mentioned earlier, standardized interviewing is designed to train interviewers to avoid

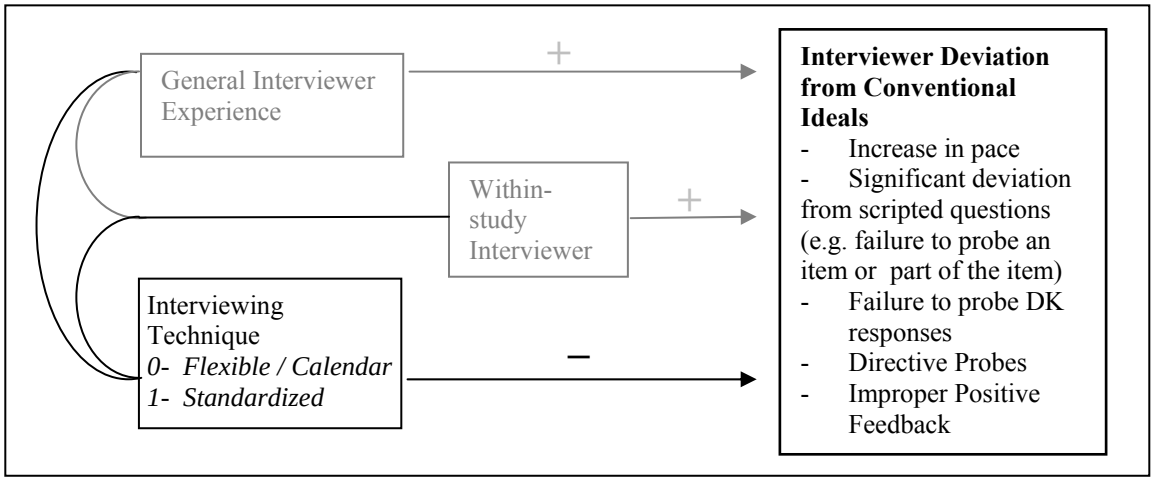
deviation from the script and conventional ideals (Beatty, 1995; Dykema et al., 1997; Fowler & Mangione, 1990; Schaeffer & Maynard, 1996). Interviewer deviations from conventional ideals have been measured via several different behaviors, most commonly used behaviors include significant deviations from the scripted questionnaire, increases in pace, directive probes, and improper feedback (Bradburn et al., 1979; Belli et al., 2001, 2004; Bilgen & Belli, 2010b; Brenner, 1982; Cannell et al., 1981; van der Zouwen, Dijkstra, & Smit, 1991; Mangione et al., 1992; Marquis, 1971; Morton-Williams, 1979). Given that calendar interviews promote conversational flexibility, not surprisingly, Belli, Lee, Stafford, and Chou (2004) and Bilgen and Belli (2010b) found that most of the deviations from standardized behaviors such as improper feedback and directive probing were used more in calendar than in standardized interviews.

Furthermore, some studies referred to interviewer pace as "interviewer words per second" (Cannell, Miller, & Oksenberg, 1981), while others, due to data limitations, used interviewing time (sometimes referred to as "interview length" in minutes) as a measure of pace (Olson & Bilgen, 2011; Olson & Peytchev, 2007). Belli et al. (2007) and Sayles, Belli, and Serrano (2010) found a small increase in interviewing time in CATI[21] calendar interviews in comparison to standardized interviews, though Belli et al. (2001) observed no differences in interviewing time in paper and pencil calendar interviews in comparison to standardized interviews. A decrease in interviewing time may indicate that interviewers may have a faster pace in standardized interviews in comparison to calendar interviews; however, interviewing time is not a perfect measure of pace (Olson & Bilgen, 2011; Olson & Peytchev, 2007). Therefore, in this dissertation the interviewing pace measure is obtained through *average number of words per minute*. To sum up, on one

[21] CATI: Computer Assisted Telephone Interviewing

hand interviewer deviation behaviors such as interviewer improper feedback, directive probes and failure to probe an item or a part of an item behaviors are more likely to be prevalent in calendar interviews in comparison to standardized interviews, on the other hand interviewer pace may be faster in standardized interviews than in calendar interviews (see Figure 2.3). Yet, as it is illustrated via "–" sign in Figure 2.3, the majority of interviewer behaviors that are measures of interviewer *deviation from conventional ideals* are used less prevalently in standardized interviews than in calendar interviews (Belli et al., 2001, 2004; Bilgen & Belli, 2010b).

Figure 2.3 Deviation Behaviors in Calendar and Standardized Interviews



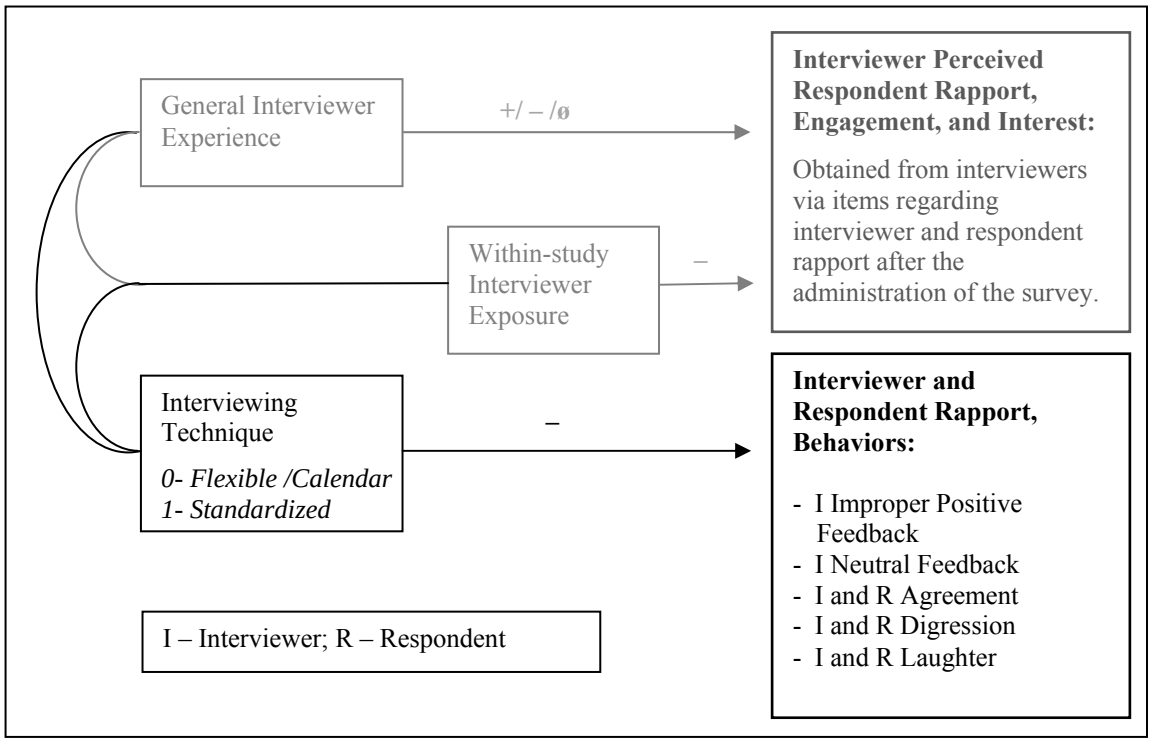
As illustrated in detail earlier, variation among interviewers' experience and exposure levels leads to deviations from the intended standardization due to higher rates of probing, feedback, reading errors, and speech variations (Bradburn et al., 1979; Cleary et al., 1981; Fowler & Manigue, 1990; O'Muircheartaigh & Campanelli, 1998). Moreover, the studies that examine the verbal behavior and interviewing technique relationship mainly focus on the questionnaire design properties and rarely focus on how interviewer characteristics play a part in this equation. No studies take into account how

interviewer experience and exposure play a role together in association with interviewing technique and the use of verbal behaviors (see Figure 2.3). Therefore, in my study I take into account the interactive nature of interviewer experience and exposure and study the relationship between interviewer experience, exposure and interviewer behaviors that are measures of *deviation from conventional ideals,* in different interviewing methods (in calendar and standardized interviews).

Studies that explored the relationship among interviewer experience, exposure, and rapport measured interviewer rapport mainly as interviewer perceived rapport due to data restrictions (Goudy & Potter, 1975-76; Hill & Hall, 1963). Belli et al. (2004) and Bilgen and Belli (2010b) are two of the few studies that explored conversational rapport via coding interviewer and respondent rapport behaviors that occur during the administration of standardized and calendar interviews (see Figure 2.4). These behavior studies identified rapport behaviors as interviewer positive or neutral feedback, interviewer and respondent digressions, agreements, and laughter (Belli, Lepkowski, & Kabeto, 2001; Belli et al., 2004; Bilgen & Belli, 2010b; Cannell et al., 1981; Lavin & Maynard, 2001). As calendar interviewing allows for more conversational rapport than standardized interviews, overall Belli et al. (2004) and Bilgen and Belli (2010b) illustrated that the majority of interviewer and respondent rapport behaviors are used more prevalently in calendar interviews than in standardized interviews (illustrated via "**—**" sign in Figure 2.4).

Specifically, both Belli et al. (2004) and Bilgen and Belli (2010b) found significantly higher rates of interviewer improper positive feedback and respondent agreement behaviors in calendar interviews. However, Bilgen and Belli (2010b) found

Figure 2.4 Differential Uses of Interviewer and Respondent Rapport Behaviors in Calendar and Standardized Interviews



higher levels of digression in calendar interviews, whereas Belli et al (2004) did not find any significant differences in the use of digressions between interviewing methods. Both studies found no significant differences in both interviewer and respondent laughter between calendar and standardized interviews, though laughter is a complicated topic. While some types of laughter such as laughing at another's joke or comment indicates rapport, in some instances, laughter occurs in uncomfortable social situations (Houtkoop-Steenstra, 2000). Houtkoop-Steenstra (2000) illustrated that interviewer laughter occasionally occurs right after respondents' negative comments or complaints regarding the interview or a specific question. Therefore, in my study I aim to disentangle the types of laughter in order to separate the rapport-related laughter from other non-rapport related laughter.

Furthermore, the studies that examine the relationship between verbal behavior and interviewing technique mainly focus on questionnaire design properties and rarely focus on how interviewer characteristics play a part in this equation. Also, to my knowledge there are no studies that have examined how interviewer experience and exposure play a role in the differential use of interviewer and respondent rapport behaviors in different types of interviewing techniques. Therefore, I plan to explore interviewer experience and exposure interactive effect on rapport behaviors that occur in calendar and standardized interviews.

The calendar literature puts a great emphasis on retrieval behaviors as these behaviors encourage more efficient retrieval of retrospective reports (Belli, 1998, 2000; Belli & Callegaro, 2009; Belli, Shay, & Stafford, 2001; Belli et al., 2004; Bilgen & Belli, 2010b; Dijkstra, Smit, & Ongena, 2009). Interviewer retrieval probes use the structure of autobiographical memory to obtain more accurate recall (Belli, 1998, 2000; Belli & Callegaro, 2009; Belli, Lepkowski, & Kabeto, 2001; Belli, Shay, & Stafford, 2001). In the interviewing situation, these retrieval probes are offered by the interviewers and retrieval strategies have been observed to be used by the respondents (Belli et al., 2004; Bilgen & Belli, 2010b). The retrieval probes include behaviors such as *parallel probes* that use contemporaneous events from one life phase (such as residence) to recall events from another life phase (such as education), *duration probes* that seek information regarding how long an event has occurred, *sequential probes* that ask respondents to recall events within the same life domain in the order of occurrence, and *timing probes* that ask respondents when an event started or ended (Belli, 1998, 2000; Belli et al., 2001a, 2001b; Belli et al., 2004; Bilgen & Belli, 2010b). Consistently, the calendar

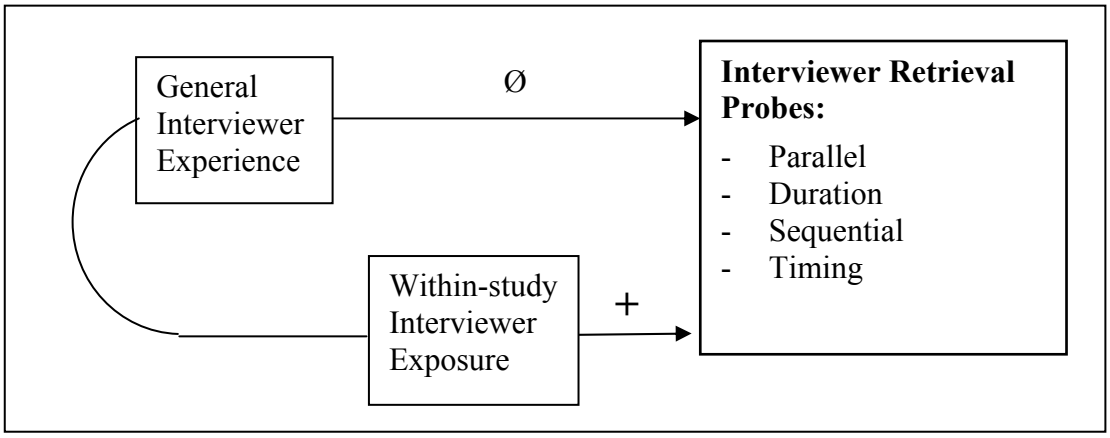
method is also designed to increase the use of respondents' spontaneous associations between events through idiosyncratically using the retrieval strategies (Belli, 1998; Belli et al., 2004; Bilgen & Belli, 2010b).

Respondent retrieval strategies are very similar to interviewer retrieval probes and include *parallel retrieval strategies* in which respondents spontaneously relate concurrent events from separate life domains, *duration retrievals* in which respondent spontaneously provide the duration of an event, *sequential retrievals* in which respondents spontaneously relate thematically similar events that happened right before or after each other, and *timing retrievals* in which respondents spontaneously provide when an event or sequence of events has started and ended (Belli, 1998, 2000; Belli et al., 2001a, 2001b, 2004; Bilgen & Belli, 2010b) (see Appendix 2 for more detailed information and examples).

To my knowledge, Bilgen, Belli, and Olson (2009) is the only study to explore the relationship among interviewer experience, exposure, and interviewer retrieval probing. No studies exist that explore the interviewer experience and exposure impact on the use of respondent retrieval strategies. This dissertation intends to fill in this research gap. Bilgen, Belli, and Olson (2009) illustrated that general interviewing experience—gained conducting standardized interviews—does not come into play in parallel, duration, and timing probes (illustrated as "ø" in Figure 2.5) and explained the reason for this as conventional interviewing techniques not promoting the use of the retrieval behaviors; therefore, interviewers, regardless of experience levels, are relatively new to the use of the retrieval probing strategies because they are relatively new to calendar interviewing. Bilgen, Belli, and Olson (2009) also found an increase in the use of interviewer retrieval

probes during the survey fielding period (illustrated via "+" in Figure 2.5) and theorized that interviewers practice using these behaviors with each interview they conduct at the beginning of the study.

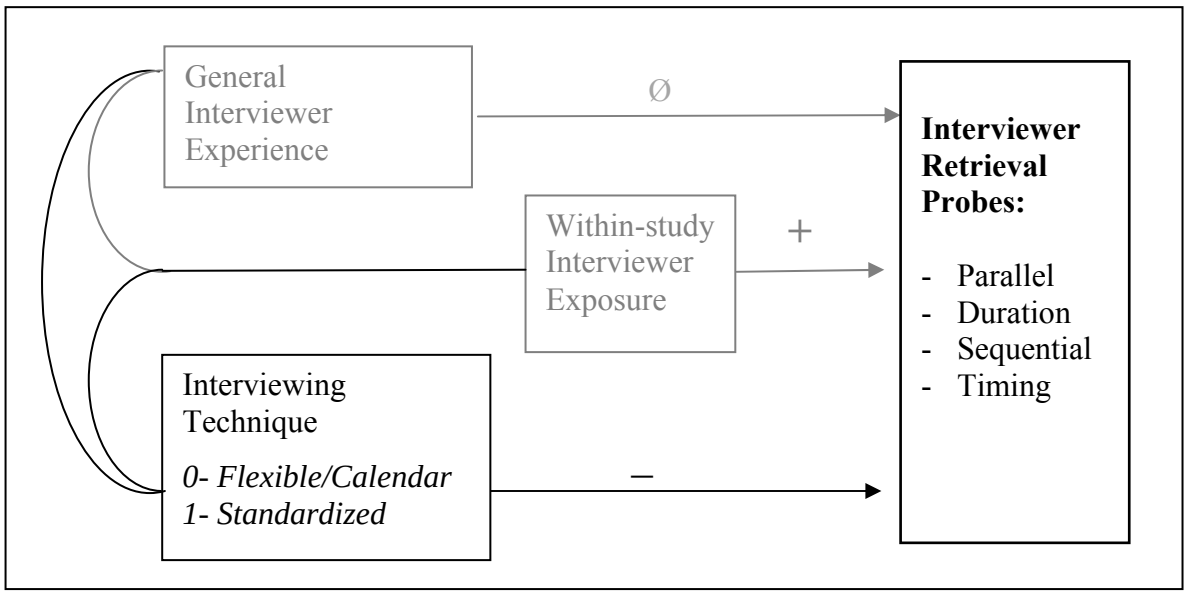
Figure 2.5 The Relationship between Experience, Exposure and Retrieval Probes



Studies that explore the prevalence of verbal behaviors among different interviewing techniques have found slightly mixed results. Belli et al. (2004) found parallel, sequential, and timing probes to be more prevalent in calendar interviews, whereas duration probes were found to be more prevalent in standardized interviews. Bilgen and Belli (2010b) found that interviewers used significantly more parallel and duration probes in calendar interviews with no statistically significant differences in the use of sequential and timing probes between calendar and standardized interviews. These results are consistent with the expectation of calendar interviews to promote the use of parallel and sequential behaviors given that these behaviors play the most crucial role in autobiographical memory. According to the findings from Belli et al. (2004) and Bilgen and Belli (2010b), interviewer retrieval probes are less prevalent in standardized interviews than calendar interviews (illustrated using "—" sign in Figure 2.6).

Many studies that explore the use of interviewer retrieval behaviors in different interviewing methods did not explore how interviewer characteristics come into play in this relationship. Bilgen, Belli, and Olson (2009) is the only study that took into account how interviewer experience and exposure play a role together in the use of interviewer retrieval behaviors in calendar and standardized interviews (see Figure 2.6). They found that overall the increase in the use of interviewer retrieval probes due to within-study exposure were higher in calendar interviews than in standardized interviews. What is yet to be explored is the relationship between interviewer experience, exposure and the differential use of respondent retrieval strategies in calendar and standardized interviews and how this relationship impacts the data quality. This dissertation intends to fill in this research gap.

Figure 2.6 The Differential Use of Interviewer Retrieval Probes in Calendar and Standardized Interviews



For the purposes of clarity and eliminate redundancy, before providing theoretical framework and hypotheses, next chapter (Chapter III) will provide information on data

sources and measures used in this dissertation. After that, Chapter IV illustrates hypotheses, and expected direction of the results for the three sets of research questions. Therefore, the aim of Chapter III is to introduce the measures and their roles in each of the hypotheses reported in Chapter IV, and the models reported in Chapters V and VI.

CHAPTER III: DATA SOURCES AND MEASURES

1. Data Description

1.1. Stage 1 – Computer-Assisted Telephone Survey Interviewing

The data for this study come from an experiment conducted in 2002. In this experiment, 632 individuals were randomly sub-sampled from the 2001 Panel Study of Income Dynamics (PSID)[22] nation-wide sample. These individuals were interviewed using CATI (computer-assisted telephone interviewing) from July to September 2002. Both interviewers and respondents were randomly assigned to *calendar* and *standardized* conditions. The selection criteria to participate were that the participants had to be members of PSID households (i.e., families) who participated in every wave of the PSID from 1980 to 1997 and the participants had to be interviewed in no less than 50 percent of the waves in which their households have participated.

First, 26 interviewers were first matched according to their general interviewing experience, and then randomly assigned to either condition. This interviewer assignment produced approximately equivalent telephone and face-to-face interviewing experience between calendar and standardized interviews (Belli, Smith, Andreski, & Agrawal, 2007; Bilgen & Belli, 2010b). Also, this assignment led to similar interviewer characteristics between calendar and standardized conditions such as similar interviewer sex (10 and 11 female interviewers in calendar and standardized conditions, respectively) and mean of

[22] The PSID is a longitudinal study with a probability sample of U.S. households (family units) that interviewed and re-interviewed members from sampled families, whether or not they were living in the same dwelling or with the same people, every year from 1968 to 1997. The PSID followed members of the households as they aged and as they formed family units of their own. For more information please see http://psidonline.isr.umich.edu/

interviewer age (see Belli, Smith, Andreski, & Agrawal, 2007). Overall, 13 interviewers and 313 participants were assigned to *calendar* and 13 interviewers and 319 participants were assigned to the *standardized* condition. Each respondent was offered a $50 incentive as a token of appreciation to participate in the study.

The standardized condition was programmed using prepackaged CATI software called *Blaise*®, which is commonly used in standardized interviews. The calendar instrument was an in-house CATI program that was generated at the University of Michigan. In both conditions, the questionnaires were designed to collect respondents' retrospective reports regarding their lifetime experiences with reference to their residence (including residential changes and addresses), marriages (e.g. the names of the spouses, number of years married, marital status changes), cohabitations (e.g. the names of the partners whom the respondent lived as married, number of years cohabited with partners, cohabitation status changes), children (e.g. number of children, name and date of birth of each children), employment (including their employers, work hours, and employment changes), unemployment (including time and duration of unemployment), and health history (including health status, weight, height, whether they were ever disabled, and whether they have ever smoked) (see Appendices 5 and 6 for more information on questionnaires used in calendar and standardized interviews). In addition, the respondents were also asked about their parents and their socio-economic status while growing up. However, these two domain are not included in the coding process as these domains include proxy responses and this dissertation focuses on respondents' self-reports which are less prone to response errors. Of the 632 interviews, approximately 93% of the interviews were audio-tape recorded with respondent permission (N_{CAL}=297; N_{STD}=291).

Among 588 tape-recorded interviews, around 4% of the interviews were problematic either due to poor tape and sound quality or a mismatch between the audio tape and the data file used to organize the audio tapes. Next, 564 non-problematic tapes (N_{CAL}=291; N_{CQ}=273) were transcribed by 15 transcribers.

1.2. Stage 2 – Verbal Behavior Data Collection

This dissertation uses verbal behaviors obtained from two different verbal behavior coding studies that used the same transcripts described above. The verbal behaviors used to test the **first set of research questions**—which aim to examine the role of *interviewer deviation from conventional ideals*—and **second set of research questions** —which aim to study the role of *interviewer and respondent interpersonal dynamics*—were collected specifically for this dissertation using the Charles Cannell Fund. The cognition-related verbal behaviors are used in response to **third set of research questions**—which aim to investigate the role of *interviewer and respondent retrieval strategies* in experience, exposure, and item non-response relationship—have been obtained from an earlier verbal behavior study with initial analyses reported by Bilgen and Belli (2010b).

In the verbal behavior study that investigated *deviations from conventional ideals* and *interpersonal dynamics*, a randomly selected 165 calendar and 162 standardized (N_{Total} = 327) transcripts (58% of the transcribed tapes) were examined. The verbal behavior coding scheme that investigates *interviewer deviation* and *interpersonal dynamics* is referred as the ***new behavior coding scheme*** throughout the dissertation. In addition, the coding scheme that focuses on *cognition-related behaviors* and used as a guide for the new behavior coding scheme is called the ***old behavior coding scheme*** throughout this dissertation. The behaviors that are used to investigate the last set of

research questions regarding *cognition-related behaviors* are a part of the coding scheme explained by Bilgen and Belli (2010b). From this coding scheme, four interviewer retrieval probes, four respondent retrieval behaviors, and one interviewer deviation from conventional ideal behavior is utilized in this dissertation. More detailed information on this can be found in Appendix 1, which includes detailed behavior definitions, examples, and coding rules regarding new behaviors, and Appendix 2, which includes detailed behavior definitions, examples, and coding rules regarding old behaviors.

New Behavior Coding Scheme – Communication Behaviors

At the beginning of fall 2009, I developed an initial *new coding scheme* with the help of a master's-level graduate student. In mid-October 2009, I hired four University of Nebraska-Lincoln (UNL) students—2 undergraduate and 2 graduate—to code verbal behaviors in the transcribed interviews. During the training process, my coding team and I improved the new coding scheme simultaneously and added additional behaviors to the initial coding scheme. The final new coding scheme includes 12 interviewer and 15 respondent behaviors (see Appendix 1). The coding team coded transcripts using the new coding scheme that includes interviewer and respondent linguistic expressions used during the interview relevant to my study. Table 3.1 illustrates a portion of the behaviors that aim to measure ***interviewer deviations from conventional ideals.*** These deviation behaviors aim to capture interviewers' deviations from the recommended standardized protocols during the data collection process (see Chapter II for detailed description). Each of these deviation behaviors is a verbal behavior that occurred during the interview except *interviewing pace* (see Table 3.1).

Table 3.1

Interviewer Deviation from Conventional Ideal Behaviors

Behaviors	Descriptions
Behaviors calculated	
Interviewing Pace	$\frac{\text{Number of words used in the interview}}{\text{Length of the Interview (in minutes)}}$ Each respondent has a different life history. The difficulty and complexity of the information obtained might affect the length of the interview (i.e., how many minutes the interview took to be completed). In order to standardize the length of the interview measure interviewing pace is measured via *average number of words used per minute for both interviewers and respondents.* This way, interviewer and respondent speech variations can be captured in a more standardized fashion.
Behaviors measured via verbal behavior coding (4 out of 5 are new behaviors)	
Failure to probe	Interviewer fails to ask an item or part of the item. *Example: R: Well, I had an accident in 1951. I: Would you say that before then it was very good? R: Very good. I: Okay, then in 1951 things changed? R: Yes. I appreciate the Lord for letting me live this long, cause lots of people thought I'm gonna die lots of times, but the Lord spared my life. I: That's wonderful. Okay, have you ever smoked cigarettes?* (Interviewer fails to ask about respondent's health status change from 1951 until the year interview took place (2002). Within this 51-year period, respondent's health might have changed again).
Significant change[23]	Interviewer significantly changes the question wording that can potentially modify the meaning of the question.

[23] There was no consensus among the coders on what was considered as *significant changes in question meaning,* especially in calendar interviews, even after I provided specific rules on what to consider as *significant change*s. Taking into consideration the subjective nature of this behavior, my coding team and I decided to exclude this from the new coding scheme. In addition, a verbal behavior coding study led by Dr. Robert F. Belli also attempted to code significant changes; the researchers noted that the significant change behaviors were not reliably coded among coders and this code was dropped out of the study (for more detailed information see Bilgen and Belli, 2010b).

Failure to probe to "don't know" (DK) response	Interviewer accepts a respondent's DK responses without providing any additional probing. *Example: R: All I can tell you is it was on street1. I don't know the address of the house. I: Okay, that is fine. I am sorry, what city did you say that was in?*
Improper positive feedback[24]	Interviewer provides feedback that carries a positive connotation. All interviewer feedback behaviors—whether neutral feedback such as "Thank you" or positive feedback—are geared toward attempts to engage respondents more with the study and demonstrate appreciation. However, *interviewers' improper positive feedbacks* may be also encouraging respondents' undesirable responses such as refusals or DK responses more than interviewer neutral feedback. *Example: R: I'm not going to be able to give you many street addresses. I: Um, that's fine. Just do as good as you can, and uh, we'll—we'll work around it.*
Directive probes[25] ***(From the old coding scheme)***	Interviewer asks a question that could potentially bias a respondent's answer, for instance, when an interviewer assumes a response and asks whether the assumed response is true. *Example: R: Probably until I was about, um, uh, 6. I: So that would make it about 1937? (In this case, a nondirective way of asking the question can be "do you remember the year?")*

Table 3.2 contains behaviors, definitions and examples of the behaviors that reflect ***interpersonal communication dynamics*** used by the interviewers and respondents during the interviews (such as *interviewer rapport*, *respondent cooperative and non-cooperative behaviors*, and *respondent expressions of difficulty*). The behaviors related to *interpersonal communication dynamics* constitute the majority of the new coding scheme. Respondent behaviors that are measures of respondents' engagement and interest

[24] In practice, interviewer's improper positive feedback can be considered as both a deviation from conventional ideal behavior and an interviewer rapport behavior. However, for the purpose of creating mutually exclusive scales, I included this behavior as a part of single scale rather than including it in both scales. In the past verbal behavior studies, it has been considered as an undesirable interviewer behavior in standardized interviews as it may encourage respondents' less than ideal responses such as respondent DK responses (Belli et al., 2001; Cannell et al., 1981; Dijkstra & Van der Zouwen, 1988). Therefore, I included the improper positive feedback as a part of interviewer's deviation from conventional ideals.

[25] The old coding scheme includes *directive interviewer probing*. Therefore, the new behavior coding scheme does not include this behavior.

with the study (i.e., willingness to help the interviewer) are referred as ***respondent cooperative behaviors***. Additionally, respondent behaviors that are measures of respondents' loss of engagement and interest are referred as ***respondent non-cooperative behaviors***. Respondent behaviors that aim to capture respondent linguistic indications of uncertainty and difficulty are referred as r***espondent expressions of difficulty***.

Table 3.2

Interpersonal Communication Behaviors

Behaviors measured via verbal behavior coding (new behaviors)

Behaviors	Descriptions
Interviewer Rapport Behaviors	
Interviewer neutral feedback	Interviewer provides a neutral phrase such as "thank you," "okay," or "fine" following a response to a study-related probe.
Interviewer joking and providing a sarcastic comment	Interviewer jokes or provides a sarcastic comment both regarding a **study-related or un-related** topic. *Example: My computer and I are not getting along right at this second.*
Interviewer empathy	Interviewer feels concern for the respondent and tries to share or understand what the respondent is thinking or feeling about the interview or about a reported event. *Example: I had to do this interview, and it was—it—it was like yours. There was a lot.*
Interviewer agreement	Interviewer agrees with respondents' both **study-related and non-study related** comments. *Example: R: This line is bad. I am having trouble. I: Yes, I can hear the static on it.*
Interviewer direct apology[26]	Interviewer apologizes from the respondent regarding the interview/task/question/questionnaire/computer program/ his/her own error by specifically saying "I am sorry"/ "sorry"/ "I apologize". *Example: I am sorry that the interview takes too long.*

[26] *Interviewer direct apology* behavior is the combination of *Interviewer apologizes from the respondent regarding the interview/ task/ question(/naire)/ computer program* and *interviewer apologies regarding his/her own error* behaviors. Both behaviors are coded when interviewers specifically said "I am sorry"/ "sorry"/ "I apologize".

Interviewer apologetic comment	Interviewer provides a comment to indirectly apologize from the respondent **without specifically saying** "I am sorry"/ "sorry"/ "I apologize" regarding the interview/task/question/questionnaire/ computer program/for his or her error. *Example: R: What did I say earlier? I: I don't, uh--.The screen goes on, and I can't see the answers.*
Interviewer laughter to a respondent joke/comment	*Example: R: I'm watching The Godfather. I can do that without sound. (Laugh-I.)*
Respondent Cooperative Behaviors	
Respondent empathy	Respondent feels concern for the interviewer. *Example: R: So, that's, uh—how you're going to record that, I don't know.*
Respondent joking and sarcasm[27]	Respondent jokes or provides a sarcastic comment both regarding a study-related or un-related topic. *Example: I: Oops. Zipcode1. New sheet. Just one minute. My computer and I are not getting along right at this second. R: Well, straighten it up.*
Respondent's spontaneous attempts to resolve difficulty	Respondent implicitly or explicitly offers help to resolve a cognitive difficulty or technical difficulty. *Example: I: Well, we got a lot of static, don't we? R: There is a lot. Let me try another phone. I: Okay.*
Respondent spontaneously offers or provides clarification	Respondent clarifies (or offers to clarify) any aspect of study-related information that he/she provided earlier. *Example: R: I was working full time. Just to explain why I'm laughing, uh, these are, um, 18 to 20 hour days.*
Respondent corrections[28]	Respondent spontaneous *corrections* of a response provided earlier or an interviewer study-related comment or assumption. *I: Example: I'm sorry. September, 1939? R: No, wait a minute. (Oh.) Uh, it was June of 1939.*
Respondent laughter	Respondent's laughter to an interviewer joke, comment or feedback. *Example: I: My computer and I are not getting along right at this second. R: Laugh-R.*

[27] During the coding scheme development, coders indicated that both interviewer and respondent jokes and sarcastic comments were not easily differentiated from each other. The coders were coding from the transcripts rather than the tapes; hence, they were not able to hear the vocal nuances that would enable them to disentangle these behaviors from each other. Therefore, in order to decrease the costs and timing the tapes decided to be not used in the coding process in addition to transcripts. Therefore, these two codes are decided to be combined at the end of the coding scheme development process.

[28] *Respondent corrections* are the combination of *respondent spontaneous corrections of a response provided earlier* and *respondent corrections of study-related interviewer comment/assumption* behaviors.

Respondent positive regard for the interview or questionnaire	Respondent indicates that the interview or the questionnaire is enjoyable or interesting. *Example: R: This is going to be a fun interview.*
Respondent Non-Cooperative Behaviors	
Respondent negative comment	Respondent implicitly or explicitly indicates that the interview/question/questionnaire is long, complicated, foolish, boring, repetitious, inappropriate etc... *Example: I: Oh, wait a minute. I just [Unintelligible]. R: I think this is foolish.*
Uncomfortable laughter	Respondent is using laughter to deal with an uncomfortable situation such as expressions of difficulty or during answering uncomfortable/sensitive questions or when either of the speakers provides a comment that can potentially increase the tension between the actors. *Example: I: And do you remember the zip code there? R: Mmm, no. Laugh-R.*
Reluctance to provide information	Respondent refusals to answer a specific question (e.g., "I won't answer this question") and respondents' indications of not being able to answer the question without thinking through a response (e.g., "I can't answer that"). *Example: Now—now, you—do you want street adder—addresses or cities? I'm not going to be able to give you many street addresses.*
Respondent Expressions of Difficulty Behaviors	
Respondent uncertainty behaviors	*Guessing* in which a respondent provides an answer while expressing that he/she does not have sufficient information to ensure accuracy. *R: She – she – she lived with us until she was married, and, uh, I think she was 18, yeah.* *Estimate* in which a respondent provides an answer expressing his/her response is close to the actual response but is not completely accurate. *R: Well, It – it must've been about 1965.*
Respondent seeks clarification	Respondent indicates more information is needed to answer the question *Example: I: Oh, sure. Um, from February, 1952 until June, 1977, did you ever have a different main job than working for employer1? R: In other words, while I was working for employer1, did I get different jobs with the company?*

Respondent's expressions of difficulty	Respondent's expressions of frustration regarding the difficulty of the question. *Example: Uh, we're going to run into a little difficulty here. I hope you have experience with this because I don't know how to handle it.*

Unresolved don't know behavior is an **item non-response** component[29] and also is a part of the new coding scheme (see Table 3.3). In the models for each research question, *unresolved don't know* is employed as the only item non-response measure. For more detailed information on the new coding scheme see Appendix 1.

Table 3.3

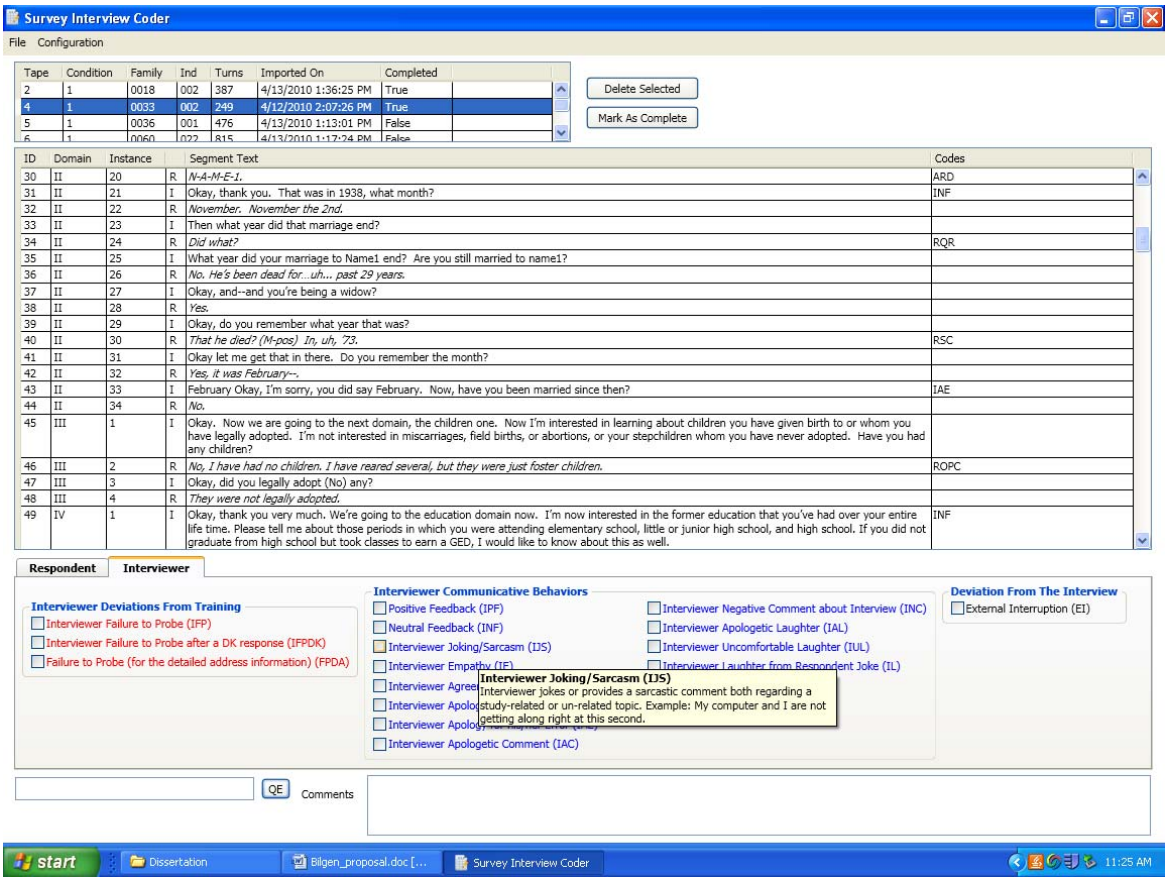
Item Nonresponse Measure

Behavior	Description and Example
Unresolved don't know	Respondent provides an explicit or implicit "don't know" response and the interviewer accepts the answer and fails to probe or fails to obtain an answer after probing for an answer. *Example 1: R: Uh… see I don't even remember the year when I broke all my ankle bones. Because then I got a metal plate and 2 screws in my hip. I: Wow, hmm, that sounds painful. R: All on the same side. I: M-pos. Yeah… R: I don't remember the year though. I: Let's just go ahead skip to the next one then. Did you ever smoke? (Interviewer accepts respondents' don't know response regarding the year of his/her health status change without providing any additional probing).* *Example 2: R: All I can tell you is it was on street1. I don't know the address of the house. I: Okay, that is fine. I am sorry, what city did you say that was in? (Interviewer accepts respondents' initial don't know response regarding one of his/her previous addresses without providing any additional probing).*

[29] Due to the low occurrences in the interviews, refusals are not included as a measure of item nonresponse.

Moreover, to implement the verbal behavior coding more efficiently, I hired a programmer to produce a coding program at the end of November 2009 (see Figure 3.1). According to the coder feedback, the coding program was able to speed up the behavior coding process and decrease coding errors as all the codes and definitions are provided in the program. This program includes the list of all interviewer and respondent codes, definitions, and acronyms that is provided in Tables 3.1, 3.2, and 3.3. In addition, for more detailed information regarding how to use the program see Appendix 3.

Figure 3.1 Coding Program Screenshot



Inter-Coder Reliabilities and Descriptive Statistics for the New Coding Scheme

In August 2010[30], the training and coding development process ended and the coding production started. During the training process, one of the coders dropped out of the study; however, a master's-level survey research student who helped with the initial coding scheme development joined the study subsequently. At the beginning of the coding production stage, I randomly assigned the transcripts at each condition to each of the four coders. The verbal behavior data collection process lasted approximately 5 months and the coding team completed the data collection on January, 2011.

Among 327 coded transcripts (165 calendar and 162 standardized interviews), I randomly selected and double-coded 10% of the transcripts from each condition (16 calendar and 16 standardized) to monitor the inter-coder reliabilities[31]. I used Kappa analyses to examine the overall inter-coder (i.e., inter-rater) reliabilities using both calendar and standardized interviews as a quality control measure. In the reliability analyses, I used turn as the level of analyses to identify potentially problematic behaviors and calculated kappa indices from 14,210 turns[32] ($N_{rel(calendar)}$ = 6,522, $N_{rel(standardized)}$ = 7,688). Kappa indices larger than 0.40 illustrate an acceptable inter-coder agreement (Bartko 1966; Cohen, 1960; Fleiss, 1971) and have been used as the reliability criterion

[30] Even though, the coding production started on May, 2010; I stopped the coding production due to the low reliabilities during the data collection monitoring stage. As a next step, the coding team discussed problematic transcripts, problematic behaviors and definitions during the retraining meetings. As a result of these meetings, the coding team updated some of the code definitions and examples and added more coding rules to the coding scheme. In our retraining meetings, we also coded several transcripts together and once the team came to an agreement on coding the problematic codes, the coding production was restarted at the end of August 2010. As the coding scheme has changed, the coders recoded the transcripts they had already coded earlier.

[31] Initially, I proposed to use a master-level student coder who was involved in the coding scheme production process to double code the transcripts for the reliability analyses. However, due to budget constraints as the principle investigator in this study, I double coded 32 transcripts for the quality control monitoring purposes.

[32] A turn is an uninterrupted stream of speech by either the interviewer or the respondent as identified by the transcribers.

in several behavior coding studies (Belli et al., 2004; Bilgen & Belli, 2010; Oksenberg et al., 1991; Presser & Blair, 1994). According to the inter-coder reliability analyses, overall Kappa values ranged between 0.31 and 0.96 (see Table 3.4) and were adequate (i.e., Kappa values were higher than 0.40) for 23 of 25 coded verbal behaviors. Overall Kappa values, which included both calendar and standardized interviews, were below 0.40 for *respondent uncomfortable laughter* and *respondent's attempts to resolve difficulty* behaviors; therefore, I excluded these variables from the further analyses.

In addition to the reliability analyses, I investigated the overall percentage of occurrence in order to exclude low occurrence variables from the further analyses. I included behaviors that occurred **at least in 100 turns or more in the 84,079 coded turns** (approximately 0.1% of the turns or more), and **occurred at least on average 1 or more times at each interview**, in the further analyses[33]. According to the turn-level means in Table 3.4, the only variable that did not fit the 0.1% criteria was the *respondent positive regard for the interview or questionnaire* behavior, which occurred in 19 turns among 84,079 turns (0.02 % and has a mean value of 0.0002). Furthermore, according to the interview-level means in Table 3.4, *respondent empathy* and *respondent laughter to interviewer's joke, sarcastic comment and feedback* occurred less than on average 1 times per interview. Therefore, I excluded *respondent positive regard*, *respondent empathy,* and *respondent laughter to interviewer's joke, sarcastic comment and feedback* from the further analyses.

[33] The exclusion criteria have been determined via examining the behaviors that did not correlate significantly with the remaining behaviors. Also, Mplus 6.1 (i.e. the statistical package used for the CFA models) treated these variables as a constant in the models due to their low means and variation.

Table 3.4

Overall Kappa Results and Descriptive Information for the Interviewer and Respondent Verbal Behaviors (New Coding Scheme)

Variable	Kappa Values	Turn level		Interview level[34]	
		MEAN	STD	MEAN	STD
Interviewer Deviation from Conventional Ideals					
I failure to probe	0.569	.009	.097	2.425	3.671
I failure to probe to DK response	0.864	.015	.122	3.872	3.872
Improper positive feedback	0.801	.028	.164	7.196	7.672
Interviewer Rapport Behaviors					
I neutral feedback	0.928	.065	.247	16.850	16.300
I joking and sarcasm	0.796	.009	.095	2.321	2.770
I empathy	0.580	.006	.076	1.508	2.363
I agreement	0.566	.005	.072	1.358	2.370
I direct apology[35]	0.808	.011	.103	2.755	3.424
I apologetic comment	0.412	.025	.155	6.336	7.867
I laughter to R joke/comment	0.856	.017	.128	4.339	5.885
Respondent Cooperative Behaviors					
R empathy	0.500	.003	.059	.887	1.892
R joking and sarcasm	0.809	.026	.160	6.737	7.813
R spontaneous attempts to resolve difficulty	0.308	.004	.062	.982	1.720

[34] In the multi-level analyses interview level is referred as respondent-level or level 2. Interview-level variables illustrate the count of each verbal behavior occurrence for each interview.

[35] *Interviewer direct apology* is the sum of *Interviewer apologizes from the respondent regarding the interview/task/question(/naire)/computer program* (IAP) and *interviewer apologies regarding his/her own error* (IAE) behaviors. Both behaviors are coded when interviewers specifically said "I am sorry"/ "sorry"/ "I apologize". Moreover, both IAP and IAE behaviors were also reliably coded (Kappa values > 0.4).

R offers or provides clarification	0.798	.218	.413	56.110	40.024
R corrections[36]	0.794	.034	.181	8.942	6.598
R laughter	0.807	.004	.061	.948	1.474
R positive regard	0.500	.000	.015	.058	.282
Respondent Non-Cooperative Behaviors					
R negative comment	0.546	.008	.087	1.982	4.527
R uncomfortable laughter	0.336	.003	.054	.737	2.162
Reluctance to provide information	0.459	.016	.125	4.064	7.092
Respondent Expressions of Difficulty Behaviors					
Guess	0.921	.086	.280	22.110	14.872
Approximation	0.944	.043	.203	11.052	10.047
R seeks clarification	0.850	.032	.175	8.159	7.507
R expressions of difficulty	0.580	.015	.120	3.768	4.468
Item Nonresponse Measure (One of the Dependent Variables)					
Unresolved DK	0.941	.037	.189	9.593	8.802
Total Number of Turns	14,210			166,346	
Total Number of Turns employed[37]	14,210			84079	
Total Number of Transcripts	32			327	

[36] *R corrections* behavior is the sum of *respondent spontaneous corrections of a response provided earlier* (RC) and *respondent corrections of study-related interviewer comment or assumption* (RCI). Both behaviors are coded when respondents spontaneously corrected a study-related comment/response. Moreover, both RC and RCI behaviors were also reliably coded (Kappa values > 0.4).

[37] In the final merged data set (which includes 327 transcripts), interviewer and following respondent turns are included as one turn as the further three-level multi-level analyses require both interviewer and respondent information to be included at each turn/case. This way, interviewer and the following respondent behaviors can be examined at the same case. My assumption here is each respondent behavior occurs right after each interviewer behavior. So, the order of the cases in data file is: I1R1, I2R2, I3R3…etc.

At the end of the verbal behavior selection process, 20 out of 25[38] (11 interviewer, 9 respondent behaviors out of 11 interviewer and 14 respondent) verbal behaviors were retained. Five behaviors, highlighted in Table 3.4, were excluded from the further analyses due to low occurrence or low inter-coder reliabilities.

Old Behavior Coding Scheme – Retrieval Behaviors

Table 3.5

Interviewer and Respondent Retrieval Strategies Behaviors coded using the old verbal behavior coding scheme

Behaviors	Descriptions
Interviewer Retrieval Probes	
Parallel	Interviewers use contemporaneous events from one life phase (such as residence) to recall events from another life phase (such as education). *Example: I: Okay, you was going to school. (Right) How many years did you live at--in a dormitory, sir? (Used in residence domain).*
Duration	Interviewers seek information regarding how long an event has occurred. *Example: Oh--okay. How long did you live at City3?*
Sequential	Interviewers ask respondents to recall events within the same life domain in the order of occurrence. *Example: Address6, City5? (Right) When you left there, where did you go?*
Timing[39]	Interviewers ask respondents to recall events within the same life domain in the order of occurrence. *Example: Do you remember the month you moved from Address7?*

[38] As it is indicated in Appendix 1, initially there are 27 (12 interviewer and 15 respondent) verbal behaviors; however as it is indicated in the following footnotes some of the behaviors are combined.
[39] In an earlier study which uses the old coding scheme and same transcripts, Belli and Bilgen (in progress) found that interviewer and respondent timing behaviors correlated poorly with the other interviewer and respondent retrieval behaviors and the authors decided to exclude interviewer timing from their retrieval scale. Therefore, I excluded interviewer and respondent timing from the further analyses.

Respondent Retrieval Strategies	
Respondent Parallel	Respondents spontaneously relate concurrent events from separate life domains. *Example: I lived at address2 until 1946, (M-pos) and then I went in the service.*
Duration Response	Respondents spontaneously provide the duration of an event. *Example: I lived there a little more than three years.*
Sequential Response	Respondents spontaneously relate thematically similar events that occurred right before or after each other. *Example: Okay, and when I came out of the Navy, I got a job at the employer3 across the river in city3.*
Timing Response[17]	Respondents spontaneously provide when an event or sequence of events started and ended. *Example: We were married until November the 8th, 1993. Uh, she passed away.*

Table 3.5 illustrates ***interviewer and respondent retrieval behaviors***, definitions and examples of these behaviors used for the cognition-related research questions. The retrieval behaviors were collected in a previous study conducted in 2007 that used the same transcripts. Five coders were randomly assigned to transcripts for each condition and coded relevant verbal behaviors in 327 transcribed tapes [for more information, see Bilgen and Belli (2010b)]. During the coding production process, the coders used a coding scheme that included 30 interviewer and 29 respondent behaviors. Among these behaviors, I used 4 interviewer and 4 respondent retrieval behaviors. Detailed definitions, further examples for each of these retrieval behaviors, and coding rules are described in detail in Appendix 2.

Inter-Coder Reliabilities and Descriptive Statistics for the Old Coding Scheme

According to Bilgen and Belli (2010b), out of 327 coded transcripts a master coder (one of the graduate students who had been involved in the research group) double-coded

randomly selected 10% of the transcripts (18 calendar and 14 standardized). The aim of the double coding was to monitor the inter-coder reliabilities, which were measured using Kappa analyses for each coded behavior. For all of the three interviewer and three respondent retrieval behaviors for the third set of research questions, the inter-coder reliability levels were adequate (Kappa values are higher than 0.40). In addition to reliability analyses, the overall percentage of occurrence has been investigated in order to exclude low occurrence variables from the further analyses. Behaviors **that occurred at least in 100 turns or more in the 84,079 coded turns** (i.e., approximately 0.1% of the turns or more), and **occurred at least on average 1 or more times at each interview**, are included in the further analyses. According to the turn-level and interview-level means illustrated in Table 3.6, all six (3 interviewer and 3 respondent) of the retrieval behaviors met the inclusion criteria and retained in the further analyses. In total, 26 (14 interviewer and 12 respondent) verbal behaviors from both old and new behavior coding scheme are used in the further analyses.

Table 3.6 *

Overall Kappa Results and Descriptive Information for the Interviewer and Respondent Verbal Behaviors (Old Coding Scheme)

Variable	Kappa Values	Turn level		Interview level[40]	
		MEAN	STD	MEAN	STD
Interviewer Retrieval Behaviors					
Parallel	0.620	.004	.067	6.691	7.135
Duration	0.852	.013	.115	7.749	6.346
Sequential	0.758	.015	.122	2.232	4.443

[40] In the multi-level analyses interview level is referred as respondent-level or level 2.

Respondent Retrieval Behaviors					
Respondent Parallel	0.648	.018	.134	8.865	9.795
Duration Response	0.750	.012	.111	6.260	5.964
Sequential Response	0.541	.010	.098	4.862	5.381
Total Number of Turns	13,968		165,795		
Total Number of Turns employed[41]	13,968		83,803		
Total Number of Transcripts	32		326[42]		

** The table is adapted from Bilgen and Belli (2010b), tables 4 & 5 on pg. 496-499.*

2. Measures

This study focuses on the response deficiencies that occur during data collection, which may potentially impact data quality. Hence, I will focus on one of the most permanently identified data quality measures, item non-response, throughout the dissertation.

2.1. Dependent Variables

Item non-response is measured using the coded *unresolved don't know (DK) response* behavior for each turn and this behavior serves as the outcome variable in the models. *Unresolved refusal response* is not included as a measure of item non-response as the coding team has decided early in the coding scheme production process that it does not

[41] In the final merged data set, interviewer and following respondent turns are included as one turn as the further three-level multi-level analyses require both interviewer and respondent information to be included at each turn/case. . This way, interviewer and the following respondent behaviors can be examined at the same case. My assumption here is each respondent behavior occurs right after each interviewer behavior. So, the order of the cases in data file is: I1R1, I2R2, I3R3…etc.

[42] During merging old and new behaviors, processing error is detected in one of the transcripts; therefore, this transcript (which includes 551 turns) is excluded from the analyses in this dissertation.

reach appropriate levels of variability due to low occurrence in the interviews. The *unresolved DK* variable is a dichotomous variable, such that a respondent either did provide or did not provide an unresolved DK response at a specific turn. Therefore, item non-response measure (*unresolved DKs*) varies among turns, respondents, and interviewers (see Table 3.7). For the three sets of research questions, separate sets of models are conducted (see Chapters V and VI). In the item non-response models, the *turn* of the speaker (interviewer or respondent) in the interviews are the unit of analyses (labeled as *ID* in Figure 3.1).

2.2. Independent Variables

The three key independent variables in this study are interviewer experience, exposure and interviewing technique. Interviewer exposure (experience gained during the administration of the survey throughout the survey fielding period) is also known as *"interview order"* in the literature (Hughes et al., 2002; Olson & Peytchev, 2007). The *interviewer exposure* variable refers to the cumulative number of interviews administered during one particular survey fielding period. Interviewer exposure is a continuous respondent-level predictor (see Table 3.7), which ranges from 1 through 37. For instance, interviewer exposure=1 when an interviewer administers his/her first interview and interviewer exposure=15 when an interviewer administers his/her fifteenth interview during a specific fielding period. In the study, this variable is constructed using the *interview date* and *time* information from the study records collected during the course of the data administration period (July through September 2002).

Interviewer experience is also retrieved from the study records and it refers to the interviewer experience with standardized interviews gained during an interviewer's

lifetime period. Studies on interviewer experience have argued that the most prevalent change in the effect of interviewer experience on data quality is between no or little general experience and some general experience (Mathiowetz & Cannell, 1980; Olson & Bilgen, 2011). Interviewer experience is a dichotomized interviewer-level predictor in which 0 is assigned to interviewers with less than one year of interviewing experience and 1 is assigned to interviewers with one or more than one year of interviewing experience (see Table 3.11). In addition, approximately 26% (29% in calendar and 22% in standardized condition) of the interviewers in this study had less than one year of experience. The last key independent variable interviewing technique is also a dichotomized interviewer-level variable, in which 0 is assigned to calendar interviews, and 1 is assigned to standardized interviews. Approximately half of the respondents were interviewed via calendar and the other half were interviewed via standardized interviews.

2.3. Control Variables

One limitation is of this study is that even though this is an experimental design, as the experiment was not geared towards exploring interviewer effects, there is a lack of interpenetration. The interpenetrated design method was developed by Mahalanobis (1946) and assigns households or respondents at random to interviewers (O'Muircheartaigh & Campanelli, 1999) in order to measure interviewer variance and separate the effects of the interviewer from the effects of other sources (such as regions). Therefore, in order to take into consideration the potential interviewer and respondent confounding effects, I include available interviewer characteristics (age, gender, and race) and respondent characteristics (age, gender, and race) into the model as control variables (see Table 3.11).

In the data set, available interviewer and respondent age are continuous measures. Whereas the mean of interviewer age is 49, the mean respondent age is 62[43]. In addition, both interviewer race (0-European-American and 1-African-American) and sex (0- men and 1- women) measures are dichotomous. About 76% of the interviewers are women and 11% of the interviewers are African-American (n=327). Furthermore, both respondent race (0-European-American and 1-African-American) and sex (0- men and 1- women) measures are also dichotomous. Approximately 47% of the respondents are women and 15% of the respondents are African-American (n=327).

2.4. Mediator variables

The interviewer and respondent behaviors that occur during the interview are used as mediator variables to explore each set of research questions. The models that examine whether these behaviors mediate the relationship between interviewer experience, exposure, and item non-response for each interviewing technique (calendar and standardized) are discussed in detail in the further chapters. Before fully explaining the mediator variables and how they play a role in the further analyses, the following sections (sections III, IV and V) discuss the construction of relevant interviewer and respondent verbal behavior scales for the purpose of synthesizing the relevant interviewer and respondent verbal behavior information for each research question. These verbal behavior scales are then used as mediator variables in the further multilevel analyses.

---

[43] The respondents in this study are all panel participants who have been Panel Survey of Income Dynamics (PSID) respondents for a relatively long period of time; therefore, the average respondent age of this study is higher than the average respondent age in several general U.S. population studies. Thus, the results of this dissertation cannot be generalized to the whole U.S. population (see Chapter VII for further discussion).

3. Psychometric Analyses for the Interviewer Verbal Behavior Scales

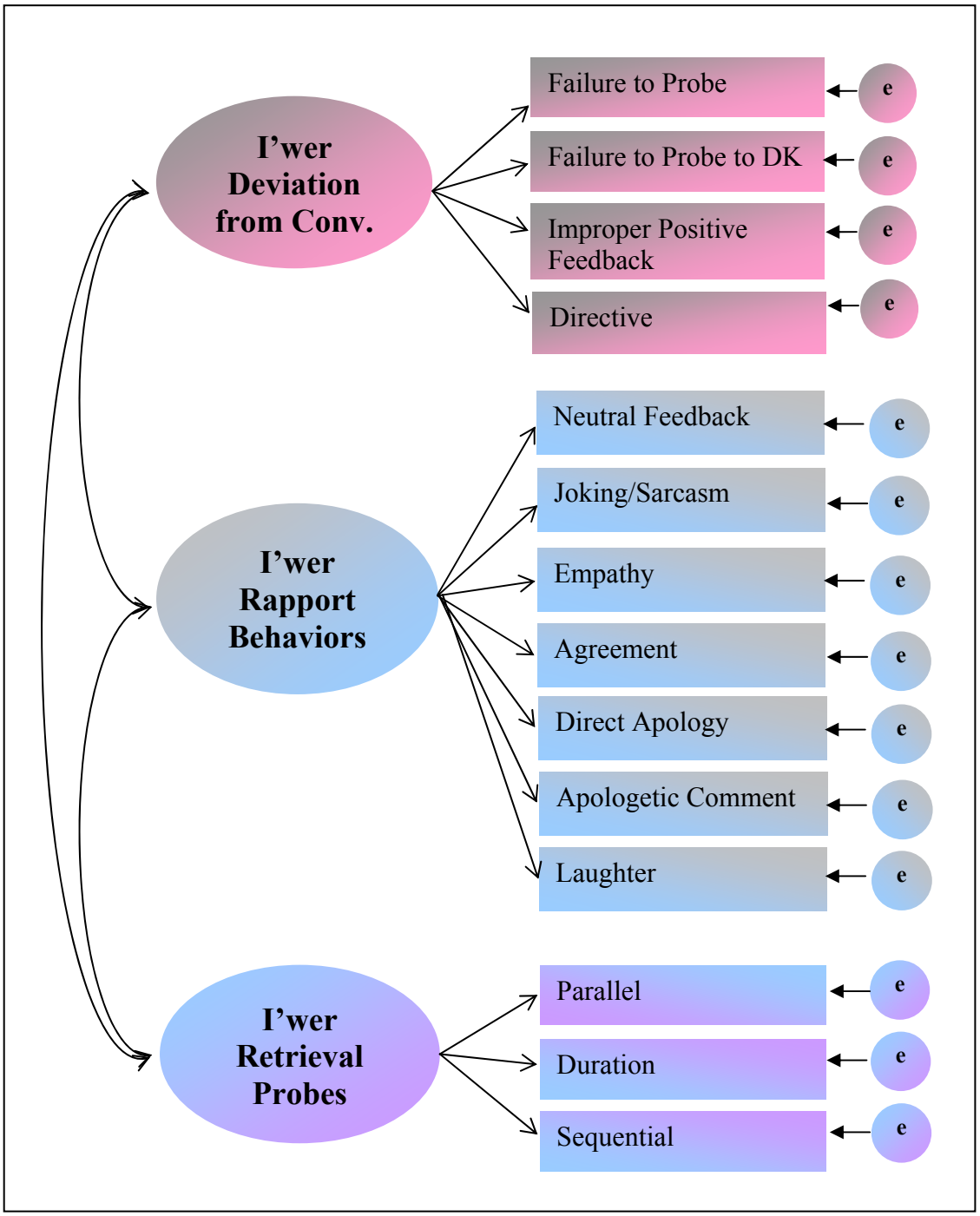
Taking into account the theoretical considerations, this study employed 14 interviewer and 12 respondent behaviors during the scale construction process. The 14 of the verbal behaviors were hypothesized to be explained by three main constructs that is a part of three sets of research questions concerning: 1- Interviewer-deviation from conventional ideals, 2-Interviewer Rapport, and 3- Interviewer Retrieval Strategies (see Figure 3.2).

As indicated earlier in the chapter, interviewer verbal behaviors were hypothesized to be multidimensional; therefore, at the first stage of the analyses, following questions were tested:

- *Are the Interviewer Verbal Behaviors Unidimensional or Multidimensional?*
  - *If it is multidimensional, does the two-factor model (retrieval versus communication behaviors, i.e., old versus new behaviors) or three-factor model (retrieval versus rapport versus deviation) fit the data best?*
- *Are the proposed verbal behaviors for each scale appropriately selected?*

As all of the turn-level verbal behaviors—planned to be used to create verbal behavior scales—were binary, initially Item Response Theory (IRT) models were proposed to be used to test: 1- whether the proposed verbal behaviors (*binary outcomes*) for each scale were appropriately selected, and 2- whether the scales were unidimensional as proposed (Embretson and Reise, 2000). However, the IRT models failed to capture the common variance of the relevant behaviors for each hypothesized scale/factor, as there was little verbal behavior variation at the turn-level (see Table 3.4 and 3.5). Therefore, the verbal behavior scale construction process continued at the interview-level (a.k.a. Level-2 or respondent-level).

Figure 3.2 Hypothesized Three-Factor Interviewer Verbal Behavior Scale Structure



Each of the interview-level (Level-2) verbal behaviors illustrated the count of each verbal behavior occurrence within each interview. During the examination of the distributions of 14 interviewer and 12 respondent verbal behaviors, the majority of the variables were observed to have zero-inflated negative binomial (ZINB) distributions. To the author's knowledge, none of the statistical packages conduct ZINB latent trait analyses efficiently. Thus, one of the limitations of this dissertation is that the verbal behaviors were assumed to be normally distributed (as almost all of the interview-level behaviors that are used to create scales were continuous). Hence, I conducted the analyses using Confirmatory Factor Analyses (CFA) approach in Mplus 6.1.

As a first step, I fitted a single-factor 14 behavior factor structure CFA model to examine whether the interviewer verbal behavior model is uni- or multidimensional. In the literature, several fit indexes were examined (Kline, 2005). In my dissertation, I use the most commonly used four modification indexes in order to evaluate the model fit. It has been indicated that *Comparative Fit Index* (CFI) and *Tucker-Lewis Index* (TLI) values higher than 0.90 indicates an adequate fit. Moreover, *Root Mean Square Residual* (RMSR) below 0.08 and *Root Mean Square of Approximation* (RMSEA) lower than 0.05 provide a good fit. However, it is also illustrated that RMSEA values below 0.08 indicates a "reasonable" model fit (Brown, 2006; Kline, 2005). The fit of the 14 interviewer behavior model was not acceptable, $\chi 2\ (77) = 534.755$, CFI = 0.524, TLI=0.437, SRMR = 0.116, RMSEA=0.135. Given that the fit of this unidimensional single-factor model was not acceptable, there was room for improving the model fit. Thus, these behaviors were divided into 2 factors: Retrieval Strategies (from the old coding scheme) versus Rapport and Deviation Behaviors (from the new coding scheme).

Table 3.7

CFA Interviewer Verbal Behavior Model Comparisons

Model	# Items	ML Chi-Square	DF	AIC	BIC	CFI	TLI	SRMR	RMSEA	Abs Chi Diff	DF Diff	Chi-Square Critical Value
Interviewer Single-Factor	14	534.755	77	27820.851	27980.029	0.524	0.437	0.116	0.135			
Interviewer Two-Factor	14	534.526	76	27822.621	27985.590	0.523	0.428	0.116	0.136	0.229	1	3.84
Interviewer Three-Factor	14	317.414	74	27609.509	27780.057	0.747	0.688	0.087	0.100	217.112	2	5.99
Interviewer Three-Factor Model Modifications												
- INF is excluded	13	230.773	62	24851.836	25011.014	0.810	0.760	0.080	0.091	86.641	12	21.03
- INF is excluded *- Directive included separately*	13	215.045	60	24840.108	25006.866	0.825	0.773	0.074	0.089	15.728	2	5.99
- IAC and INF are excluded *- Directive included separately*	12	153.032	49	22562.537	22717.925	0.875	0.832	0.06	0.081	62.013	11	19.68
DIR sep, w/out IAC and INF (WITH 1 error corr)	12	130.148	48	22541.653	22700.831	0.901	0.864	0.055	0.072	22.884	1	3.84
*DIR sep, w/out IAC and INF**(WITH 2 error corr)*	12	101.122	47	22514.627	22677.595	0.935	0.909	0.047	0.059	29.026	1	3.84

** Best Fitting Model

Given that the one-factor unidimensional model was nested within the two-factor model, these two models were compared to examine if the two-factor model improved the model fit. The fit of the two-factor model was also not acceptable, $\chi 2$ (76) = 534.526, CFI = 0.523, TLI=0.428, SRMR = 0.116, RMSEA=0.136. Also, the two factor model was not significantly better than the unidimensional one-factor model, $\chi 2$ difference (1) = 0.229 < 3.84 (see Table 3.7). Given that the fit of both single and two-factor models were poor, there was room for model improvement. Thus, the initial hypothesized model illustrated in Figure 3.2, *in which the behaviors were divided into 3 factors: 1- Retrieval Strategies, 2- Rapport Behaviors, and 3- Deviation Behaviors*, was tested.

As two-factor model was nested within the three-factor model, these two models were compared to examine if the three-factor model improved the model fit. The fit of the three-factor model was also not acceptable, $\chi 2$ (74) = 317.414, CFI = 0.747, TLI=0.688, SRMR = 0.087, RMSEA=0.100. However, this model was significantly better than the two-factor model, $\chi 2$ difference (2) = 217.112> $\chi 2$ table value=5.99 (see Table 3.7). This indicated that the interviewer verbal behavior was a multi-dimensional three-factor model.

The global fit of the 14-behavior three-factor model was not acceptable; therefore, there was still room for model improvement. As a result, the second step was to examine the local model fit by inspecting the standardized model residuals to identify specific problems regarding the correlation between each verbal behavior and its' corresponding predictor factor. The model residuals provide how far off the item correlations are from what the factor predicts (Kline, 2005). According to the model residuals, *Interviewer Neutral Feedback* behavior did not fit well with *Interviewer Rapport* (the predictor

factor) and there should have been a higher relation between the behavior and the factor than what was predicted in the model. This may be because interviewers were observed to use neutral feedback as silence fillers (such as "okay" or "fine") rather than attempt to increase rapport. Overall, this behavior did not correlate well with any of the verbal behaviors; therefore, it was decided to be excluded from the model.

In addition, according to the model residuals, *Interviewer Apologetic Comment* behavior did not also fit well with *Interviewer Rapport* and there should have been a higher relation between the behavior and the factor than what was predicted in the model. Also, it did not correlate well with either *Interviewer Direct Apology* or the other verbal behaviors in the model. This may be because coders were observed to include task-related feedback behaviors (such as, "hold on a minute, let me write that down") as *Interviewer Apologetic Comment*. Moreover, what is observed as *apologetic behaviors* by the coders may have been silence fillers; therefore, these behaviors may not be necessarily relevant for the purposes of this dissertation. As a result, *Interviewer Apologetic Comment* behavior was excluded from the model.

Lastly, according to local fit index (model residuals and modification index), directive behavior did not fit well with *Interviewer Deviation* (the predictor factor) and was observed not to correlate well with the other interviewer verbal behaviors that are predicted by *Interviewer Deviation* Factor (see Figure 3.2). Rather, *Directive* Behavior had a higher correlation with the retrieval behaviors. This may be because both interviewer directive behavior and interviewer behaviors, which were predicted by interviewer retrieval, belonged to the old coding scheme and were collected through the same study. Past studies indicated that directive behaviors are not a part of retrieval

strategies (Belli et al., 2004; Belli & Bilgen, in progress). Therefore, directive behavior was included as a separate single observed behavior in the model. Each of these changes in the model improved the model fit (see Table 3.7). However, there was still room for model improvement as the model fit for the three-level model with 12 interviewer verbal behaviors was mediocre (CFI = 0.875, TLI=0.832, SRMR = 0.060, RMSEA=0.081).

Thus, the modification indices have been examined to see how much of the $\chi 2$ would decrease by adding a particular model parameter. One of the suggestions of the modification indices was to correlate the residuals for *Interviewer Positive Feedback* and *Interviewer Jokes or Sarcasm* because some of the correlation between these behaviors was not explained by the three-factor model. The coding team observed that interviewers used improper positive feedback to build rapport with respondents, while deviating from conventional ideals. As a result, *Interviewer Positive Feedback* and *Interviewer Jokes or Sarcasm* behaviors were correlated in the new model.

The fit of the three-factor model with one error correlation was fairly acceptable, $\chi 2$ (48) = 130.148, CFI = 0.901, TLI=0.864, SRMR = 0.055, RMSEA=0.072 and was significantly better than the model without the error correlation, $\chi 2$ difference (1) = 22.884 > $\chi 2$ table (1) = 3.84 (see Table 3.7). One other suggestion from the modification indices was to correlate the residuals for *Sequential Probing* and *Interviewer Failure to Probe to DK responses*, because some of the correlation between these behaviors was not explained by the three-factor with one error correlation.

Also, the standardized model residuals indicated that there should have been a higher correlation between *Sequential Probing* and *Interviewer Failure to Probe to DK* behaviors than what was predicted. Sequential Probing requires interviewer ask about

thematically similar events in a chronological order. Therefore, the coding team observed that when interviewers started asking questions on similar events that occurred in a chronological order, interviewers increased their question administration pace, and sometimes ignored respondents' DK response. As a result, the residuals for the *Sequential Probing* and *Interviewer Failure to Probe to DK* behaviors were correlated in this model. The fit of the three-factor model with the two error correlation was adequate, $\chi 2$ (47) = 101.122, CFI = 0.935, TLI=0.909, SRMR = 0.047, RMSEA=0.059 and was significantly better than the three-factor model with one error correlation, $\chi 2$ difference (1) = 29.026 > $\chi 2$ table (1) = 3.84. Therefore, three-factor model with two error correlations was decided to be selected as the best-fitting model (see Table 3.7).

According to Table 3.8, the standardized factor loadings for the best fitting model ranged from 0.23 to 0.87 (and all of them were significant). This indicated that the common behavior correlation was significantly explained by the three factors. Also, all the estimates were within the bounds and there were no negative variances or covariances and the standardized factor loadings were smaller than one (see Table 3.8).

4. Psychometric Analyses for the Respondent Verbal Behavior Scales

The 12 of the verbal behaviors were hypothesized to be explained by four main constructs: 1- Respondent Cooperative Behaviors, 2-Respondent Non-Cooperative Behaviors, 3- Respondent Difficulty Behaviors, and 4- Respondent Retrieval Strategies (see Figure 3.3).

As indicated earlier in the chapter, respondent verbal behaviors were hypothesized to be multidimensional; therefore, at the first stage of the analyses, following questions were tested:

- *Are the Respondent Verbal Behaviors Unidimensional or Multidimensional?*
  - *If it is multidimensional, does the two-factor model (retrieval behaviors versus the rest of the behaviors), or three-factor model (cooperative versus difficulty and non-cooperative versus retrieval behaviors), or four-factor model (cooperative versus difficulty versus non-cooperative versus retrieval behaviors) fit the data best?*
- *Are the proposed verbal behaviors for each scale appropriately selected?*

In order to assess the extent to which the 12 behavior factor structure is unidimensional, a single-factor CFA model is fitted. The fit of this 12 single-factor behavior model was not acceptable, $\chi 2$ (54) = 304.231, CFI = 0.802, TLI=0.758, SRMR = 0.078, RMSEA=0.119. Given that the fit of this single-factor model was not acceptable, there was room for improving the model fit. Thus, these behaviors were divided into two factors: Retrieval Strategies (from the old coding scheme) versus Cooperation, Non-cooperation and Difficulty Behaviors (from the new coding scheme).

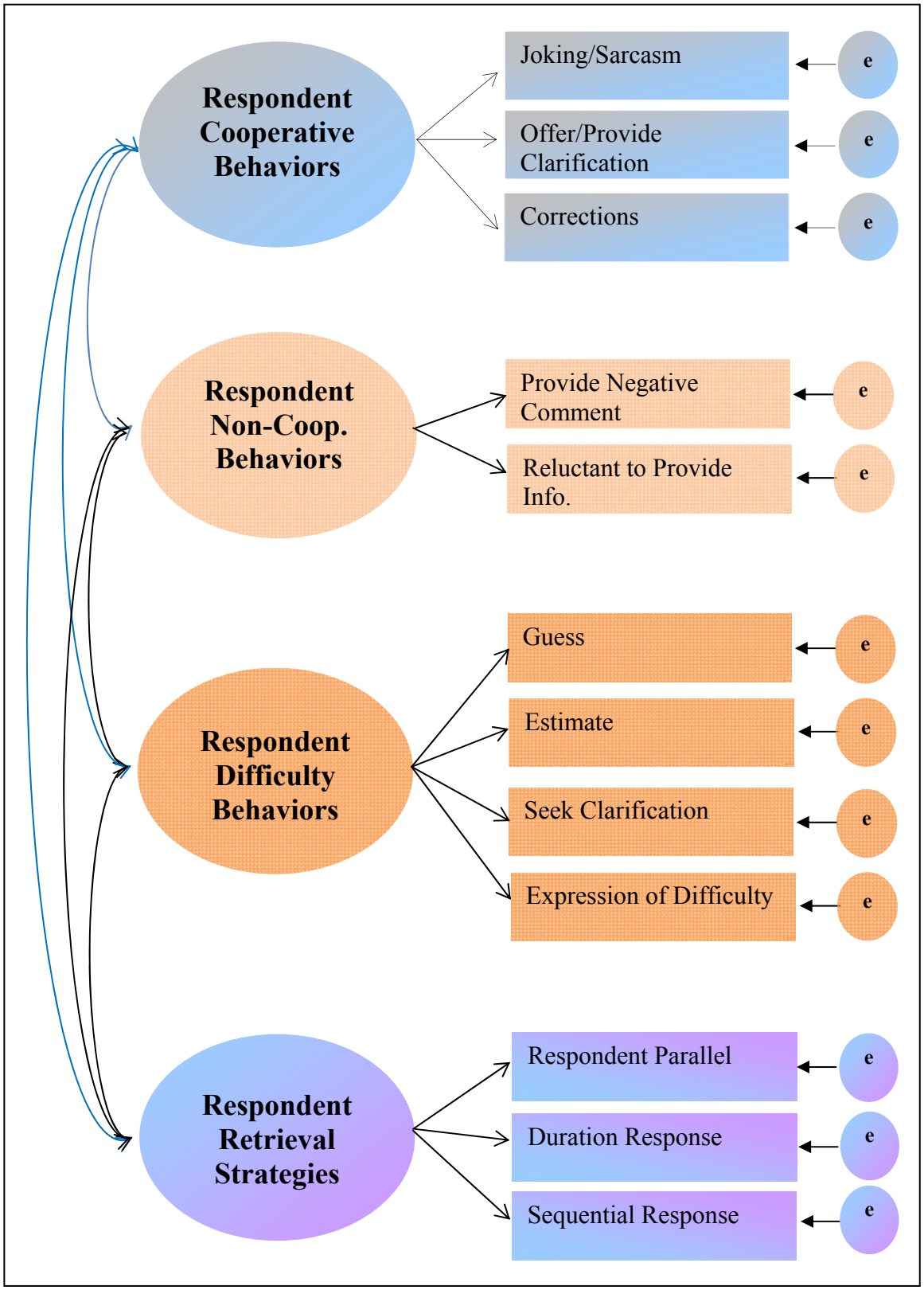
Table 3.8

Standardized Factor Loading Estimates from the Best Fitting Interviewer Verbal Behavior Model

Factors	Interviewer Verbal Behaviors	Loading Estimates	SE	p-values
I'wer Verbal Behavior (VB) Deviation	Failure to Probe	0.278	0.095	0.004
	Failure to Probe to DK	0.238	0.080	0.003
	Improper Positive Feedback	0.543	0.121	0.000
I'wer Rapport	Joking/Sarcasm	0.506	0.054	0.000
	Empathy	0.675	0.045	0.000
	Agreement	0.647	0.045	0.000
	Direct Apology	0.233	0.061	0.000
	Laughter	0.718	0.044	0.000
I'wer Retrieval	Parallel	0.506	0.045	0.000
	Duration	0.866	0.033	0.000
	Sequential	0.302	0.054	0.000
Factor 1 = I'wer VB Deviation Variance		1.000	0.000	999.000
Factor 2 = I'wer Rapport Variance		1.000	0.000	999.000
Factor 3 = I'wer Retrieval Variance		1.000	0.000	999.000
Directive Variance**		1.000	0.000	999.000
Factor 1 and Factor 2 Correlation		0.307	0.108	0.004
Factor 1 and Factor 3 Correlation		0.451	0.120	0.000
Factor 2 and Factor 3 Correlation		0.315	0.069	0.000
Directive and Factor 1 Correlation		0.466	0.107	0.000
Directive and Factor 2 Correlation		0.265	0.060	0.000
Directive and Factor 3 Correlation		0.881	0.033	0.000
Positive Feedback and Joking/Sarcasm Corr.		0.317	0.063	0.000
Failure to Probe to DK and Sequential Corr.		0.300	0.052	0.000

**Please note that directive is included in the model as a separate single observed behavior because it does not correlate well with the hypothesized deviation behaviors.

Figure 3.3 Hypothesized Four-Factor Respondent Verbal Behavior Scale Structure



As the single-factor unidimensional model was nested within the two-factor model, the two models were compared to examine whether the two-factor model improved the model fit. The fit of the two-factor model was also not acceptable, $\chi 2$ (53) = 292.205, CFI = 0.811, TLI=0.765, SRMR = 0.078, RMSEA=0.117. However, the two factor model was significantly better than the single-factor unidimensional model, $\chi 2$ difference (1) = 12.026 > $\chi 2$ table (1) = 3.84 (see Table 3.9). Since the fit of both single and two-factor models were poor, there was still room for model improvement. Thus, the 12 respondent verbal behaviors were divided into three factors: 1- Respondent Cooperative Behaviors, 2- Respondent Difficulty and Non-Cooperative Behaviors, and 3- Respondent Retrieval Strategies (see Figure 3.3)

As the two-factor model was nested within the three-factor multidimensional model, these two models were compared to examine if the three-factor model improved the model fit. The fit of the three factor-model was also not acceptable, $\chi 2$ (51) = 278.714, CFI = 0.820, TLI=0.767, SRMR = 0.076, RMSEA=0.117. However, the three-factor model was significantly better than the two-factor model, $\chi 2$ difference (2) = 13.491> $\chi 2$ table (2) = 5.99 (see Table 3.9). As the three-factor model did not have a good fit, there was still room for improvement in the model. Therefore, the initial hypothesized model illustrated in Figure 3.3, *in which the behaviors were divided into four factors:* 1- Respondent Cooperative Behaviors, 2-Respondent Difficulty Behaviors, 3- Respondent Non-Cooperative Behaviors, and 4- Respondent Retrieval Strategies (see Figure 3.3), was tested.

Table 3.9

CFA Respondent Verbal Behavior Model Comparisons

Model	# Items	ML Chi-Square	DF	AIC	BIC	CFI	TLI	SRMR	RMSEA	Abs Chi Diff	DF Diff	Chi-Square Critical Value
Respondent Single-Factor	12	304.231	54	26760.622	26897.06	0.802	0.758	0.078	0.119			
Respondent Two-Factor	12	292.205	53	26750.597	26890.825	0.811	0.765	0.078	0.117	12.026	1	3.84
Respondent Three-Factor	12	278.714	51	26741.105	26888.914	0.820	0.767	0.076	0.117	13.491	2	5.99
Respondent Four-Factor	12	180.778	48	26649.169	26808.348	0.895	0.856	0.060	0.092	97.936	3	7.81
Respondent Three-Factor - RPI is excluded	11	164.777	41	24535.341	24671.779	0.893	0.856	0.056	0.096	113.937	10	14.07
Respondent Three-Factor - RPI is excluded (WITH 1 error corr)	11	133.201	40	24505.765	24645.993	0.919	0.889	0.052	0.084	31.576	1	3.84
Respondent Three-Factor** - RPI is excluded (WITH 2 error corr)	11	112.48	39	24487.043	24631.062	0.936	0.910	0.048	0.076	20.721	1	3.84

** Best Fitting Model

As the three-factor model was nested within the four-factor model, these two models were compared to examine if the four-factor model improved the model fit. The fit of the four-factor model was also not acceptable, $\chi 2$ (48) = 180.778, CFI = 0.895, TLI=0.856, SRMR = 0.060, RMSEA=0.092. However, the four factor model was significantly better than the three-factor model, $\chi 2$ difference (3) = 97.936> $\chi 2$ table (3) = 7.81 (see Table 3.9). This indicated that the respondent verbal behavior was a multidimensional model.

The global fit of the 12-behavior four-factor model was not acceptable; therefore, there was still room for model improvement. As a result, the second step was to examine the local model fit by inspecting the residuals which provide the information regarding the correlation between each verbal behaviors and the predictor factor (Kline, 2005). According to the model residuals, *Reluctant to Provide Information* (RPI) behavior did not fit well with *Respondent Non-Cooperative Behaviors* (the factor it was hypothesized to predict) and there should have been a higher relation between the behavior and the factor than what was predicted in the model. This has been a problematic behavior as the coders indicated that this was not understood clearly. Even though we attempted to clarify this behavior by providing coding rules and specific definition and examples, the CFA models illustrated that this behavior did not correlate well with any of the respondent verbal behaviors. Therefore, it was excluded from the model. The exclusion of RPI behavior improved the model fit (see Table 3.7). However, this exclusion indicated that the four-factor model won't exist as one of the two variables, which were predicted by Respondent Non-Cooperative Behaviors, is excluded from the model. Therefore, the model fit for the three-level model with 11 respondent verbal behaviors,

and which excluded RPI, was tested. The exclusion of the RPI improved the three-factor model fit ($\chi 2$ difference (10) = 113.937> $\chi 2$ table (10) = 14.07). However, there was still room for model improvement, as the model fit for this three-factor model (which excluded RPI) was not acceptable (CFI = 0.893, TLI=0.856, SRMR = 0.056, RMSEA=0.096).

Thus, the modification indices have been examined to see how much of the $\chi 2$ decreased by adding a particular model parameter. One of the suggestions of the modification indices was to correlate the residuals for *Respondent Negative Comment* and *Guess* behaviors, because some of the correlation between these behaviors was not explained by the three-factor model. One of the coders indicated that she observed that some respondents got aggravated and provided a negative comment if their life history was relatively difficult and they did not know the exact response of many questions, which then usually led to guessing responses. As a result, the residuals for the *Respondent Negative Comment* and *Guess* behaviors were correlated in the new model. The fit of the three-factor model with one error correlation was fairly acceptable, $\chi 2$ (40) = 133.201, CFI = 0.919, TLI=0.889, SRMR = 0.052, RMSEA=0.084 and was significantly better than the model without the error correlation, $\chi 2$ difference (1) = 31.576 > $\chi 2$ table (1) = 3.84 (see Table 3.9).

One other suggestion from the modification indices was to correlate the residuals for the *Duration Response* and *Response Estimation* behaviors, because some of the correlation between these behaviors was not explained by the three-factor with one error correlation. Duration Response was coded when respondents spontaneously indicated

Table 3.10

Standardized Factor Loading Estimates from the Best Fitting Respondent Verbal Behavior Model

Factors	Respondent Verbal Behaviors	Loading Estimates	SE	p-values
Respondent Cooperative Behaviors	Joking/Sarcasm	0.632	0.038	0.000
	Offer/Provide Clarification	0.895	0.023	0.000
	Corrections	0.602	0.040	0.000
Respondent Difficulty and Non-Cooperative Behaviors	Guess	0.750	0.040	0.000
	Estimate	0.564	0.044	0.000
	Seek Clarification	0.426	0.052	0.000
	Expression of Difficulty	0.438	0.051	0.000
	Provide Negative Comment*	0.415	0.058	0.000
Respondent Retrieval	Parallel Response	0.839	0.027	0.000
	Duration Response	0.634	0.038	0.000
	Sequential Response	0.591	0.041	0.000
Factor 1 = R Cooperative Behaviors Mean		1.000	0.000	999.000
Factor 2 = R Difficulty Behaviors Mean		1.000	0.000	999.000
Factor 3 = R Retrieval Mean		1.000	0.000	999.000
Factor 1 and Factor 2 Correlation		0.798	0.044	0.000
Factor 1 and Factor 3 Correlation		0.883	0.031	0.000
Factor 2 and Factor 3 Correlation		0.792	0.044	0.000
Negative Comment and Guess Corr.		-0.438	0.081	0.000
Duration Response and Estimate Corr.		0.273	0.057	0.000

*Reluctant to Provide Information is excluded from the model. Therefore, Respondent Provides Negative Comment Behavior included as a part of the Difficulty Behavior Scale/Factor.

how long an event has occurred. Therefore, the coders observed that when respondents provided a length of time period rather than a specific time point, they tended to be uncertain about the exact time and were more likely to use phrases such as “about” or

"around," which indicates that respondents' answers were an estimate. Thus, the residuals for the *Duration Response* and *Response Estimation* behaviors are correlated in this model. The fit of the three-factor model with the two error correlation was adequate, $\chi 2$ (39) = 112.480, CFI = 0.936, TLI=0.910, SRMR = 0.048, RMSEA=0.076 and was significantly better than the three-factor model with one error correlation, $\chi 2$ difference (1) = 20.721 > $\chi 2$ table (1) = 3.84. To sum up, three-factor model with two error correlations was decided to be selected as the best-fitting model for the respondent verbal behaviors (see Table 3.9). According to Table 3.10, the standardized factor loadings for the best fitting model ranged from 0.41 to 0.89 (and all of them are significant). This indicated that the common behavior correlation was significantly explained by the three-factors. Also, all of the estimates were within the bounds, there were no negative variances or covariances, and the standardized factor loadings were smaller than one (see Table 3.10).

5. Construction of Mediator Variables/Scales

The first set of research questions were explored using interviewing pace and a group of verbal behaviors that intended to measure interviewers' deviation from conventional ideals. The interviewing pace[44] was constructed via *number of words used in the interview/ length of interview (in minutes)*, and was a continuous variable that varied among respondents and interviewers. Whereas the interviewer verbal behavior deviation

[44] Olson and Bilgen (2011) investigated whether interview length—which was explored as a measure of interviewing pace—mediated the relationship between interviewer experience and data quality. The authors found that after accounting for the interviewing length, the relationship between the interviewer experience and data quality remained. Authors suggested that there may be other mediating mechanisms that can explain the relationship between the experience and data quality. Therefore, the idea of studying the interviewer and respondent verbal behaviors as mediating mechanisms in my dissertation is an extension of the mediation analyses in Olson and Bilgen (2011) paper.

scale was created using the sum of: *failure to probe (an item or part of an item), failure to probe a DK response,* and *improper positive feedback*. In addition, *directive* verbal behavior was examined by itself as it was explained earlier in the chapter. These verbal behaviors initially were all dichotomous variables at the turn-level, in which 0 was assigned when a behavior was not observed, and 1 was assigned when behavior was observed in a specific turn, and varied among turns, respondents, and interviewers. However, as explained earlier, due to low occurrences at the turn-level they were included at the aggregate level (i.e., number of behaviors occurred at the interview/respondent-level). Hence, these behaviors were included as count measures and they varied among respondents and interviewers in the models (see Table 3.11).

Three sets of behaviors, 1-Interviewer rapport, 2-Respondent cooperative, and 3-Respondent difficulty and non-cooperative behaviors, were used in the models conducted to explore interviewer and respondent interpersonal communication. The interviewer rapport scale was created using the sum of interviewer behaviors such as *joking/sarcasm, empathy, agreement, direct apology,* and *laughter to a respondent joke or comment.* Each of these five interviewer rapport behaviors were binary variables in which 0 was assigned when interviewer behavior was not observed and 1 was assigned when the interviewer behavior was observed in a specific turn.

Similarly, the respondent cooperation scale was created via the sum of respondent behaviors such as *joking and sarcasm*, *spontaneously offering or providing clarification,* and *spontaneous corrections.* These respondent cooperation variables were all dichotomous variables (0-not observed; 1-observed in a specific turn). Lastly, the respondent difficulty and non-cooperation scale was created using respondent difficulty

Table 3.11

Measures and Levels of the Measures Employed in this Study

MEASURES	LEVEL(S) MEASURE VARIES		
	LEVEL 1 (Turn)	LEVEL 2 (Respondent)	LEVEL 3 (Interviewer)
Dependent Variable			
Item non-response (Unresolved DK) *0- Not occurred,* *1- Occurred within the turn*	X	X	X
Independent Variables			
LN (Interviewer Exposure)		X	X
Interviewer Experience *0- Less than 1 year,* *1- 1 year or more*			X
Interviewing Technique *0- Calendar,* *1- Standardized*			X
Mediator Variables			
1- Interviewer's Deviation From Conventional Ideals			
Interviewing Pace		X	X
I'wer VB Deviation Scale	X	X	X
I'wer Directive Behavior	X	X	X
2- Interviewer and Respondent Interpersonal Communication			
I'wer Rapport Scale	X	X	X
R Cooperation Scale	X	X	X
R Difficulty and Non-Coop Scale	X	X	X
3- Interviewer and Respondent Retrieval Behaviors			
Interviewer Retrieval Scale	X	X	X
Respondent Retrieval Scale	X	X	X
Control Variables			
Interviewer Age, Sex, Race			X
Respondent Age, Sex, Race		X	X

behaviors such as *uncertainty* (*guess* and *approximation)*, *expressions of difficulty*, and *seeking clarification* and respondent non-cooperative behavior such as *negative comment about the interview/task/questionnaire, uncomfortable laughter.* Initially, these behaviors were all dichotomous and collected at the turn level. However, similar to deviation behaviors, due to low occurrences at the turn-level they were included at the aggregate level (i.e., number of behaviors occurred at the interview/respondent-level). Hence, these behaviors were included as count measures and they varied among respondents and interviewers in the explored models (see Table 3.11).

Lastly, in the models that examined retrieval strategies, two sets of behaviors, 1- Interviewer retrieval probes and 2- respondent retrieval behaviors were explored. The interviewer retrieval scale was created using the sum of an interviewer's *duration, parallel,* and *sequential probes.* The respondent retrieval scale was constructed via the sum of a respondent's *parallel*, *duration*, and *sequential retrieval strategies.* Initially, all of the interviewer and respondent retrieval verbal behaviors were binary variables (*0- not observed, 1- observed*) and collected at the turn level. However, similar to interpersonal communication behaviors, due to low occurrences at the turn-level they were included at the aggregate level (i.e., number of behaviors occurred at the interview/respondent-level). Hence, these behaviors were included as count measures and they varied among respondents and interviewers in the explored models (see Table 3.11).

As explained in detail earlier, all scales (interviewer and respondent retrieval scales, interviewer rapport, respondent cooperation, and respondent difficulty and noncooperation scales) were examined at the interview-level (i.e., level 2 / respondent-level) due to low occurrences at the turn-level. Nevertheless, item non-response

dependent variable (*respondent's unresolved DK behavior*) was included at the turn-level, because response errors occur at the question-level. Also, the unexplained within-interview (i.e., within-respondent) variation among the item nonresponse measure has been taken into account by the turn-level models (see Table 3.11).

CHAPTER IV: THEORETICAL FRAMEWORK

HYPOTHESES AND EXPECTED FINDINGS FOR THREE SETS OF RESEARCH QUESTIONS

1. Hypotheses regarding Deviation from Conventional Ideals

Interviewer experience is a continually changing measure, as experience levels of interviewers change while gaining exposure during the survey fielding period of a study (Hughes et al., 2002; Olson & Peytchev, 2007). Therefore, this study focuses on exploring the interactive effect of experience and exposure rather than their individual effects. The first set of analyses focus on how interviewers shape their behaviors as they gain experience and exposure. In particular, as illustrated in Chapter III, this study examines the *interviewer verbal behavior deviation scale, interviewer directive probes* and *interviewing pace* –i.e., measures of ***deviation from conventional ideals***–individually in the models that investigate the effect of interviewer experience and exposure on item non-response in calendar and standardized interviews.

Although this study evaluates the three measures of deviation from conventional ideals individually, I expect each of the three measures to behave similarly in the analyses. Therefore, Figure 4.1 indicates that the direction of the hypotheses, regarding each of the three "deviation from conventional ideals" measures' relationship with experience and exposure, are similar and each of the three measures are illustrated as one measure and referred as "deviation from conventional ideals" in Figure 4.1. **Specific**

hypotheses regarding the models in which *interviewer deviation from conventional ideals* measures play an intervening role:

Pre-hypothesis 1: *In standardized interviews, as within-study interviewer exposure increases, interviewers who are inexperienced will significantly deviate from conventional ideals* ***($\beta_1 > 0$ in Figure 4.1)*** *as there is no general experience to mitigate the impact of the within-study exposure and there is more opportunity to learn and develop their own behaviors and strategies. In standardized interviews, as the experienced interviewers will rely on their habitual interviewing behaviors, there is less room for them to learn new behaviors. Therefore, experienced interviewers will not significantly change their behaviors during the interview because of within-study exposure* ***($\beta_A = 0$ in Figure 4.1)****.*

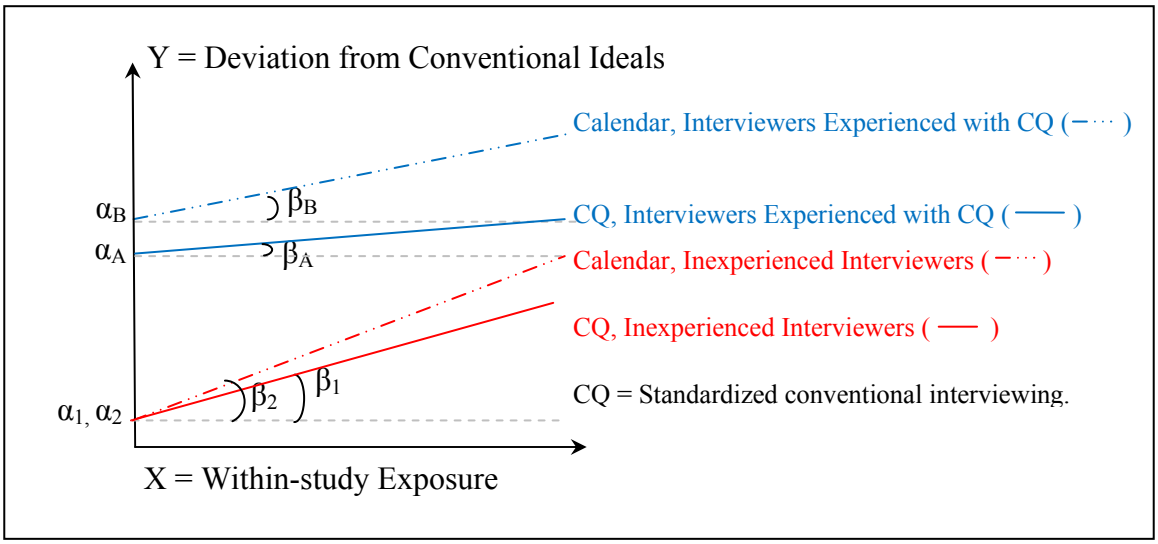
Pre-hypothesis 2: *Due to its flexible nature, there is more room for interviewer behavior change during the survey fielding period in calendar interviews than in standardized interviews, regardless of interviewing experience. Hence, in calendar interviews, both experienced and inexperienced interviewers will significantly deviate from conventional ideals due to within-study interviewer exposure* ***($\beta_2 > 0$ and $\beta_B > 0$ in Figure 4.1).***

H1: In both calendar and standardized interviews, interviewers with general experience will rely on their interviewing habits more than inexperienced interviewers. Thus, interviewers with general experience will deviate more at the beginning of the survey fielding period than inexperienced interviewers **($\alpha_A > \alpha_1$ and $\alpha_B > \alpha_A$ in Figure 4.1).**

H2: In both calendar and standardized interviews, at the end of the data collection period the deviation gap decreases between the experienced and inexperienced interviews.

Inexperienced interviewers deviate more due to within-study exposure during the study fielding period from the conventional ideals in comparison to experienced interviewers ($\beta_1 > \beta_A$ and $\beta_2 > \beta_B$ **in Figure 4.1**, indicates a plausible significant interaction effect between general interviewer experience and within-study interviewer exposure).

Figure 4.1[45] **Expected Direction of the Hypotheses 1 through 3**



Calendar interviews aim to allow conversational flexibility and interviewer independence. Consistently, overall interviewer deviations from conventional ideals were found to be more prevalent in calendar interviews than standardized interviews (Belli, et al., 2004; Bilgen & Belli, 2010b). This dissertation continues examining these findings and explores how different types of experience relate to interviewer deviations from conventional ideals when they are administering different types of interviews.

[45] Even though the relationship has been illustrated as linear in this graph; the relationship between *deviation from conventional ideals* (Y-axis) and *within-study interviewer exposure* (X-axis) might be a log-linear or polynomial relationship. As the interviewers would learn more from their experiences at the beginning of the data collection period, the increase in the deviation from conventional ideals might eventually decelerate during the data collection period (depending on the length of the data collection process) and the increase in the deviation from conventional ideals might eventually decrease or stop completely. This is discussed more in detail in the results sections of the dissertation (Chapters V and VI).

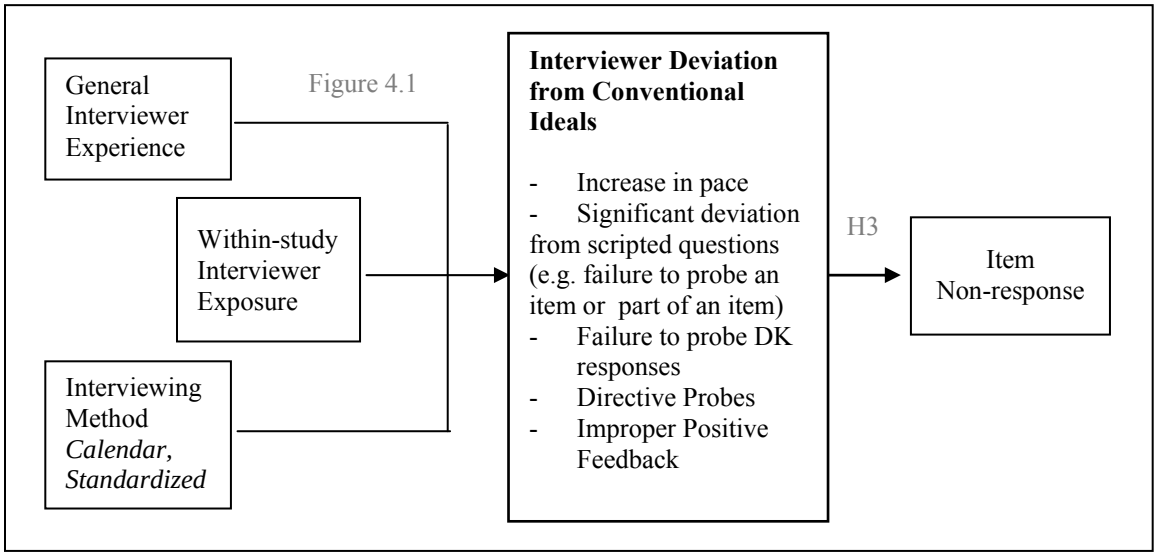
H3: Thus, calendar interviewers will deviate significantly more from conventional ideals due to within-study interviewer exposure in comparison to standardized interviewers, regardless of general interviewer experience **($\beta_2 > \beta_1$ and $\beta_B > \beta_A$ in Figure 4.1).**

As emphasized in the literature review, there is no clear evidence on how the change in interviewer behaviors due to experience and exposure affect response errors that may potentially impact data quality. Research on the relationship between interviewer behavior and data quality illustrates that interviewer deviations from the script (such as failure to probe, decrease in providing feedback, and increased digressions) may increase (Cannell, Miller, & Oksenberg, 1981; Henson, Cannell, & Lawson, 1976) or slightly decrease (Dykema, Lepkowski, & Blixt, 1997) response errors. Specifically, one data quality measure–item nonresponse–has been found to relate to deviation from conventional ideals (Bradburn, Sudman, & Associates, 1979; Cannell et al., 1981). However, the findings regarding this relationship are mixed. For instance, Bradburn, Sudman, and Associates (1979) point out that providing positive feedback may encourage respondents to feel more comfortable and provide additional responses to threatening questions, whereas Cannell et al. (1981) mention that positive feedback provided after an item non-response earlier in the interview may encourage and increase respondents' item nonresponse behavior later in the interview. Moreover, Bradburn, Sudman, and Associates (1979) also suggest that the speech variations and probing failures/errors are interviewers' reflection of respondent anxiety and uneasiness cues; hence, related to increased item non-response. However, Henson et al. (1976) argue that speech variations and probing failures/errors are interviewers' solution to decrease respondent uneasiness and difficulties; hence, decrease item nonresponse (for more

detailed discussion see Chapter II). To sum up, the main idea is that the theories regarding interviewer deviation and item nonresponse relationship are mixed. Hence, I would expect:

H4: The deviation from conventional ideals *(such as increase in interviewing pace, failure to probe, increase in interviewers' improper feedback and directive probing)* significantly changes (i.e., may increase or decrease) *item non-response*.

Figure 4.2 Expected Role of *Interviewer Deviation from Conventional Ideals* in the Item NR Models for Calendar and Standardized Interviews



To my knowledge none of the studies fully explored these two pieces together (1- interviewer experience, exposure and deviation behavior association, and 2- deviation associated behaviors and item non-response). This study aims to fill in this gap (see Figure 4.2). Additionally, all of the studies that examine the relationship between interviewer experience, exposure, and item non-response (or other response errors) focus on standardized interviews. As it is unclear how this relationship plays a role in flexible interviewing techniques, it appears promising to explore how interviewer strategies,

which are developed when interviewers gain experience and exposure, relate to item nonresponse when researchers use different interviewing techniques (namely, standardized versus flexible calendar interviews).

Lastly, I combine all the pieces of the first four hypotheses together:

H5: Overall ***(regardless of interviewer exposure)***, rate of ***item non-response*** change due to ***interviewer deviation behaviors*** will be:

- Higher among experienced interviewers than inexperienced interviewers (in both calendar and standardized interviews).
- Higher among calendar interviewers than standardized interviewers (in both interviewer experience levels).

H6: *Rate of* ***item non-response*** change due to ***interviewer deviation behaviors*** in early studies versus in later studies (low vs. high within-study interviewer exposure) will be:

- Higher among inexperienced interviewers with higher exposure than the experienced interviewers with higher exposure (in both calendar and standardized interviews).
- Higher among calendar interviewers with higher exposure than the standardized interviewers with higher exposure (regardless of the experience levels).

2. Hypotheses regarding Interviewer and Respondent Interpersonal Dynamics

One of the purposes of this dissertation is to explore how interviewer experience and exposure interaction play a role in shaping interviewer and respondent interpersonal dynamics. Figure 4.3 illustrates the hypotheses regarding how I would expect overall interviewer rapport behaviors (measured via *interviewer rapport scale*) to work in the

models that aim to explore the second set of research questions on interviewer and respondent interpersonal dynamics that occur during an interview.

My specific hypotheses are:

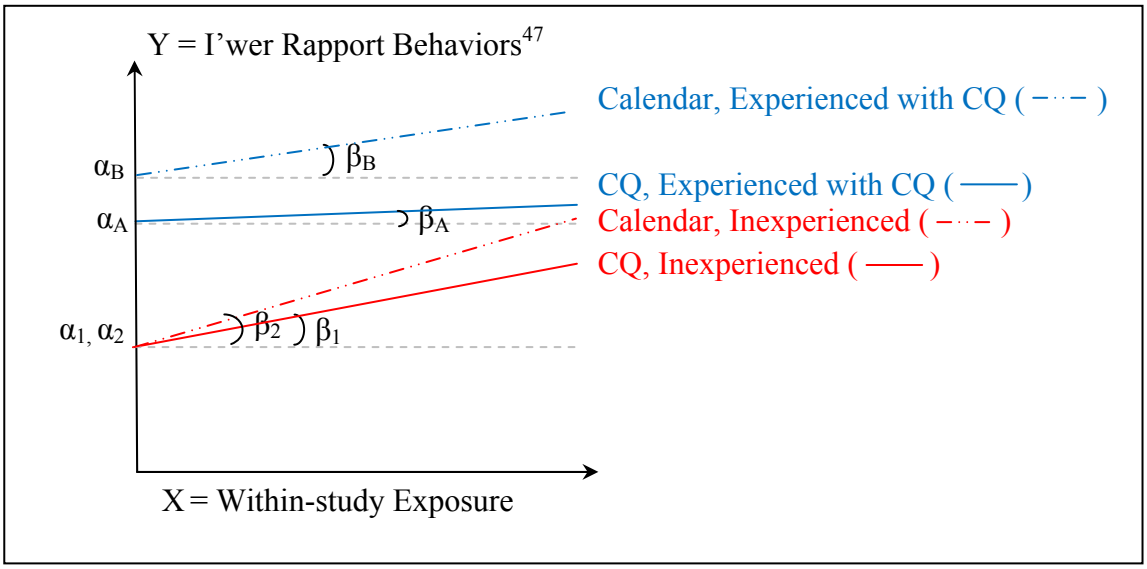
> ***Pre-hypothesis 3:*** *In standardized interviews, inexperienced interviewers develop rapport behaviors and learn how to communicate better during the administration of a particular survey fielding period. Hence, inexperienced interviewers will use significantly more rapport behaviors at the end of the data collection period in comparison to the beginning of the data collection period* ***(β1 > 0 in Figure 4.3)****. In standardized interviews, as experienced interviewers use their old set of rapport behaviors acquired or adapted in earlier studies, there is slight room for the modification of these behaviors. Hence, experienced interviewers are not expected to significantly change their rapport behaviors during the survey fielding period* ***(βA = 0 in Figure 4.3).***
>
> ***Pre-hypothesis 4:*** *All calendar interviewers, regardless of the general interviewing experience levels, will significantly use more rapport behaviors due to within-study interviewer exposure* ***($\beta_2 > 0$ and $\beta_B > 0$ in Figure 4.3).***

H7: In both standardized and calendar interviews, as experienced interviewers bring their previously learned and modified set of rapport behaviors into the new interviews, they will use higher numbers of rapport behaviors at the beginning of the data collection period than inexperienced interviewers **($\alpha_A > \alpha 1$ and $\alpha_B > \alpha_A$ in Figure 4.3).**

H8: In both standardized and calendar interviews, the gap in the use of rapport behaviors will diminish between the experienced and inexperienced interviewers at the end of the data collection period. The increase in inexperienced interviewers' use of rapport

behaviors will be higher in comparison to experienced interviewers due to within-study interviewer exposure during the study fielding period **($\beta_1 > \beta_A$ and $\beta_2 > \beta_B$ in Figure 4.3)**.

Figure 4.3[46] Expected Direction of the Hypotheses 7 through 9



All of the studies that examine the relationship between interviewer experience, exposure, and data quality focus on standardized interviews. Hence, it is promising to evaluate how interviewer strategies, which are developed as they gain experience and exposure, relate to item non-response when researchers use different interviewing techniques (i.e., standardized versus calendar interviews). Calendar interviews allow conversational flexibility and interviewer independence, so overall respondent rapport (i.e., cooperative) behaviors are more prevalent in calendar interviews and respondent expressions of difficulty and non-cooperative behaviors are more common in

[46] Even though the relationship has been illustrated as linear in this graph, the relationship between *interviewer rapport* (Y-axis) and *within-study interviewer exposure* (X-axis) might be a log-linear or polynomial relationship (see Chapters V and VI for further discussion).

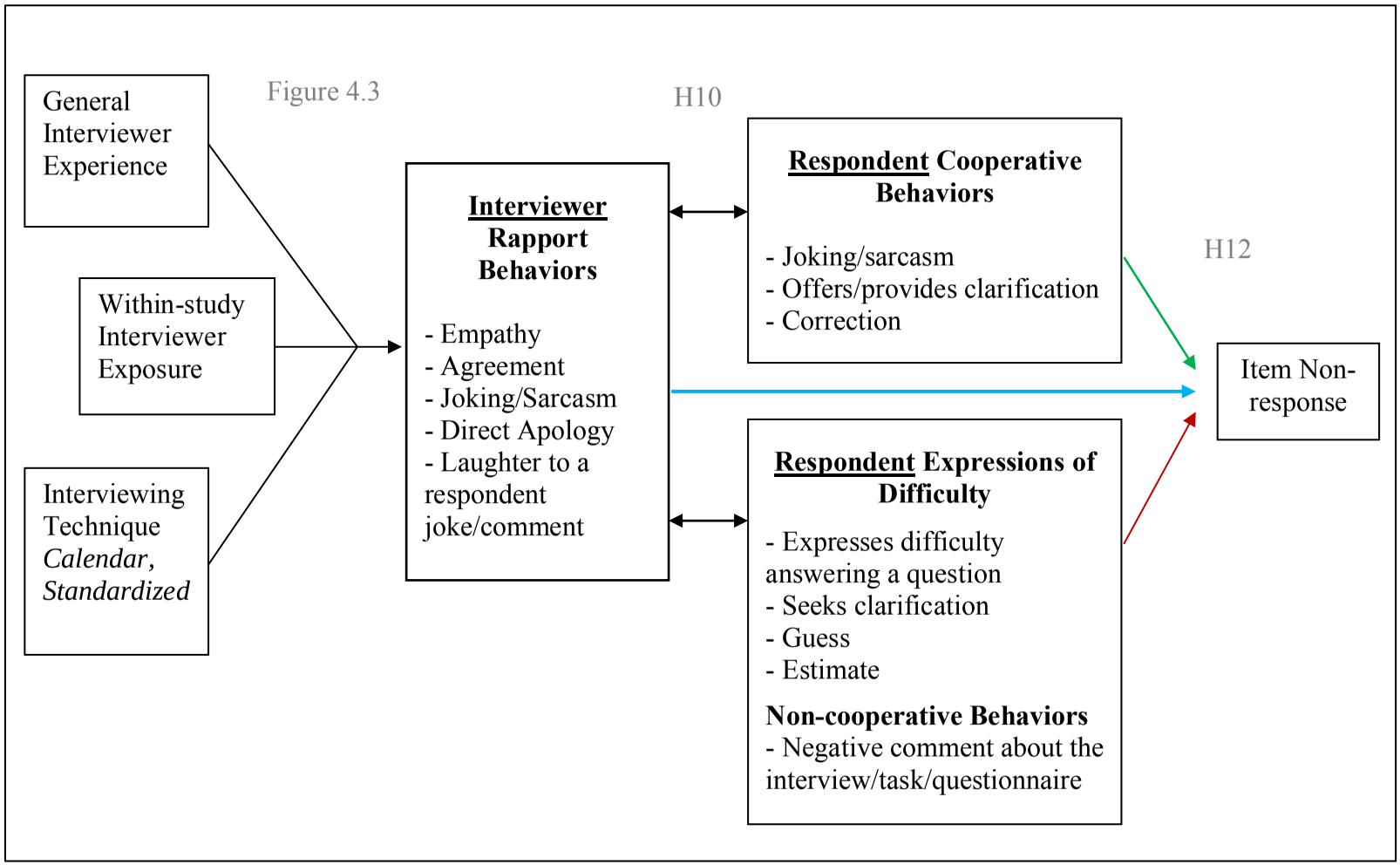
[47] The studies in the survey research literature mainly focused on experience, exposure, and respondent rapport, interest, or motivation. To my knowledge, none of these studies focused on the association between experience, exposure, and interviewer rapport. This study also intends to fill in this gap.

standardized interviews (Belli, et al., 2004; Bilgen & Belli, 2010b). I will continue examining these findings by exploring how interviewer experience and exposure interaction affect interviewer rapport behaviors when they are administering calendar and standardized interviews. Thus, the specific hypothesis regarding this is:

H9: As there is more room for interviewer behavior change in calendar interviews than in standardized interviews throughout the survey fielding period, the increase in the use of interviewer rapport behaviors due to within-study interviewer exposure is larger in calendar interviews than in standardized interviews, regardless of general interviewing experience **($\beta_2 > \beta_1$ and $\beta_B > \beta_A$ in Figure 4.3).**

As discussed in detail in the literature review, no clear evidence exists on how the interviewer and respondent rapport behaviors relate to change in item non-response due to interviewer experience and exposure. No studies empirically test how all of these pieces relate to each other. Thus, as a second step I propose to test the relationship between interviewer rapport behaviors and item non-response controlling for the respondent cooperative, non-cooperative, cognitive difficulty behaviors (see Figure 4.4). As discussed in detail earlier in Chapter III, each set of behaviors will be combined to produce three measures (1- Interviewer rapport scale, 2- Respondent cooperation scale, 3- Respondent non-cooperation, and difficulty scale). The expected direction in Figure 4.4 indicates how I would expect the interviewer and respondent interpersonal behaviors to interact with interviewer experience and exposure and how this interaction translates into the change in item nonresponse.

Figure 4.4 Expected Role of *Interviewer and Respondent Interpersonal Dynamics* in the Item Non-response Models for Calendar and Standardized Interviews



Hence, my specific hypotheses are:

H10: There is a positive correlation between interviewer rapport behaviors and respondent cooperative behaviors, and a negative correlation between interviewer rapport behaviors and respondent expression of difficulty or non-cooperative behaviors.

H11: The increase in interviewer rapport and respondent cooperative behaviors AND decrease in respondent expression of difficulty and non-cooperative behaviors will decrease *item non-response* (see Figure 4.4).

Lastly, I combine all the pieces of the first five hypotheses together (see Figure 4.4):

H12: Overall ***(regardless of interviewer exposure)***, the rate of ***item non-response*** decrease due to ***interviewer and respondent communicative behaviors*** will be:

- Higher among experienced interviewers than inexperienced interviewers (in both calendar and standardized interviews).
- Higher among calendar interviewers than standardized interviewers (in both interviewer experience levels).

H13: The rate of ***item non-response*** decrease due to ***communicative behaviors*** in early studies versus in later studies (low vs. high within-study interviewer exposure) will be:

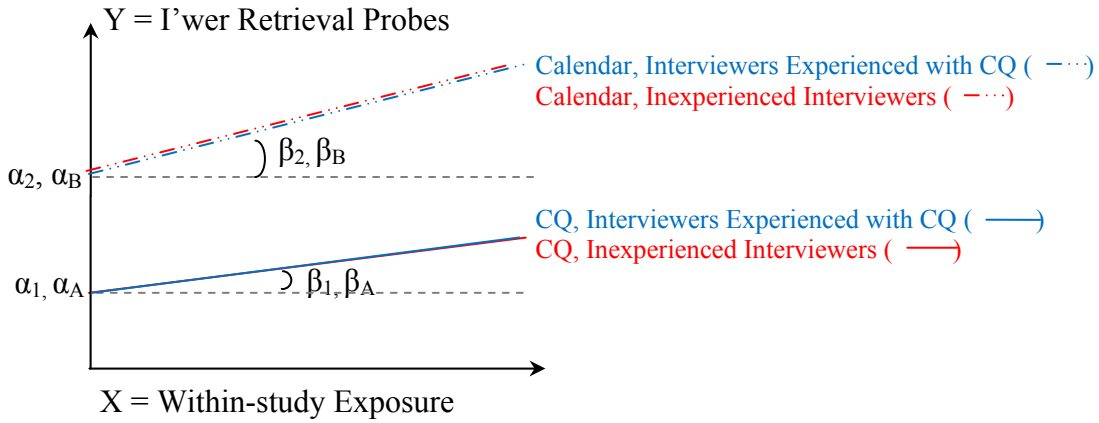
- Higher among inexperienced interviewers with higher exposure than the experienced interviewers with higher exposure (in both calendar and standardized interviews).
- Higher among calendar interviewers with higher exposure than the standardized interviewers with higher exposure (regardless of the experience levels).

3. Hypotheses regarding Retrieval Strategies and Probes

Bilgen, Belli, and Olson (2009) examined how interviewer experience and exposure interact in relation to interviewer retrieval probes. The last set of research questions is a

continuation of the Bilgen, Belli, and Olson (2009) findings that are illustrated in detail in the literature review section of Chapter II. Calendar interviewing is designed to use the structure of autobiographical memory. It has been found that **overall** both interviewer retrieval probes and respondent retrieval strategies are used more prevalently in the calendar interviews in comparison to standardized interviews (Belli et al., 2004; Bilgen & Belli, 2010b). Moreover, Bilgen, Belli, and Olson (2009) examined how interviewer experience and exposure play a role together during the use of interviewer retrieval probes in calendar and standardized interviews. The authors found that, at the beginning of the study there are no differences in interviewers' use of retrieval probes between the inexperienced and experienced interviewers in both calendar and standardized interviews.

Figure 4.5[48] **Overall Findings of Bilgen, Belli and Olson (2009)**



Bilgen, Belli, and Olson (2009) examined four interviewer retrieval probes (timing, parallel, sequential and duration) separately and in 2 of the 4 retrieval behaviors (parallel and duration) they found: $\alpha_2 = \alpha_B > \alpha_1 = \alpha_A$

Also, 3 of the 4 retrieval behaviors (parallel, duration and timing) they found: $\beta_2 = \beta_B > \beta_1 = \beta_A = 0$

[48] The examination of these findings is not the scope of this study. These are provided because hypotheses regarding the use of retrieval strategies aim to build on these findings.

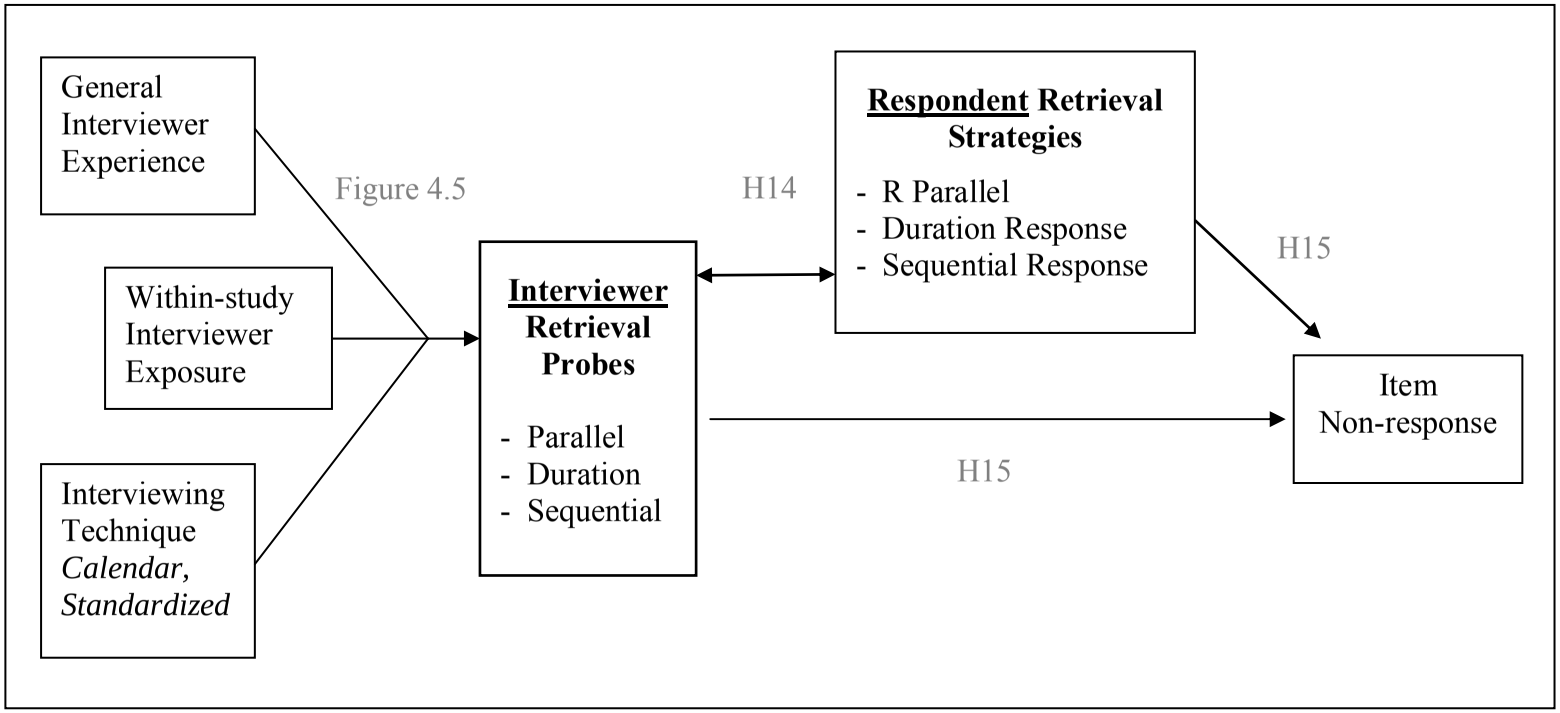
Similar to earlier findings, they found that calendar interviewers were using retrieval probes more often than standardized interviewers, regardless of their experience levels **($\alpha_2 = \alpha_B > \alpha_1 = \alpha_A$ in Figure 4.5).** Bilgen, Belli, and Olson (2009) found that general interviewing experience does not come into play in the majority of the interviewer retrieval probes during the survey fielding period in both calendar and standardized interviews. In addition, they found that in calendar interviews the use of interviewer retrieval probes significantly increased as the interviewers gained exposure during the fielding period, whereas in standardized interviews the use of interviewer retrieval probes does not significantly increase as the interviewers gained exposure during the study fielding period **($\beta_2 = \beta_B > \beta_1 = \beta_A = 0$ in Figure 4.5)**.

Moreover, as mentioned earlier in Chapter I, no studies explore how interviewer experience and exposure affect the use of respondent retrieval strategies. This dissertation intends to fill this research gap. As discussed earlier in Chapter III, each set of behaviors will be combined to produce two measures (1- Interviewer retrieval scale, 2- Respondent retrieval scale). The expected direction in Figure 4.6 indicates the hypotheses regarding how I would expect interviewer retrieval probes and respondent retrieval strategies to interact with interviewer experience and exposure and how this interaction translates into the change in item non-response due to interviewer experience and exposure. Specifically, my hypotheses are:

H14: There is a significant positive correlation between interviewer retrieval probes and respondent retrieval strategies.

H15: The increase in interviewer retrieval probes and respondent use of retrieval strategies both decrease *item non-response* (see Figure 4.6).

Figure 4.6 Expected Role of *Interviewer and Respondent Retrieval Strategies* in the Item Non-response Models for Calendar and Standardized Interviews



In addition, it has been hypothesized that retrieval behaviors aid respondents to recall events more productively (Belli, 1998; Belli, Lee, Stafford, & Chou, 2004; Bilgen & Belli, 2010a). Both interviewer retrieval probes and respondent retrieval strategies have been found to increase the accuracy of retrospective reports in calendar interviews and to be inconsequential in standardized interviews (Belli, et al., 2004). However, what is yet to be explored is how the interviewer and respondent retrieval strategies relate to change in item non-response due to interviewer experience and exposure in calendar and standardized interviews. This dissertation is also intended to fill this research gap.

Hence, I combine all the pieces regarding earlier findings and earlier mentioned hypotheses together (see Figure 4.6):

H16: Overall, the rate of ***item non-response*** decrease due to ***interviewer and respondent retrieval behaviors (regardless of interviewer exposure)***:

- Will not differ among experienced and inexperienced interviewers (in both calendar and standardized interviews).
- Will be higher among calendar interviewers than standardized interviewers (in both interviewer experience levels).

H17: The rate of ***item non-response*** decrease due to ***retrieval behaviors*** in early studies versus in later studies (low vs. high within-study interviewer exposure):

- Will not differ among experienced and inexperienced interviewers with higher exposure levels (in both calendar and standardized interviews).
- Will be higher among calendar interviewers with higher exposure than the standardized interviewers with higher exposure (regardless of the experience levels).

CHAPTER V: DATA ANALYSES PLAN AND INITIAL FINDINGS

1. Data Analyses Plan: Initial and Main Analyses

As illustrated in Chapter III, each of the nine outcome measures varies either across two-levels—as respondents are nested within interviewers—or in three-levels—as turns are nested within respondents and within interviewers (see Table 3.11). Simple regression analyses do not take into account the hierarchical structure of the data, which would yield underestimation of standard errors and increase in Type I errors (i.e., false positive/ incorrect rejection of null hypotheses) in clustered data. Thus, by using multilevel models I aimed to more accurately measure standard errors and appropriately account for dependency among respondents who are interviewed by the same interviewer and among turns that are coming from the same interview (Hox, 1994; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). I used SAS 9.2 in the multilevel data analyses illustrated in the next sections because it is a flexible and powerful program that is suitable for exploring generalized linear mixed models (Hedeker, 2005; Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006).

The multilevel analyses consist of a three step process in this dissertation. At the first step, I examined the intraclass correlations (ρ_{int}) to assess whether there is a significant interviewer variation effect on the two-level measures from the random intercept only models (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Moreover, for the measures that vary in three-levels, I computed three intraclass correlations from the random intercept only logistic models in order to take into account the three levels of

nesting. Hence, I examined the variation related to turns (level 1), respondents (level 2), and interviewers (level 3). Step 2 analyses examined the relationship between interviewer experience, exposure, and interviewer and respondent behaviors that occur in calendar and standardized interviews, and contained eight interviewer and respondent behavior measures as outcome variables. Step 2 models also took into account the available interviewer and respondent characteristics (age, sex, and race) as control variables. Lastly, Step 3 analyses (illustrated in Chapter VI) contained the key findings as they answer the "so what" question in this dissertation via including the item non-response data quality measure in the models. Hence, Step 3 analyses first examined the relationship between interviewer experience, exposure, and item non-response in calendar and standardized interviews; then, assessed whether this relationship was mediated by several different interviewer and respondent behaviors that occur in calendar and standardized interviews. These models also took into account the available interviewer and respondent characteristics (age, sex, and race). Hence in these key models, item non-response (respondent's unresolved "don't know" behaviors) is the outcome measure.

2. Variation Across Interviewers: Intraclass Correlations Coefficients

For the two-level measures that are utilized as the outcome measures in step 2 analyses, I assessed the baseline random intercept only models to examine whether there is a significant and meaningful correlation among interviewers via intraclass correlation coefficient (also known as ρ_{int} or ICC) (Groves & Magilavy, 1986; O'Muircheartaigh & Campanelli 1998; Kish, 1962; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

Table 5.1 Intraclass Correlation Coefficients for the Interviewer and Respondent Behaviors that are used as Outcome Measures in the Initial and Main Analyses

3-Level Outcome	**ICC**
Item Non-Response	
ICC1 for turns within respondents (and interviewers) [(BI Variance) + (BR Variance) / (Total Variance)] *(Dependency between turns among the same survey interview)*	0.345
ICC2 for turns within interviewers [(BI Variance) / (Total Variance)] *(Dependency among turns from the interviews collected by the same interviewer)*	0.045
ICC3 for respondents within interviewers [(BI Variance) / (BI Variance) + (BR Variance)] *(The dependency among R's who are interviewed by the same I'wer)*	0.131
2-Level Outcomes	**ICC**
1- Interviewer's Deviation from Conventional Ideals	
Interviewing Pace	0.537
Interviewer Verbal Behavior Deviation Scale	0.214
Interviewer Directive Behaviors	0.613
2- Interviewer and Respondent Interpersonal Communication	
Interviewer Rapport Scale	0.262
Respondent Cooperation Scale	0.124
Respondent Difficulty and Non-Coop Scale	0.053
3- Interviewer and Respondent Retrieval Behaviors	
Interviewer Retrieval Scale	0.337
Respondent Retrieval Scale	0.077

In these analyses, the intraclass correlations indicated the percentage of between interviewer variance from the two-level measures:

$$\text{ICC}\ (\rho_{int}) = \frac{\text{Between Interviewer Variance}}{\text{Between Interviewer Variance} + \text{Within Interviewer Variance}}$$

and was calculated via random intercept only unconditional means models in SAS PROC MIXED using the Maximum Likelihood (ML) estimation method (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). As explained in Chapter III, due to the low occurrence at the turn level I examined all the hypothesized mediation variables (i.e., interviewer deviation from conventional ideals, interviewer and respondent interpersonal dynamics, and interviewer and respondent retrieval behaviors measures) at the respondent level. Hence, Table 5.1 illustrates the ICC (ρ_{int}) values for these two-level outcome measures.

According to the ICC (ρ_{int}) values for the **interviewer deviation from conventional ideals** measures, 53.7% of the unexplained variance in the interviewing pace measure ($\chi^2 = 160.46$, df =1, $p < 0.001$) resulted from the between-interviewer variation. This indicates that the interviewing pace is more correlated (i.e., similar) among the interviews collected by the same interviewer than among the interviews collected by different interviewers.

Consistent with the literature, the dependency in interviewing pace among the interviews collected by the same interviewer may suggest that the interviewers play a role in setting the pace of the interviews (Cannell et al., 1981; Kahn and Cannell, 1957; Olson & Peytchev, 2007). Moreover, 21.4% of the unexplained variance in the verbal behavior (VB) deviation scale measure ($\chi^2 = 34.04$, df =1, $p < 0.001$) and 61.3% of the unexplained variance in the interviewer directive behavior measure ($\chi^2 = 211.75$, df =1,

$p < 0.001$) resulted from the between-interviewer variation. These ICC results for the interviewer deviation from conventional ideals measures illustrate that these interviewer behaviors are significantly less variant (more similar) among the interviews collected by the same interviewers than among the interviews collected by different interviewers.

The ICC (ρ_{int}) values for the **interviewer and respondent interpersonal communication** behaviors indicate that 26.2% of the unexplained variance in the interviewer rapport scale ($\chi^2 = 51.20$, df =1, $p < 0.001$), 12.4% of the unexplained variance in the respondent cooperation scale ($\chi^2 = 13.08$, df =1, $p < 0.001$), and 5.3% of the unexplained variance in the respondent difficulty and non-cooperation scale ($\chi^2 = 3.68$, df =1, $p = 0.055$) resulted from the between-interviewer variation. The ICC (ρ_{int}) values for the **interviewer and respondent retrieval behaviors** illustrate that 33.7% of the unexplained variance in the interviewer retrieval scale ($\chi^2 = 74.25$, df =1, $p < 0.001$) and 7.7% of the unexplained variance in the respondent retrieval scale ($\chi^2 = 6.83$, df =1, $p = 0.009$) resulted from the between-interviewer variation. These ICC (ρ_{int}) values indicate that interviewer rapport and retrieval behaviors or probes and respondent cooperation and retrieval behaviors are significantly alike among interviews collected by the same interviewer than among interviews collected by different interviewers.

The interviewer behaviors that occur during the interview are an attribute of interviewers; hence, respondent behaviors are less prone to interviewer variation than interviewer behaviors. Therefore, not surprisingly, ICCs that measure the between-interviewer variation in interviewer behaviors are larger than the ICCs that measure the between-interviewer variation in respondent behaviors for the two-level outcome

measures (see Table 5.1). Lastly, it needs to be noted that the ρ_{int} values for the interviewer behaviors (and the interviewing pace measure) are larger than the generally reported intraclass correlations in the literature (Groves & Magilavy, 1986; O'Muircheartaigh & Campanelli, 1998; Olson & Bilgen, 2011). This may indicate that respondent-level respondent and interviewer behaviors are more prone to interviewer variation than responses to specific items, as the reported ICCs in the literature indicate the interviewer variation for responses to specific items, whereas in these analyses the reported ICCs indicate the variation for specific interviewer and respondent behaviors.

3. Variation Between Interviewers and Respondents: Intra-class Correlation Coefficients for the Three-Level Outcome Measure

For the item non-response measure, which vary in 3-levels, I examined three ICCs from the three-level logistic random intercept unconditional means models via SAS PROC GLIMMIX (using the Laplace estimation method) to specify the three levels of nesting and account for the unexplained variation related to turns, respondents, and interviewers (Hedeker, 2005; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

The first ICC, illustrated in Table 5.1 expresses the correlation (or dependency) of turns that are obtained from the same interview (i.e., from the same interviewer and the same respondent). This ICC is the percentage of between respondent (BR) + between interviewer (BI) variance in the item non-response measure (Hedeker, 2005):

$$ICC1 = \frac{(\text{BI Variance}) + (\text{BR Variance})}{\text{Total Variance}} = \frac{(\text{BI Variance}) + (\text{BR Variance})}{[(\text{BI Variance}) + (\text{BR Variance}) + (\pi^2/3)]}$$

According to ICC1 illustrated in Table 5.1, 34.5% of the unexplained variation in the item nonresponse measure at the turn-level resulted from variation among interviewers and respondents. This indicates that there is a correlation between the respondent's unresolved "don't know" behaviors among the turns that are collected from the same respondent. The second ICC illustrated in Table 5.1, expresses the similarity in item nonresponse among the turns in the interviews that are collected via same interviewer. This ICC is the percentage of between interviewer (BI) variance in the item non-response measure (Hedeker, 2005):

$$ICC2 = \frac{(\text{BI Variance})}{\text{Total Variance}} = \frac{(\text{BI Variance})}{[(\text{BI Variance}) + (\text{BR Variance}) + (\pi^2/3)]}$$

According to ICC2, 4.5% of the unexplained variation in the item nonresponse measure at the turn-level resulted from variation among interviews. This indicates that there is some dependency in respondent's unresolved "don't know" behaviors among the turns and interviews that are collected via the same interviewer.

Lastly, the last ICC illustrated in Table 5.1 expresses the dependency in item nonresponse among respondents who are interviewed by the same interviewer.

$$ICC3 = \frac{(\text{BI Variance})}{[(\text{BI Variance}) + (\text{BR Variance})]}$$

According to ICC3, 13.1% of the unexplained variation in the item nonresponse measure at the respondent-level resulted from variation among interviewers. That means there is some similarity in item non-response measure between the respondents who are interviewed by the same interviewer. However, it also needs to be noted even though this is an experimental design, there is a lack of interpenetration as the experiment was not geared towards exploring interviewer effects. The interpenetrated design method was

developed by Mahalanobis (1946) and this design assigns households or respondents at random to interviewers (O'Muircheartaigh & Campanelli, 1999) to measure interviewer variance and separate the effects of the interviewer from the effects of other sources (such as region, neighborhood, and area). Therefore, the dependency among turns and respondents discussed above may not be solely due to interviewers. This is one of the limitations of this dissertation, which is also discussed in detail in the conclusion and discussion chapter (see Chapter VII).

4. The Relationship between Interviewer Experience, Exposure, and Interviewer and Respondent Behaviors that occur during the Collection of Calendar and Standardized Interviews

As a second step, the analyses continue with ***eight intermediary models*** in which eight interviewer and respondent behaviors (1- Interviewing Pace, 2- Interviewer VB Deviation Scale, 3- Interviewer Directive Behavior, 4- Interviewer Rapport Scale, 5- Respondent Cooperation Scale, 6- Respondent Difficulty and Non-Cooperation Scale, 7- Interviewer Retrieval Scale, and 8- Respondent Retrieval Scale) are utilized as the outcome measures.

All of the interviewer and respondent behaviors (including the interviewing pace – number of words per minute) are ***count measures,*** which vary among respondents and interviewers. Due to the overdispersion (i.e., variances >> means) in all of the eight outcome measures, Poisson distribution is not an appropriate assumption (which assumes equal variances and means). Hence, I used negative binomial multi-level models via SAS PROC GLIMMIX procedure. The specification of the negative binomial distribution (via including *dist=negbin* in the MODEL statement) in PROC GLIMMIX procedure transforms the

dependent variables into a natural logarithm form (i.e., default link function=log) in order to be able to use it as a continuous measure in the models.

In addition, as the outcome measures vary among respondents and interviewers, two-level[49] random intercept only models are used in the second step of this study. The random intercept models allowed the interviewer effects (i.e., means) to be random and the linear effect of the predictors that describe the individual differences in change (slopes) among interviewers to be fixed. In other words, random intercept models took into account the dependency among respondents within interviewers by allowing interviewer means (i.e., intercepts) to be random, and fixed the variation (i.e. allow systematic variation) in the slopes that describe the variation in the effect of the predictors among interviewers.

In the initial models, I explored two types of interviewer experience (i.e., general experience and within-study exposure) and interviewing technique (i.e., condition) as the three main independent variables. The general interviewer experience (IExpe) is a dichotomized[50] *(0-Less than 1 year, 1- 1 year or more)* interviewer-level measure, which varies among interviewers. Interviewing technique (IT) is also a dichotomous *(0-calendar, 1-standardized)* interviewer-level measure. Hence, both of these variables were dummy coded to simplify the interpretation of the estimates and are included at the interviewer level (BIIExpe and BIIT) in the models below. The within-study interviewer exposure variable is a respondent-level measure that varies between interviewers and between respondents. In order to capture the log-linear nature of the relationship between

[49] Seven out of the eight intermediary outcome measures (interviewer and respondent behavior scales) occur at the turn level (see Table 3.11). However, I analyze these at the respondent level throughout the dissertation due to the low occurrences and variation at the turn level (see Chapter III).

[50] The general interviewer experience is dichotomized as the change in interviewer behaviors due to experience is expected to be non-linear (see Chapter II for further discussion). Hence, the change in their behaviors is expected to be steeper before they hit the 1 year mark, then this change is expected to decrease when they gain experience for 1 year or more (Olson & Bilgen, 2011; Singer, Frankel & Glassman, 1983; Tu & Liao, 2007)

the interviewer exposure and interviewer and respondent behaviors, the exposure measure was transformed into a natural logarithm (LN)[51] format.

The interviewer exposure variable measures the "interview order" (Hughes, et al., 2002; Olson & Peytchev, 2007), in which "1" is designated to the first interview and each 1 unit increase indicates an increase in exposure. The natural logarithm of the exposure measure starts from 0 and each unit increase indicates an increase in the LN(interviewer exposure). Hence, LN (interviewer exposure) is a "conceptually grand-mean centered" measure (Hoffman, 2011 personal communication) and has two-components in the model due to the variation in interviewer and respondent-levels: The respondent–level component [WILN (IExpo) =LN ($IExpo_{ri}$)] compares the effect of the exposure measure relative to other interviews. The between interviewer-level component indicates the additional effect of between-i'wer exposure increase [BILN (IExpo) =$LN(\overline{IExpo}_{.i})$]. The between-interviewer exposure is included as a control variable in the models. In this study, each interviewer conducted different numbers of interviews during the survey fielding period. In other words, there is a variation in interviewers' overall exposure levels. Hence, the unexplained variation in outcome measures due to the between-interviewer exposure variation is taken into account via including the between-interviewer exposure component in the models.

Lastly, interviewer and respondent characteristics (age, sex, and race) are included as control variables in order to take into consideration the potential interviewer and respondent confounding effects. Both interviewer and respondent sex and race

[51] In six out of eight multi-level models illustrated below, the models which include natural logarithm of *Interviewer Exposure* measure fit better than the models which include *Interviewer Exposure* as a linear predictor. This is also consistent with earlier findings (Olson & Peytchev, 2007). Therefore for the purpose of consistency, the natural logarithm of interviewer exposure is employed in all of the further analyses (see Chapters I and II for further discussion).

measures are dichotomous dummy coded measures. Interviewer age is a continuous measure and grand mean centered. As interviewer characteristics only vary in interviewer-level, they are only included at the interviewer level in the models. As respondent sex and race vary both among interviewers and respondents, they are "conceptually grand-mean centered" (Hoffman, 2011 personal communication) and have two-components in the model: The respondent–level component compares the effect of respondent sex and race on interviewer and respondent behaviors (WIRsex =$Rsex_{ri}$ and WIRrace=$Rrace_{ri}$) and the interviewer-level component indicates the incremental between-interviewer effect of respondent sex and race (BIRsex=$\overline{Rsex}_{.i}$ and BIRrace=$\overline{Rrace}_{.i}$).

In this study, there are significantly higher numbers of female and European-American interviewers (see Chapter III). These higher numbers lead to the higher likelihood of demographic match of female interviewers with female respondents and of European-American interviewers with European-American respondents. Therefore, the between-interviewer (BI) effect of respondent race and sex is included in subsequent models to control for the uneven likelihood of demographic match. To explore the overall between-interviewer effect of respondent sex and race (e.g., to explore whether the interviewers with more female respondents differ in their behaviors than interviewers with more male respondents), the within- and contextual- components (obtained via group-mean centering) are combined to create a between-interviewer effect using the ESTIMATE statement in SAS 9.2.

Moreover, the respondent age is a continuous measure and has two-components in the model: The respondent–level group-mean centered component compares the effect of

respondent age on interviewer and respondent behaviors (WIRage = $Rage_{ri} - \overline{Rage}_{.i}$). Moreover, as respondent age ranges between 46 and 98, there is a higher likelihood of demographic match between older interviewers and older respondents. As many of the experienced interviewers also are older than average, the between interviewer effect of both interviewer and respondent age are included in subsequent analyses. The interviewer-level grand-mean centered component indicates the overall between-interviewer effect of respondent age on interviewer and respondent behaviors (BIRage=$\overline{Rage}_{.i} - \overline{Rage}_{..}$). To summarize, all the models (illustrated in Tables 5.2 through 6.7) include the group-mean centered within-interviewer components and the grand-mean centered between-interviewer components of the measures due to the lack of randomization and imperfections in the experimental design in this study (see Chapter III for further information).

The initial models include the three-way interaction effects of interviewer experience, interviewer exposure, and interviewing technique. However, in the further tables, I only report the final models that are established via backward elimination method. To conclude, overall the **initial models** for each of the eight interviewer and respondent outcome measures are:

Respondent-level:

$$LN(Y_{ri}) = \beta_{0i} + \beta_{1i}\, WILN(IExpo_{ri}) + \beta_{2i}\, WIRage_{ri} + \beta_{3i}\, WIRsex_{ri} + \beta_{4i}\, WIRrace_{ri} + U_{ri}$$

Interviewer-level:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}\, BIIExpe_i + \gamma_{02}\, BIIT_i + \gamma_{03}\, BIIExpe_i * BIIT_i + \gamma_{04}\, BILN(IExpo_{.i})$$
$$+ \gamma_{05}\, BIRage_{.i} + \gamma_{06}\, BIRsex_{.i} + \gamma_{07}\, BIRrace_{.i} + V_{0i}$$

$\beta_{1i} = \gamma_{10} + \gamma_{11}$ BIIExpe$_i$ + γ_{12} BIIT$_i$ + γ_{13} BIIExpe$_i$*BIIT$_i$

$\beta_{2i} = \gamma_{20}$

$\beta_{3i} = \gamma_{30}$

$\beta_{4i} = \gamma_{40}$

This dissertation is organized around three sets of interviewer and respondent behaviors: 1- Interviewers' deviation from conventional ideals; 2-Interviewer and respondent interpersonal dynamics; and 3-interviewer and respondent retrieval strategies. Subsequent sections in this chapter illustrate the examination of the effect of interviewer experience and exposure on each of the three sets of interviewer and respondent behaviors that occur during calendar and standardized interviews.

DEVIATION FROM CONVENTIONAL IDEALS

The models examining the relationship between **interviewer experience[52], exposure, and deviation from conventional ideals in calendar and standardized interviews** evaluate Hypotheses 1, 2 and 3 in Chapter IV. Deviation from conventional ideals has three components: *Interviewing pace*, *interviewer directive behaviors,* and *interviewer verbal behavior deviation behaviors*[53]. Table 5.2 illustrates the findings from the **final models** regarding the relationship between interviewer experience, exposure, and each of the deviation components in different interviewing techniques. On the one hand, inconsistent with the hypotheses, interviewer experience does not significantly interact

[52] This measure refers to the experience with standardized interviews gained during an interview's lifetime period regardless of their interviewing condition (calendar or standardized interviewing).
[53] Interviewer verbal behavior deviation behaviors include: *Failure to probe a question or a part of the question, failure to probe after a "don't know" response, and improper positive feedback* (see Chapter III for further details on construction of the deviation scale).

Table 5.2

Interviewer's Deviation from Conventional Ideals[a]

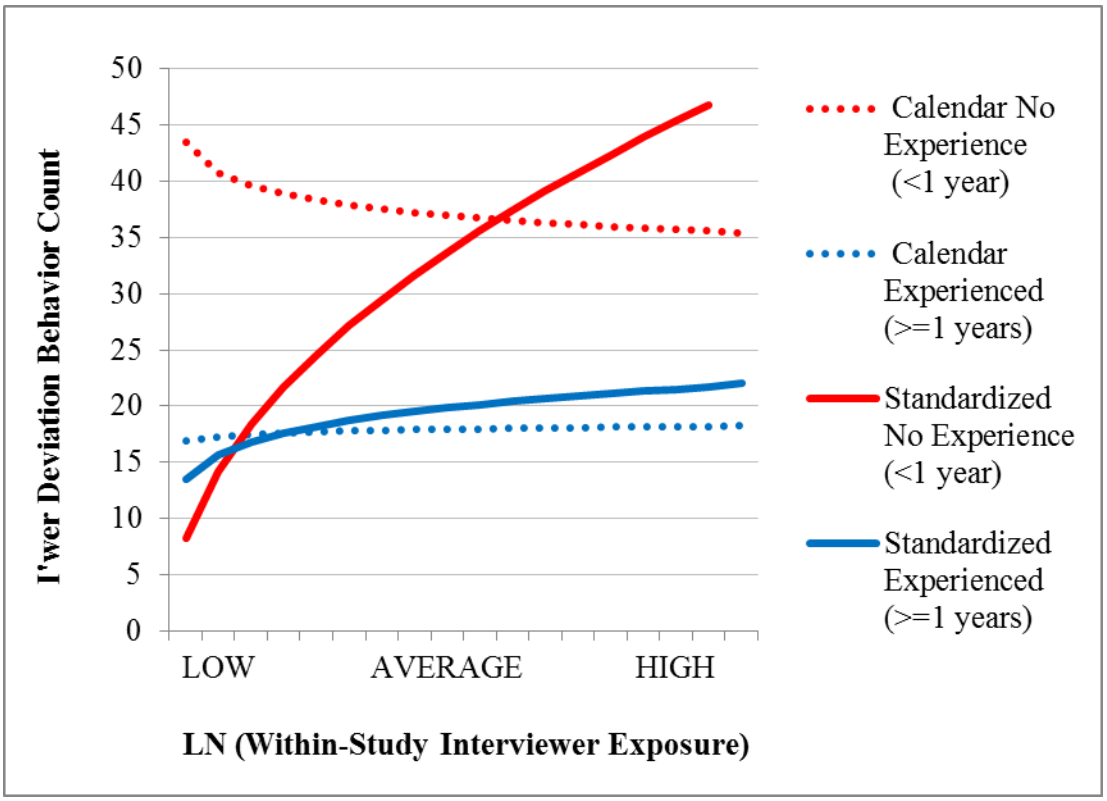
	Interviewing Pace		I'wer Directive Behaviors		I'wer VB Deviation	
Parameters	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects						
Intercept	5.053***	0.156	3.592***	0.415	3.771***	0.434
WI LN(I Expo)	0.018	0.011	0.030	0.044	-0.057	0.110
BI LN(I Expo)	-0.129*	0.058	-0.013	0.154	-0.271	0.181
BI IT = *Standardized*	0.177***	0.046	-1.491***	0.122	-1.660**	0.521
BI IExpe =*1 year or +*	-0.260***	0.073	-0.729***	0.200	-0.944**	0.332
BIIT*WILN(I Expo)					0.553**	0.204
BIIT*BIIExpe					1.439*	0.589
BIIExpe*WILN(I Expo)					0.078	0.137
3-way interaction					-0.439+	0.235
Control Fixed Effects						
Respondent-level						
WI R Age *(grp-mean)*	0.000	0.001	0.011**	0.003	0.011**	0.004
WI R Sex = *Women*	0.032+	0.019	0.059	0.078	0.099	0.087
WI R Race = *A-A*	-0.023	0.025	-0.044	0.101	0.026	0.113
Interviewer-level						
BI R Age *(0= 62)*	-0.012	0.008	-0.015	0.022	-0.063***	0.018
BI R Sex = *% Women*	-0.017	0.133	0.128	0.372	-0.352	0.308
BI R Race = *% A-A*	0.504+	0.290	2.469**	0.828	1.354+	0.709
BI I Age *(0= 49)*	-0.005*	0.002	-0.002	0.007	0.002	0.006
BI I Sex = *Women*	-0.057	0.072	0.029	0.194	0.033	0.161
BI I Race = *A-A*	-0.295***	0.086	-0.852**	0.268	-1.005***	0.204
Variance Components						
Residual Variance	0.013	0.002	0.277	0.034	0.356	0.037
Intercept Variance	0.009	0.003	0.048	0.028	0.014	0.016
Model Fit						
-2Log Likelihood	2447.15		2090.81		2013.33	
AIC	2479.15		2122.81		2053.33	
BIC	2498.65		2142.31		2077.71	
N (Sample Size)	285		292		292	

[a] ***Within-interviewer (WI) components:*** Group-mean centered *(grp-mean)*,
Between-interviewer (BI) components: Grand-mean centered

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001; -- Not Applicable;

with either interviewer exposure or interviewing technique in the interviewing pace and interviewer directive behavior models. On the other hand, consistent with the hypotheses, in the interviewer verbal behavior deviation models, interviewer experience interacts with interviewer exposure and interviewing technique (see Table 5.2). According to the interviewing pace and interviewer directive behavior models in Table 5.2, in both calendar and standardized interviews regardless of their exposure levels, inexperienced interviewers—who have less than 1 year of experience—are overall significantly faster (i.e., deliver higher numbers of words per minute) and use significantly higher numbers of directive behaviors than experienced interviewers—who have 1 year or more experience.

Figure 5.1 The Relationship between Interviewer Experience, Exposure, and Interviewers' VB Deviation Behaviors in Calendar and Standardized Interviews



Consistent with the hypotheses illustrated in Figure 4.1, in standardized interviews at the beginning of the study, inexperienced interviewers use slightly lower numbers of verbal behavior deviation behaviors (such as *failure to probe a question or a part of the question, fail to probe after a "don't know" response, and provide improper positive feedback*) than experienced interviewers. However, inexperienced interviewers' use of VB deviation behaviors drastically increases at the later interviews and their VB deviation behaviors increase more rapidly than experienced interviewers. Therefore, they use significantly higher numbers of deviation behaviors than experienced interviewers at the middle and end of the study (see Figure 5.1).

As the general interviewer experience is measured as interviewers' experience with standardized interviews throughout their lifetime, it is not surprising that, overall, experienced interviewers deviated less from conventional standardized interviewing ideals than inexperienced interviewers. These interviewers may be simply following what they have been told to do throughout their careers (such as following a slower pace, avoiding directive probing, and reading the questions as scripted). However, it also needs to be taken into account that even though the "deviation from conventional ideal" behaviors are labeled as "undesirable behaviors" in the standardized interviewing literature, these "so-called undesirable" behaviors may not be detrimental to data quality. More findings on this are illustrated in the following chapter.

Moreover, in calendar interviews inexperienced interviewers use significantly higher numbers of deviation behaviors than experienced interviewers in both early and later interviews during the study fielding period (see Figure 5.1). This is inconsistent with the expected findings (see Figure 4.1). Also, surprisingly inexperienced interviewers' use

of verbal behavior deviation behaviors decreases over the survey fielding period in calendar interviews. Calendar inexperienced interviewers may be using high deviation behaviors in earlier interviews during the fielding period as they are not familiar with this relatively complex interviewing technique. Also, as calendar interviewing encourages interviewers to provide higher number of retrieval behaviors (such as parallel, duration, and sequential probing), this may lead to trade-offs between "unconventional calendar" and "conventional standardized" behaviors (Bilgen & Belli, 2010; Bilgen, Belli & Olson, 2009) in the earlier interviews. Therefore, while inexperienced interviewers are trying to use higher retrieval behaviors (provide unconventional calendar techniques), they may be deviating from the training and do not use conventional standardized behaviors. However, as the experienced interviewers are trained via standardized interviews, they are more reluctant to use the deviation behaviors (such as *failure to probe a question or a part of the question, fail to probe after a "don't know" response, and provide improper positive feedback),* which are considered "not desirable" in standardized interviews. Thus, calendar interviewers, who are experienced with standardized interviews, may be simply not able to unlearn what they have learned throughout their careers.

Consistent with the hypotheses in Chapter IV, both calendar and standardized inexperienced interviewers modify their behaviors in later interviews during the study fielding period. Moreover, the change in inexperienced interviewers' deviation behaviors was steeper than the change in experienced interviewers' deviation behaviors regardless of the interviewing technique. This is also consistent with the findings in the literature that indicate that inexperienced interviewers are more inclined to adapt their use of behaviors in later interviews in comparison to earlier interviews during the study fielding

period. However, experienced interviewers bring a "package of habits" to the interviewing situation that are used throughout the study; hence, they may not feel the need to modify or adapt their existent behaviors throughout the study fielding period (Hughes et al., 2002; Olson & Peytchev, 2007; Pickery & Loosveldt, 2001).

Lastly, according to Table 5.2 interviewer race consistently associates with all of the deviation from conventional ideals measures. According to these findings, African-American interviewers are overall significantly slower (i.e., deliver fewer words per minute) and use significantly fewer directive and deviation behaviors than the European-American interviewers. Moreover, interviewers who are older than average (interviewer age average=49) are significantly slower (i.e., deliver fewer words per minute). Also, interviewers provide significantly higher numbers of directive and deviation behaviors when they are interviewing respondents who are older than average (respondent age average=62).

According to the interviewer-level respondent characteristics, interviewers who interview more African-American respondents provide higher numbers of directive behaviors and interviewers who interview higher numbers of older respondents provide lower verbal behavior deviation behaviors. However, these findings may not fully explain the role of interviewer and respondent characteristics on the use of interviewer and respondent behaviors. There may be several other meaningful interactive effects that could be explored (e.g., the effect of interviewer and respondent characteristic match on interviewer and respondent behaviors). However, it is not the scope of this paper to examine the role of interviewer and respondent characteristics in calendar and standardized interviews. These measures are included for the purpose of controlling the

confounding effects of interviewer and respondent characteristics due to the lack of interpenetration (see Chapter III for more discussion).

INTERVIEWER AND RESPONDENT INTERPERSONAL DYNAMICS

The models regarding the relationship between **interviewer experience, exposure, and survey actors' interpersonal communication in calendar and standardized interviews** respond to Hypotheses 7, 8 and 9 provided in Chapter IV. The interviewer and respondent interpersonal behaviors have three components: ***Interviewer rapport behaviors***[54], ***respondent cooperative behaviors***[55] and ***respondent expression of difficulty and non-cooperation behaviors***[56]. Table 5.3 illustrates the findings from the **final models** regarding the relationship between interviewer experience, exposure, and each of the interpersonal behaviors in different interviewing techniques.

According to Table 5.3, neither the three main predictors (interviewer experience, exposure and interviewing technique), nor the two- or three-way interactions significantly relates to interviewer rapport behaviors. However, interviewer experience and interviewing technique have an interactive (and significant) effect on both respondent cooperation behaviors and respondent difficulty and non-cooperative behaviors. Moreover, the relationship between interviewer experience and respondent difficulty and non-cooperation varies among interviewers with different exposure levels. However, this relationship is marginally significant (see Table 5.3).

---

[54] **Interviewer Rapport behaviors include**: *Interviewer empathy, agreement, joking or sarcasm, direct apology, and laughter to a respondent joke or comment (*see Chapters III and IV for further discussion).
[55] **Respondent Cooperative behaviors include**: *Respondent's jokes or sarcastic comments, spontaneous clarifications, and corrections.*
[56] **Respondent Expressions of difficulty and non-cooperative behaviors include**: *Respondent expressions of difficulty answering a question, seeking clarification, guesses, estimates, and negative comments about the interview/task/questionnaire.*

Table 5.3

Interviewer and Respondent Interpersonal Communication Dynamics[a]

	I'wer Rapport		R Cooperation		R Difficulty and Non-Cooperation	
Parameters	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects						
Intercept	2.750**	0.715	4.511***	0.283	3.502***	0.296
WI LN(I Expo)	-0.041	0.052	0.057	0.036	0.155*	0.075
BI LN(I Expo)	-0.339	0.262	-0.144	0.112	0.042	0.124
BI IT = *Standardized*	-0.235	0.212	0.211	0.174	0.356**	0.133
BI IExpe =*1 year or +*	-0.418	0.343	-0.107	0.152	0.422+	0.215
BIIT*WILN(I Expo)						
BIIT*BIIExpe			-0.367+	0.205	-0.495**	0.160
BIIExpe*WILN(I Expo) 3-way interaction					-0.146+	0.085
Control Fixed Effects						
Respondent-level						
WI R Age *(0=grp)*	-0.003	0.004	0.013***	0.003	0.009**	0.003
WI R Sex = *Women*	0.242**	0.092	0.024	0.064	0.009	0.067
WI R Race = *A-A*	-0.334**	0.121	-0.078	0.083	0.024	0.086
WI I Rapport *(0=grp)*	--	--	0.034***	0.003	0.020***	0.003
Interviewer-level						
BI R Age *(0= 62)*	-0.001	0.037	-0.004	0.014	-0.008	0.012
BI R Sex = *% Women*	0.746	0.624	0.305	0.244	0.318	0.198
BI R Race = *% A-A*	1.192	1.374	0.566	0.542	0.400	0.446
BI I Rapport (0=12)	--	--	0.029***	0.008	0.019***	0.006
BI I Age *(0= 49)*	0.001	0.011	0.004	0.004	0.003	0.004
BI I Sex = *Women*	0.415	0.341	-0.127	0.137	-0.059	0.107
BI I Race = *A-A*	-1.188**	0.409	-0.242	0.168	-0.117	0.138
Variance Components						
Residual Variance	0.386	0.043	0.217	0.020	0.226	0.020
Intercept Variance	0.196	0.076	0.013	0.010	--[b]	--[b]
Model Fit						
-2Log Likelihood	1936.51		2822.60		2606.43	
AIC	1968.51		2860.60		2644.43	
BIC	1988.01		2883.76		2667.59	
N (Sample Size)	292		292		292	

[a] ***WI components:*** Group-mean centered (0=grp), ***BI components:*** Grand-mean centered

[b] The G-matrix (random intercept variance) is estimated to be zero; hence, could not be kept in the model.

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001; -- Not Applicable

Figure 5.2 The Relationship between Interviewer Experience, Interviewing Technique and R Cooperation Behaviors (regardless of interviewer exposure)

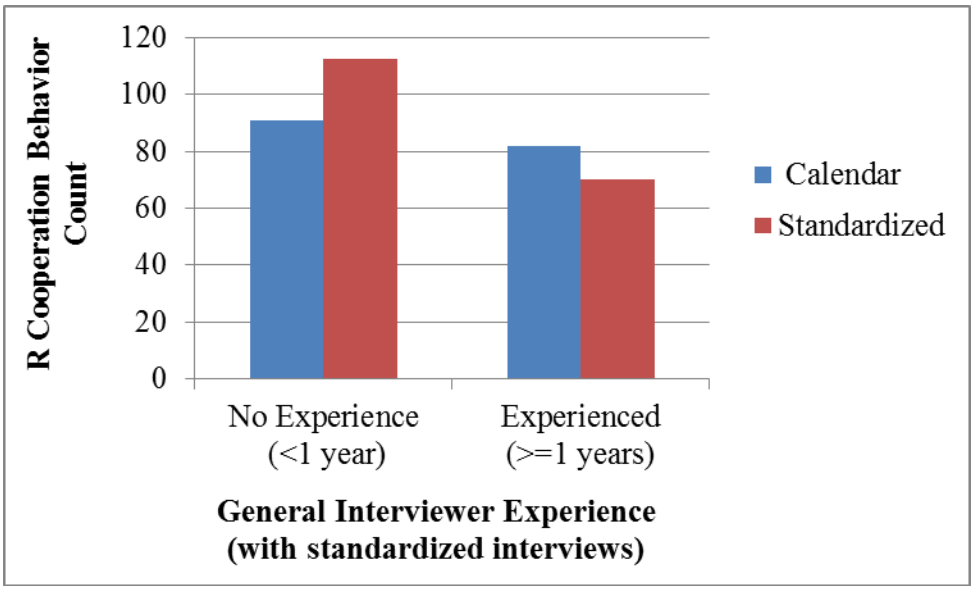
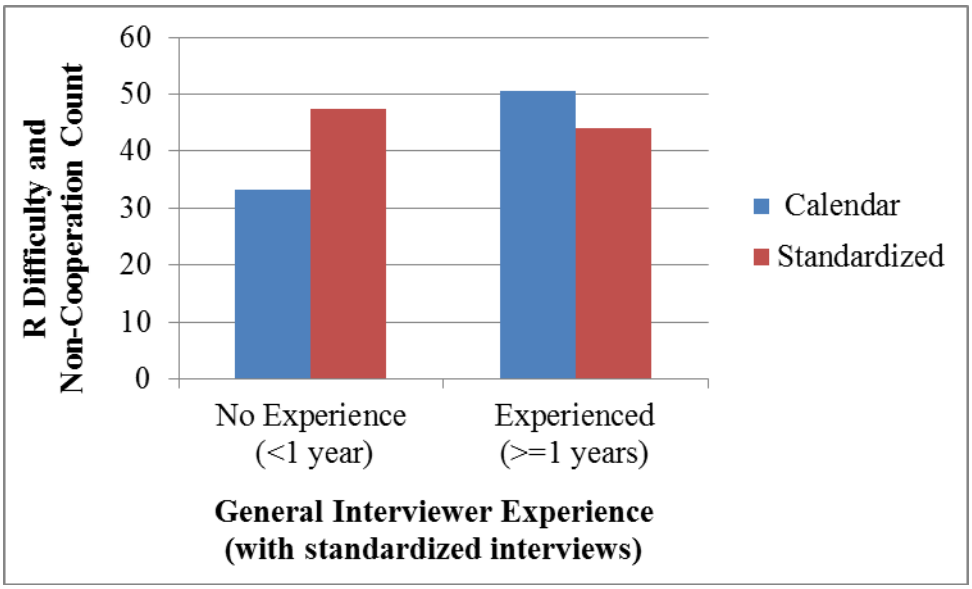


Figure 5.3 The Relationship between Interviewer Experience, Interviewing Technique and R Difficulty Behaviors (regardless of interviewer exposure)



According to the Figures 5.2 and 5.3, while experienced calendar interviewers obtain higher respondent cooperation, difficulty and non-cooperation behaviors than experienced standardized interviewers; inexperienced calendar interviewers obtain lower

respondent cooperation, difficulty and non-cooperation behaviors than inexperienced standardized interviewers. The respondent cooperation and difficulty/non-cooperation models may indicate that the interviewers who enable an open communicative interaction may be getting more feedback from respondents regarding their interviewing experience, regardless of the pleasantness/un-pleasantness, or difficultness of the interviewing situation. Hence, the "undesirable respondent behaviors" may be in fact desirable as the respondent difficulty and non-cooperation behaviors may take place as respondents feel more comfortable providing their opinions about the interview and expressing their frustration when they encounter a problem while answering questions. Moreover, another finding that may support this notion is that the significant positive relationship between interviewer rapport and respondent difficulty and non-cooperative behaviors. Consistently, there is a significant positive association between interviewer rapport and respondent cooperation when interviewer and respondent characteristics are taken into account in the models (see Table 5.3).

These findings may suggest that while experienced interviewers are more approachable in calendar interviews than in standardized interviews; inexperienced interviewers are more approachable in standardized interviews than in calendar interviews. Moreover, according to the Figures 5.2 and 5.3, inexperienced interviewers overall obtain higher respondent interpersonal communication behaviors (respondent rapport, difficulty and non-cooperative behaviors) than their experienced colleagues in standardized interviews. One possibility is that in standardized interviews, while inexperienced interviewers are more approachable and they enable an open communication, experienced interviewers put more emphasis on efficiency and are less

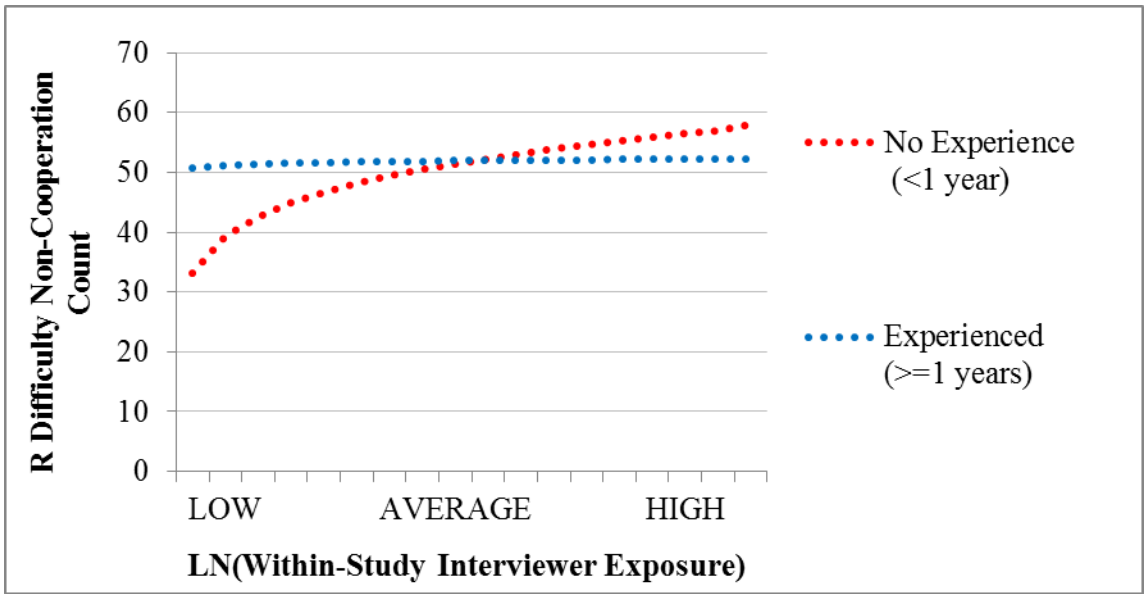
likely to allow respondent feedback and communication (Groves et al. 2004; Olson & Bilgen, 2011). Consistently, earlier findings in the literature also have shown that experienced interviewers deviate less from the standardized conventional ideals. Therefore, the experienced interviewers may be simply following what they have been instructed to do by the researchers throughout their careers. However, whether this is detrimental or beneficial to data quality is still an unanswered research question. This dissertation tackles whether this is detrimental or beneficial to item response probabilities in the next chapter.

The relationship between respondent communicative behaviors and interviewer experience is more complex and can be explained via different theories in calendar interviews. Inexperienced calendar interviewers obtain higher rapport behaviors, and lower difficulty and non-cooperation behaviors than experienced calendar interviewers. One explanation is that inexperienced calendar interviewers put more emphasis on respondent happiness and comfort during the interview (Belli, 2011 personal communication). Hence, the perceptual distinction between respondents' "good" (i.e., respondent cooperation) and "bad" (i.e., respondent difficulty and non-cooperation) behaviors may be more pronounced among inexperienced calendar interviewers than experienced calendar interviewers. Another explanation for these findings is that inexperienced interviewers are providing a more enjoyable and less difficult interviewing situation than experienced interviewers in calendar interviews and rewarded by the respondents due to their accomplishment. However, respondents' communication behaviors may not be a good proxy to determine respondents' feeling about the

interviewing experience, as not every respondent may be vocal about their interviewing experiences.

In addition, according to the respondent difficulty and non-cooperation models, inexperienced interviewers on average attain lower numbers of respondent difficulty and non-cooperative behaviors than experienced interviewers in earlier interviews. Consistent with the hypotheses illustrated in Chapter IV, the behavior gap between inexperienced and experienced interviewers diminishes as inexperienced interviewers gain exposure by the middle of the study fielding period. While inexperienced interviewers obtain higher difficulty and non-cooperation behaviors in later studies, experienced interviewers attain consistent amount of these behaviors throughout the study fielding period (see Figure 5.4 and Table 5.3).

Figure 5.4 The Relationship between Interviewer Experience, Exposure and R Difficulty and Non-Cooperative Behaviors (regardless of interviewing technique)



This is consistent with the notion that inexperienced interviewers modify their behaviors, and put more emphasis on efficiency rather than respondent happiness and

comfort at each interview they conduct throughout the study, while experienced interviewers are more consistent throughout the study, as their interviewing style has become more crystallized (Bilgen, Belli & Olson, 2009; Hughes et al., 2002; Olson & Peytchev, 2007; Pickery & Loosveldt, 2001).

Lastly, according to the interviewer-level respondent characteristics in Table 5.3, interviewer race significantly associates with interviewer rapport behaviors. Accordingly, African-American interviewers significantly provide less rapport behaviors than European-American interviewers. Moreover, interviewers use higher numbers of rapport behaviors when they are interviewing female or European-American respondents. As there are more female interviewers and more European-American respondents and interviewers in the study (see Chapter III), these results may be due to higher likelihood of demographic match between female or European-American interviewers and respondents. In addition, older respondents provide higher respondent cooperation, difficulty and non-cooperation behaviors. This may indicate that older respondents are more vocal about their interviewing experiences.

RETRIEVAL STRATEGIES AND PROBES

The relationship between interviewer experience, exposure, and interviewer retrieval probes[57] in calendar and standardized interviews has been already explored by Bilgen, Belli, and Olson (2009) who analyzed the same cases of the data set used in this

[57] Bilgen, Belli, & Olson (2009) explore interviewer retrieval behaviors (parallel, duration, and sequential) separately. However, the combined interviewer retrieval provide very similar findings when they are included in the models as one outcome variable; hence, I will only report Bilgen, Belli, & Olson (2009) findings in order to avoid repetition.

dissertation. Hence, first I will report their findings, and then explore interviewer experience, exposure, and respondent retrieval behaviors[58] in the current section.

According to the interviewer retrieval behavior findings in Bilgen, Belli, and Olson, (2009), calendar interviewers use more retrieval behaviors than standardized interviews. In calendar interviews both experienced and inexperienced interviewers use retrieval behaviors (duration, parallel and sequential behaviors) more in the later interviews over the study fielding period. However, in standardized interviews the use of duration and parallel retrieval behaviors does not change over the course of fielding period. In addition, standardized interviewers use sequential probes more than duration and parallel behaviors (Bilgen & Belli, 2010). Therefore, in standardized interviews there is more room for modification of sequential probing throughout the fielding period than duration and parallel probing. Hence, experienced standardized interviewers use higher sequential behaviors **in later interviews** during the fielding period as they learn to use these behaviors more spontaneously by following the examples in the standardized scripts. However, while inexperienced standardized interviewers use higher sequential behaviors than experienced interviewers **in early interviews** as they follow the training, they tend to get more careless and use less sequential behaviors throughout the fielding period (Bilgen, Belli & Olson, 2009).

Table 5.4 illustrates the findings from the **final models** regarding the relationship between interviewer experience, exposure, and respondent retrieval in different interviewing techniques. According to Table 5.4, both interviewing technique and interviewer experience have a significant effect on respondent retrieval behaviors.

[58] Respondent retrieval strategies include: *Respondent parallel, duration response, and sequential response behaviors (*see Chapters III and IV for further discussion).

Table 5.4

Respondent Retrieval Strategies[a]

Parameters	R Retrieval Behaviors (w/out I'wer Retrieval) Beta	SE	R Retrieval Behaviors (w/ I'wer Retrieval) Beta	SE
Main Fixed Effects				
Intercept	3.160***	0.353	2.853***	0.324
WI LN(I Expo)	0.079	0.057	0.032	0.052
BI LN(I Expo)	-0.064	0.129	-0.120	0.119
BI IT = *Standardized*	-0.260**	0.095	0.290	0.175
BI IExpe =*1 year or +*	-0.299+	0.158	-0.153	0.146
BIIT*WILN(I Expo)				
BIIT*BIIExpe				
BIIExpe*WILN(I Expo)				
3-way interaction				
Control Fixed Effects				
Respondent-level				
WI R Age *(0=grp)*	0.017***	0.004	0.015***	0.004
WI R Sex = *Women*	0.263*	0.103	0.146	0.092
WI R Race = *A-A*	-0.252+	0.132	-0.233+	0.119
WI I Retrieval *(0=grp)*			0.031***	0.004
Interviewer-level				
BI R Age *(0= 62)*	0.012	0.019	0.045*	0.020
BI R Sex = *% Women*	0.145	0.277	0.284	0.247
BI R Race = *% A-A*	1.480*	0.636	1.107+	0.582
BI I Retrieval (0=17)			0.040***	0.012
BI I Age *(0= 49)*	-0.001	0.005	-0.002	0.005
BI I Sex = *Women*	0.090	0.148	0.037	0.137
BI I Race = *A-A*	-0.811***	0.199	-0.669***	0.186
Variance Components				
Residual Variance	0.534	0.048	0.417	0.039
Intercept Variance	--[b]	--[b]	--[b]	--[b]
Model Fit				
-2Log Likelihood	2279.40		2213.73	
AIC	2309.40		2247.73	
BIC	2327.68		2268.45	
N (Sample Size)	292		292	

[a] ***WI components:*** Group-mean centered (0=grp), ***BI components:*** Grand-mean centered

[b] The G-matrix (random intercept variance) is estimated to be zero; hence could not be kept in the model.

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001; -- Not Applicable

Inexperienced interviewers obtain higher respondent retrieval behaviors and consistent with the Bilgen and Belli (2010) findings, interviewers attain higher respondent retrieval in calendar interviews. However, when interviewer retrieval is added to the model, both effects disappear. This is consistent with the hypotheses that interviewer retrieval mediates the relationship between respondent retrieval behaviors and interviewer experience and interviewing technique. Also, as expected there is a significant positive association between the interviewer and respondent retrieval behaviors at each respondent and interviewer level. In other words, at the respondent level, the interviews who use more interviewer retrieval behaviors than average also obtain more respondent retrieval behaviors. In addition, at the interviewer level, the interviewers who use more retrieval behaviors than average obtain higher levels of respondent retrieval behaviors.

Lastly, several interviewer and respondent characteristics significantly relate to respondent retrieval strategies (see Table 5.4). According to the respondent retrieval model that controls for the interviewer retrieval behaviors, African-American respondents provide lower retrieval behaviors and African-American interviewers attain lower numbers of respondent retrieval behaviors. However, interviewers who interview more European-American respondents obtain lower numbers of retrieval behaviors. Again, the inconsistency in these findings may be due to several interactive mechanisms between interviewer and respondent race effect on respondent retrieval. However, it is not the scope of this dissertation to explore these effects. Moreover, respondents who are older than average provide higher retrieval behaviors and interviewers who interview higher numbers of older respondents obtain higher respondent retrieval behaviors. This may be due to the relationship between response difficulty and respondent age. In these

interviews, respondents are asked regarding their life-time histories; hence, respondents who are older than average are asked to remember events that happened longer ago than respondents who are younger than average. Moreover, as the long term memory declines as one gets older (Neisser & Hyman, 2000), the respondents (who are older than average=62) may need to utilize the retrieval strategies more than the younger respondents (Belli, Stafford, & Alwin, 2009).

CHAPTER VI: ITEM NON-RESPONSE FINDINGS

1. Introduction to Item Non-Response Models

The main purpose of this chapter is to explore how interviewer and respondent behaviors shape the relationship between interviewer exposure, experience, and item non-response in calendar and standardized interviews. While item non-response is utilized as the outcome variable in the further analyses, the interviewer and respondent behaviors are included as the intermediate measures. Item non-response[59] is measured via whether ***unresolved DK response*** occurred at a turn or not. Due to the dichotomous nature of the outcome measure, I used the SAS PROC GLIMMIX procedure, which transforms the item non-response measure via the logit link function in order to be able to analyze it as a continuous measure in the generalized linear mixed models (Hedeker, 2005):

$$\textbf{Transformed } INR_{tri} = \textbf{logit } (P_{INRtri=1}) = \textbf{LN } [P_{INRtri=1} / (1- P_{INRtri=1})],$$

in which LN indicates the ***natural logarithm of*** $P_{INRtri=1} / (1- P_{INRtri=1})$.

$P_{INRtri=1}$ refers to the probability of the respondent providing an unresolved DK at a turn, and $(1- P_{INRtri=1}) = P_{INRtri=0}$ indicates the probability of the respondent NOT providing an unresolved DK at a turn (Snijders & Bosker, 1999).

As the item non-response measure varies among turns, respondents, and interviewers; the further analyses use three-level[60] random intercept only models. These

[59] Item non-response measure does NOT include respondent refusals due to low occurrences (see Chapter III for further discussion).

[60] Seven out of the eight intermediary outcome measures (interviewer and respondent behavior scales) occur at the turn level (see Table 3.11). However, I analyze these at the respondent level throughout the dissertation due to the low occurrences and variation at the turn level (see Chapter III).

models allow the respondent and interviewer effects (i.e., means) to be random and the linear effect of the predictors that describe the individual differences in change (slopes) among respondents and interviewers to be fixed (i.e., to vary systematically). Therefore, in the tables below the G-matrix has three components: Level 1 residual variance, Level 2 (respondent) intercept variance and level 3 (interviewer) intercept variance. In logistic multi-level models the level 1 residual term is constant ($\pi^2/3$) (Hedeker, 2005) and is thus not included in the tables below. The level 2 intercept variance takes into account the dependency among turns within respondents by allowing respondent means (i.e., intercepts) to be random. The level 3 intercept variance takes into account the dependency among respondents within interviewers by allowing interviewer means (i.e., intercepts) to be random. Initially, in the three-level random intercept models "*the estimated G matrix was not positive definite"* due to the interviewer random effects. This computational issue occurred because the level 3 (interviewer) random intercept variance component is estimated as zero (SAS support, 2011). Hence, I specified the use of Cholesky algorithm in PROC GLIMMIX; as, it constraints the G matrix to be non-negative and is a more powerful, stable and computationally efficient algorithm (Bates, 2011; Davis, 2005; SAS institute Inc., 2008).

In each of the models below I explore two types of interviewer experience (i.e., general experience and within-study exposure) and interviewing technique (i.e., condition) as the three main independent variables. As illustrated in Chapter V, both general interviewer experience (IExpe) and interviewing technique (IT) are dummy coded to simplify the interpretation of the estimates and are included at the interviewer level (BIIExpe and BIIT) in the models below. The within-study interviewer exposure

variable is a respondent-level measure and varies between interviewers and respondents. In order to capture the log-linear nature of the relationship between interviewer exposure and item non-response, the exposure measure is transformed into a natural logarithm (LN)[61] format. As illustrated in Chapter V, the LN (interviewer exposure) is a "conceptually grand-mean centered" measure (Hoffman, 2011 personal communication) and has two-components in the model due to the variation in interviewer and respondent-levels: The respondent–level component [WILN (IExpo) =LN (IExpo$_{ri}$)] and the between interviewer-level component [BILN (IExpo) =$LN(\overline{IExpo}_{.i})$]. It is not the scope of this dissertation to explore the latter; however, the between component is included as a control variable in the models, as each interviewer vary in their exposure levels at the end of the study fielding period.

Moreover, the eight intermediate interviewer and respondent verbal behavior measures are included in the models as moderators. As illustrated in Chapters IV and V, the behavior counts are included at the respondent level into the model due to low occurrences at the turn level; hence, they vary among respondents and interviewers. All behaviors are continuous measures and have two components in the item non-response models: The *respondent-level group-mean centered component* compares the between-interview/within-interviewer effect of each interviewer and respondent behaviors on item non-response (WIVB =$VB_{ri}-\overline{VB}_{.i}$) and *the interviewer-level grand-mean* centered component compares the <u>overall</u> between-interviewer effect of interviewer and respondent behaviors on item non-response (BIVB=$\overline{VB}.i-\overline{VB}..$). Lastly, as illustrated in Chapter V, interviewer and respondent characteristics (age, sex, and race) are included as

[61] The majority of the models illustrated in Chapter V and VI which include natural logarithm of *Interviewer Exposure* fit better than the models which include *Interviewer Exposure* as a linear predictor.

control variables in order to take into consideration the potential interviewer and respondent confounding effects. For interpretation purposes, all models illustrated in the tables below report the group-mean centered within-interviewer components and the grand-mean centered between-interviewer components of the measures.

The first models in this chapter explore the experience, exposure, and item non-response relationship in calendar and standardized interviews without including the interviewer and respondent behaviors in the models (see Section 2). Next, Section 3 explores the main effects of each respective behavior on item non-response while controlling for the main independent variables in the models. However, the main aim in Section 3 is to examine whether the effect of interviewer experience and exposure on item non-response is moderated by several different kinds of interviewer and respondent behaviors in calendar and standardized interviews. While the initial models include the interaction effects of main and intermediary independent variables, the final models illustrated in the further tables below are established via backward elimination method. To summarize, the generic version[62] of the model for item non-response outcome measure is:

Turn-Level (Level 1):

$$LN\left[P_{INRtri=1}/\left(1-P_{INRtri=1}\right)\right]=\pi_{0ri}+e_{tri}$$

Respondent-level (Level 2):

$$\pi_{0ri}=\beta_{00i}+\beta_{01i}\,WILN(IExpo_{ri})+\beta_{02i}\,WIVB_{ri}+\beta_{03i}\,[WILN(IExpo_{ri})*WIVB_{ri}]+\beta_{04i}\,WIRage_{ri}$$
$$+\beta_{05i}\,WIRsex_{ri}+\beta_{06i}\,WIRrace_{ri}+U_{0ri}$$

[62] The models in Sections 2, 3 and 4 are derived from the item non-response model illustrated via equations above.

Interviewer-level (Level 3):

$$\beta_{00i} = \gamma_{000} + \gamma_{001}\,BIIExpe_i + \gamma_{002}\,BIIT_i + \gamma_{003}\,BIVB._i + \gamma_{004}\,(BIIExpe_i{*}BIIT_i)$$
$$+ \gamma_{005}\,(BIIExpe_i{*}\,BIVB._i) + \gamma_{006}\,(BIIT_i{*}\,BIVB._i) + \gamma_{007}\,(BIIExpe_i{*}BIIT_i{*}\,BIVB._i)$$
$$+ \gamma_{008}\,BILN\,(IExpo._i) + \gamma_{009}\,BIRage._i + \gamma_{0010}\,BIRsex._i + \gamma_{0011}\,BIRrace._i + V_{00i}$$

$$\beta_{01i} = \gamma_{010} + \gamma_{011}\,BIIExpe_i + \gamma_{012}\,BIIT_i + \gamma_{013}\,BIVB._i + \gamma_{014}\,(BIIExpe_i{*}BIIT_i)$$

$$\beta_{02i} = \gamma_{020} + \gamma_{021}\,(BIIExpe_i) + \gamma_{022}\,(BIIT_i) + \gamma_{024}\,(BIIExpe_i{*}BIIT_i)$$

$$\beta_{03i} = \gamma_{030} + \gamma_{031}\,(BIIExpe_i) + \gamma_{032}\,(BIIT_i) + \gamma_{034}\,(BIIExpe_i{*}BIIT_i)$$

$$\beta_{04i} = \gamma_{040}$$

$$\beta_{05i} = \gamma_{050}$$

$$\beta_{06i} = \gamma_{060}$$

2. The Relationship between Interviewer Experience, Exposure, and Unresolved "Don't Know" Responses in Calendar and Standardized Interviews

Table 6.1 illustrates the findings for whether interviewers with different experience and exposure levels differ in obtaining unresolved don't know responses in calendar and standardized interviews when the interviewer and respondent verbal behaviors are not taken into account. The initial model contain the experience, exposure and interviewing technique interaction; however, the final model—which is obtained via backward elimination—indicates that the interviewer experience and exposure interaction does not significantly differ in calendar and standardized interviews. According to Table 6.1, the final model with an experience and exposure interaction reveals that in both calendar and standardized interviews, while inexperienced interviewers are less likely to obtain don't know responses than experienced interviewers at the beginning of the study fielding period, the item non-response gap between inexperienced and experienced interviewers

Table 6.1

Item NR Models without Interviewer and Respondent Behaviors

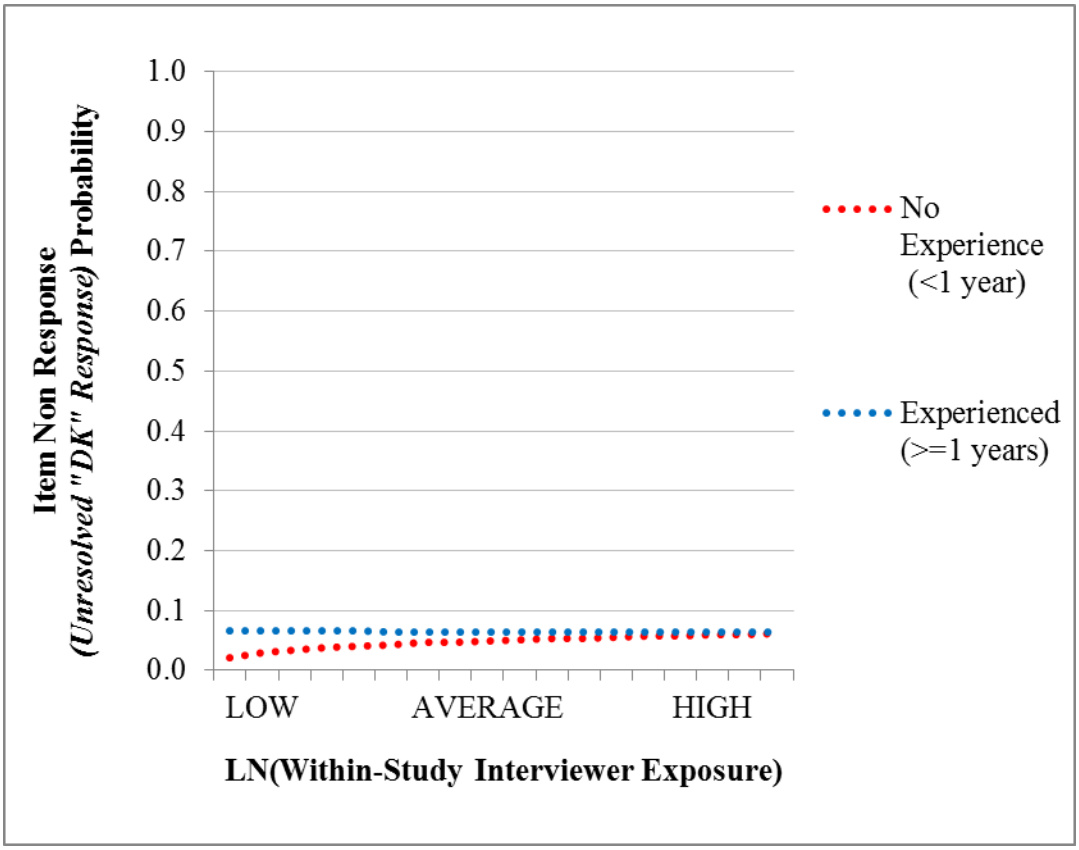
	Initial "Don't Know" Response Models [a]		Final "Don't Know" Response Models [a]	
Parameters	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects				
Intercept	-3.851***	0.457	-3.849***	0.421
WI LN(I Expo)	0.326*	0.137	0.299**	0.109
BI LN(I Expo)	0.108	0.189	0.101	0.167
BI IT = *Standardized*	0.615	0.555	0.380***	0.095
BI IExpe =*1 year or +*	1.144**	0.379	1.128***	0.300
BIIT*WILN(I Expo)	-0.090	0.223		
BIIT*BIIExpe	-0.097	0.612		
BIIExpe*WILN(I Expo)	-0.281+	0.160	-0.303*	0.124
BIIT*BIIExpe*WILN(I Expo)	-0.016	0.249		
Control Fixed Effects				
Respondent-level				
WI R Age *(grp-mean)*	0.015***	0.004	0.015***	0.004
WI R Sex = *Women*	0.140	0.095	0.146	0.095
WI R Race = *A-A*	0.006	0.123	0.010	0.123
Interviewer-level				
BI R Age *(0= 62)*	0.000	0.015	0.000	0.015
BI R Sex = *% Women*	-0.727**	0.255	-0.730**	0.256
BI R Race = *% A-A*	-0.256	0.637	-0.098	0.591
BI I Age *(0= 49)*	0.005	0.005	0.004	0.005
BI I Sex = *Women*	-0.097	0.144	-0.094	0.143
BI I Race = *A-A*	0.223	0.195	0.214	0.193
Variance Components [b]				
Level 2 Intercept Variance	0.374	0.047	0.378	0.047
Level 3 Intercept Variance	0.000	0.105	0.000	0.095
Model Fit				
-2Log Likelihood	23219.37		23220.77	
AIC	23259.37		23254.77	
BIC	23283.75		23275.49	
N (Sample Size)	76021		76021	

[a] ***Within-interviewer (WI) components:*** Group-mean centered *(grp-mean)*,
Between-interviewer (BI) components: Grand-mean centered

[b] The G-matrix (intercept variance) is estimated via Cholesky algorithm

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001

Figure 6.1 The Relationship between Interviewer Experience, Exposure and Item Non-Response (Unresolved "DK" Responses)[63]



diminishes at the end of the study fielding period (see Figure 6.1). However, one caveat in these findings is that the probability of obtaining an unresolved "don't know" response in a turn is very low. Thus, the change in this item non-response measure is significant but very small throughout the study fielding period. One theory that can justify the precision of the significance in the interactions is that the small probabilities may occur because item non-response measure is a turn-specific variable (rather than item-specific, which occurs at a higher level). This may cause under-estimation in the item non-

[63] The unresolved DK measures are collected at the turn-level. Therefore, it is the probability of an unresolved DK response to occur at a turn among all turns respondents are providing (including the turns he/she are digressing, providing clarifications, asking for clarifications, etc.). Hence, as the unresolved DK measure is a turn-specific measure (not item-specific); this may cause an under-estimation in the item non-response probabilities illustrated in the figure above.

response probabilities. As indicated earlier, the issue of using turn as the unit of analyses needs to be examined further. It is not the scope of this dissertation to do so; however, this issue is acknowledged as a limitation of this dissertation (see Chapter VII).

Nevertheless, consistent with the earlier findings in Chapter V, experienced interviewers obtain similar numbers of unresolved don't know behaviors from respondents throughout the study fielding period regardless of the interviewing technique they are using during questionnaire administration. As the literature hypothesized, this may indicate that, on the one hand, as experienced interviewers bring their own interviewing habits and behaviors, they do not modify these habits and behaviors throughout the study fielding period (Hughes et al., 2002; Olson & Peytchev, 2007; Pickery & Loosveldt, 2001). On the other hand, inexperienced interviewers may modify their behaviors, and hence obtain different levels of unresolved "don't know" responses in earlier and later studies throughout the study fielding period. In order to test these theories, the next sections include interviewer and respondent behaviors as intermediary measures into the final model that is illustrated in Table 6.1.

Lastly, according to Table 6.1, respondent age significantly relates to respondent's unresolved "don't know" responses. Older respondents consistently provide higher levels of unresolved "don't know" responses, which may be due to the relationship between question difficulty and respondent age. In these interviews, respondents are asked regarding their life-time histories; hence, respondents who are older than average are asked to remember events that happened longer ago than respondents who are younger than average. Also, the significant effect of interviewer level respondent sex indicates that interviewers who interview more female respondents obtain lower numbers

of unresolved "don't know" responses. This effect may be due to the higher numbers of female interviewers in the study (see Chapter III); hence, there may be a higher likelihood of demographic match between female interviewers and respondents (more discussion is provided later in this chapter).

3. The Role of Interviewer and Respondent Behaviors in the relationship among Interviewer Experience, Exposure and Item Non-Response in Calendar and Standardized Interviews

This dissertation is organized around three sets of interviewer and respondent behaviors: 1- Interviewers' deviation from conventional ideals; 2-Interviewer and respondent interpersonal dynamics; and 3-interviewer and respondent retrieval strategies. This section explores when each of the interviewer and respondent behaviors are included as control variables whether there is a change in the interviewer experience, exposure, and item non-response relationship illustrated in Section 2 above. Moreover, the effect of the within and between components of each of these behaviors on item non-response will be explored separately in the models below.

DEVIATION FROM CONVENTIONAL IDEALS

The models exploring the role of deviation from conventional ideals correspond to hypothesis 4 in Chapter IV. Table 6.2 illustrates the findings from the **final models** regarding the relationship between interviewer experience, exposure, and item non-response when each of the deviation components is included separately in the models. The deviation from conventional ideals has three components: *Interviewing pace*,

interviewer directive behaviors, and *interviewer verbal behavior deviation behaviors*[64]. In Table 6.2, each of the three deviation behavior component is examined separately and "WI Verbal Behavior" and "BI Verbal Behavior" refer to the within and between interviewer main effects on item non-response for each of the deviation behaviors.

Consistent with the hypothesis 4 in Chapter IV, each of the *within-interviewer deviation behaviors* significantly relate to item non-response (i.e., the differential use of deviation behaviors at an interview relate to the probability of obtaining "unresolved don't know" responses at the interviews). However, none of the *between-interviewer deviation behaviors* relate to item non-response. In other words, interviewers who use deviation behaviors more than average do not significantly obtain "unresolved don't know" responses than interviewers who use deviation behaviors less than average (see Table 6.2). Overall, interviews that are faster than average are more likely to have lower unresolved "don't know" responses. However, in the interviews in which interviewers use more directive probing and verbal behavior deviation behaviors are more likely to obtain higher "don't know" responses than the interviews in which interviewers use lower than average directive and deviation behaviors. Although fast pace relates to lower item non-response in these analyses, it still does not mean that faster pace interviews obtain higher quality retrospective reports. One explanation is that in the faster pace interviews, respondents may not be able to take time to think about the responses in order to not to interrupt the question-answer flow. In addition, the positive association between directive probes, verbal behavior deviations, and item non-response illustrates that the respondent level deviation behaviors increase item non-response (which is one data

[64] Interviewer verbal behavior deviation behaviors include: *Failure to probe a question or a part of the question, failure to probe after a "don't know" response, and improper positive feedback* (see Chapter III for further details on construction of the deviation scale).

Table 6.2

Item NR Models with Interviewers' Deviation from Conventional Ideals[a] (1)

	w/ Interviewing Pace		w/ I'wer Directive Behaviors		w/ I'wer VB Deviation	
Parameters	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects						
Intercept	-3.939***	0.467	-3.812***	0.433	-4.016***	0.418
WI LN(I Expo)	0.308**	0.106	0.291**	0.108	0.257**	0.099
BI LN(I Expo)	0.145	0.184	0.097	0.165	0.157	0.157
BI IT = *Standardized*	0.356**	0.120	0.352	0.250	0.395***	0.087
BI IExpe =*1 year or +*	1.241***	0.310	1.108***	0.304	1.229***	0.282
BIIExpe*WILN(I Expo)	-0.319**	0.121	-0.303*	0.122	-0.329**	0.113
WI Verbal Behavior	-0.007*	0.003	0.011*	0.005	0.032***	0.004
BI Verbal Behavior	0.001	0.004	-0.001	0.009	0.008	0.009
Control Fixed Effects						
Respondent-level						
WI R Age *(grp-mean)*	0.015***	0.004	0.014**	0.004	0.010**	0.004
WI R Sex = *Women*	0.208*	0.095	0.139	0.094	0.108	0.087
WI R Race = *A-A*	0.016	0.121	0.009	0.122	-0.012	0.112
Interviewer-level						
BI R Age *(0= 62)*	0.008	0.015	0.000	0.015	0.005	0.016
BI R Sex = *% Women*	-0.714**	0.253	-0.721**	0.266	-0.619*	0.242
BI R Race = *% A-A*	-0.645	0.618	-0.078	0.587	-0.128	0.545
BI I Age *(0= 49)*	0.005	0.005	0.004	0.005	0.004	0.005
BI I Sex = *Women*	-0.084	0.145	-0.089	0.144	-0.118	0.130
BI I Race = *A-A*	0.363+	0.215	0.198	0.201	0.339	0.210
Variance Components[b]						
Level 2 Intercept Var.	0.352	0.045	0.367	0.046	0.288	0.037
Level 3 Intercept Var.	0.000	0.382	0.000	0.095	0.000	0.161
Model Fit						
-2Log Likelihood	22398.37		23214.18		23165.33	
AIC	22436.37		23252.18		23203.33	
BIC	22459.53		23275.34		23226.49	
N (Sample Size)	73293		76021		76021	

[a] ***Within-interviewer (WI) components:*** Group-mean centered *(grp-mean)*, ***Between-interviewer (BI) components:*** Grand-mean centered

[b] The G-matrix (intercept variance) is estimated via Cholesky algorithm

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001

quality dimension)[65]. On the surface, the deviation and directive behaviors may seem "undesirable"; however, the relationship between interviewer deviation behaviors and item non-response is not as simple as it seems, as the structure of this association changes when *the effect of differential use of behaviors among interviewers with different levels of experience and exposure on item non-response* is examined in the further findings.

In addition, Table 6.2 indicates that the deviation behaviors do not mediate the relationship between interviewer experience, exposure and item non-response as this relationship (illustrated via BIIExpe*WILN (I Expo) in Table 6.2) gets stronger when controlling for the interviewing pace and interviewer verbal behavior deviations. Consistent with the hypothesis 5 and 6 in Chapter VI the deviation behaviors affect the relationship between interviewer experience, exposure, and item non-response. Further analyses correspond to hypothesis 5 and 6 in Chapter IV and illustrate whether deviation from conventional behaviors moderate the effect of experience and exposure on item non-response in calendar and standardized interviews via examining the role of within and between interviewer deviations:

The moderator effect of between-interviewer deviation:

- *(BIIT * BIIEXPE * BIDEV)*[66] tests whether the effect of the deviation behaviors on item non-response differs among interviewers with differential experience

[65] One interesting finding is that when pace is included in the models, the relationship between interviewing technique and item non-response weakens and even disappears when directive behaviors are included in the models. This indicates that pace and deviation behaviors mediate the relationship between interviewing technique and item non-response. However, it is neither the scope of this paper, nor a new idea to explore this relationship. Rather, further analyses in this chapter explore whether *the differential use of behaviors among interviewers with different levels of experience and exposure affect item non-response* ***in calendar and standardized interviews***.

[66] Please note that the **initial** models that include the 3-way interactions illustrated above also include the relevant two-way interactions and main effects. Moreover, as the **final** models (illustrated in Table 6.3) include significant three-way interactions, these models also include the relevant two-way interactions and the main effects.

levels in different interviewing techniques, regardless of interviewers' exposure levels (i.e. regardless of the usage of deviation behaviors in early versus later interviews during the study fielding period).

The moderator effect of within-interviewer deviation:

- *(BIIEXPE*WI[LN(IEXPO)]*WIDEV)* tests whether the effect of the interviewer deviation behaviors on item non-response differs among interviewers with different experience levels and among respondents who are interviewed by interviewers with differential exposure levels, regardless of the interviewing technique.
- *(BIIT * WI[LN(IEXPO)] * WIDEV)* tests whether the effect of the interviewer deviation behaviors on item non-response differs across respondents who are interviewed by interviewers with differential exposure levels in calendar and standardized interviews, regardless of the interviewer experience.

According to these models, the between-interviewer (BI) deviation behaviors (**between interviewer interviewing pace and verbal behavior deviation**) moderate the relationship among interviewer experience and item non-response in calendar and standardized interviews; this moderation effect is not observed among the within-interviewer (WI) deviation behaviors. According to Table 6.3, the final moderation models indicate that two of the three BI deviation behaviors (BI interviewing pace and BI interviewers' verbal behavior deviation) moderate the relationship among interviewer experience, interviewing technique, and item non-response in a consistent manner. In standardized interviews, experienced interviewers with faster BI interviewing pace than average and who use higher BI deviation behaviors than average are more likely to obtain

Table 6.3

Item NR Models with Interviewers' Deviation from Conventional Ideals[a] (2)

	w/ Interviewing Pace		w/ I'wer VB Deviation	
Parameters	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects				
Intercept	-4.066***	0.545	-3.918***	0.549
WI LN(I Expo)	0.319**	0.108	0.270**	0.101
BI LN(I Expo)	0.194	0.201	0.134	0.205
BI IT = *Standardized*	1.219*	0.600	0.227	0.279
BI IExpe =*1 year or +*	1.357***	0.368	1.251***	0.321
BIIExpe*WILN(I Expo)	-0.336**	0.124	-0.339**	0.116
WI Behavior (Pace or Deviation)	-0.007**	0.003	0.032***	0.004
BI Behavior (Pace or Deviation)	0.009	0.014	0.006	0.020
BIIT*BIIExpe	-0.897	0.596	0.129	0.315
BIIT*BI Behavior	-0.073	0.048	-0.088	0.056
BIIExpe*BI Behavior	-0.013	0.015	-0.007	0.024
BIIT*BIIExpe*BI Behavior	0.083+	0.048	0.106+	0.055
Control Fixed Effects				
Respondent-level				
WI R Age *(grp-mean)*	0.015***	0.004	0.010**	0.004
WI R Sex = *Women*	0.212*	0.094	0.106	0.087
WI R Race = *A-A*	0.017	0.119	-0.012	0.112
Interviewer-level				
BI R Age *(0= 62)*	0.000	0.016	-0.002	0.018
BI R Sex = *% Women*	-0.546*	0.262	-0.527*	0.267
BI R Race = *% A-A*	-1.245	0.757	-0.338	0.594
BI I Age *(0= 49)*	0.000	0.006	0.002	0.005
BI I Sex = *Women*	-0.167	0.149	-0.173	0.137
BI I Race = *A-A*	0.335	0.355	0.146	0.251
Variance Components[b]				
Level 2 Intercept Variance	0.341	0.044	0.288	0.037
Level 3 Intercept Variance	0.000	0.083	0.000	0.078
Model Fit				
-2Log Likelihood	22392.48		23161.01	
AIC	22438.48		23207.01	
BIC	22466.51		23235.04	
N (Sample Size)	73293		76021	

[a] ***Within-interviewer (WI) components:*** Group-mean centered *(grp-mean)*, ***Between-interviewer (BI) components:*** Grand-mean centered

[b] The G-matrix (intercept variance) is estimated via Cholesky algorithm

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001

Figure 6.2 Interviewer Experience, Interviewing Pace and Item Non-Response (Unresolved "DK" Responses) in Calendar and Standardized Interviews

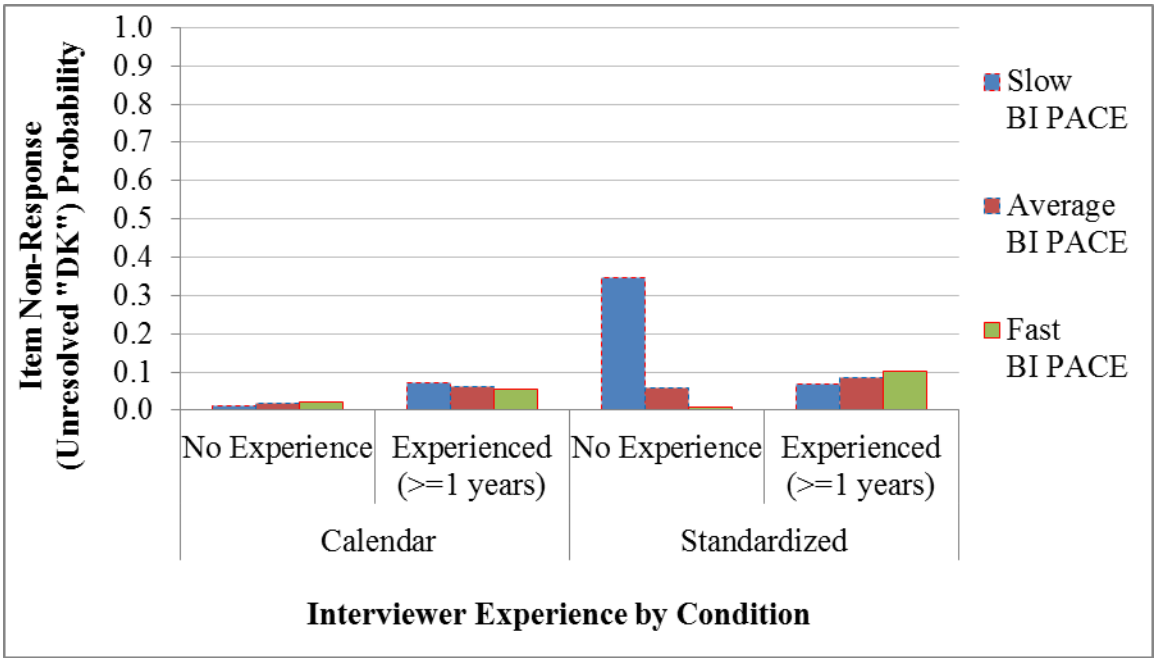
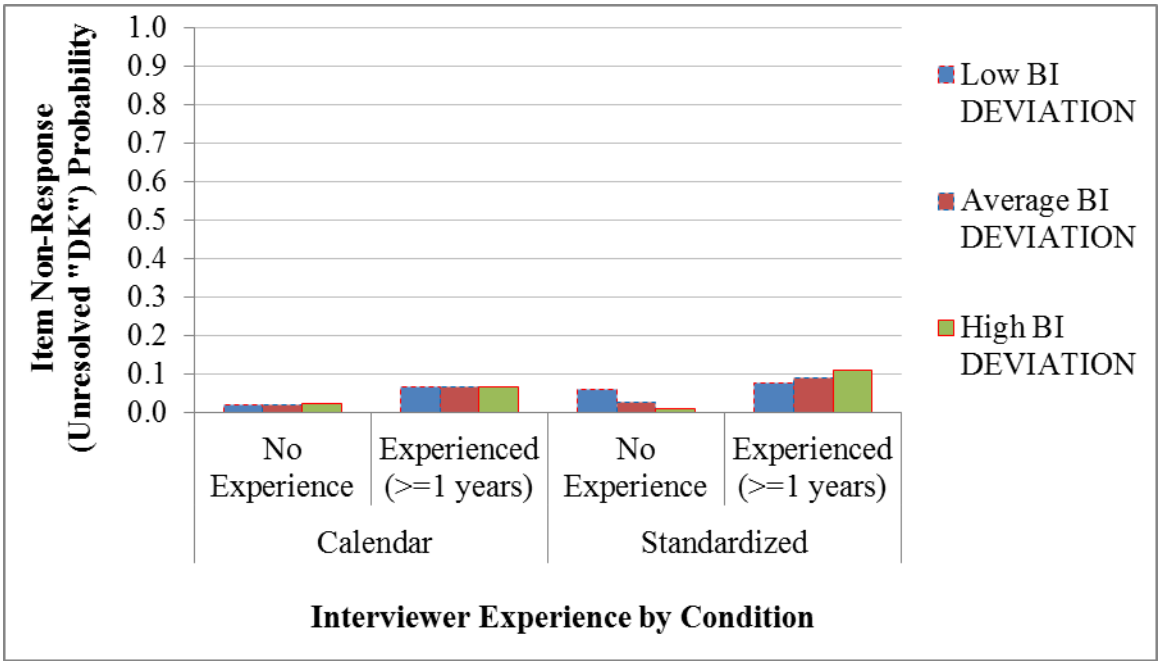


Figure 6.3 Interviewer Experience, Interviewers' VB Deviation and Item Non-Response (Unresolved "DK" Responses) in Calendar and Standardized Interviews[67]



[67] As the unresolved DK measure is a turn-specific measure (not item-specific); this may cause an under-estimation in the item non-response probabilities illustrated in the figures above.

higher unresolved "don't know" behaviors than experienced interviewers with slower BI pace and deviation behaviors than average (see Figures 6.2 and 6.3). However, again in standardized interviews the direction of the between-interviewer deviation behaviors and item non-response relationship is the opposite among inexperienced interviews. In other words, inexperienced interviewers with faster BI pace and higher levels of deviation obtain lower unresolved "don't know" behaviors than inexperienced interviewers with slower BI pace and lower levels of deviation behaviors than average.

The results illustrate that in standardized interviews different mechanisms play a role among interviewers with different experience levels. The literature suggests the use of high deviation behaviors (fast pace and high numbers of verbal behavior deviations) among experienced interviewers is an indication of interviewer carelessness as they increase item non-response, whereas the slow pace and low deviation behaviors among inexperienced interviewers may be an indication of problems with transition from training to practice. This is an important finding, as this adds another dimension to the standardized interviewing studies that explore *"which behaviors affect data quality?"*

The findings suggest that the "so-called undesirable" interviewer behaviors (such as deviation from conventional ideals) may become desirable when inexperienced interviewers are using these behaviors. Hence, the question rather becomes *"who uses which behaviors and how does this affect data quality?"* especially in standardized interviews. Moreover, another interesting finding is that in calendar interviews, the differences in the use of deviation behaviors among inexperienced and experienced interviewers are more similar in comparison to standardized interviews. In other words, the change in item non-response probabilities due to the differences in deviation

behaviors among experienced and inexperienced interviewers is significantly lower in calendar interviews than in standardized interviews. .

INTERVIEWER AND RESPONDENT INTERPERSONAL DYNAMICS

The models exploring the role of interviewer and respondent interpersonal dynamics respond to hypothesis 11 in Chapter IV. Table 6.4 illustrates the findings from the **final models** regarding the relationship between interviewer experience, exposure, and item non-response when the interviewer and respondent interpersonal behaviors are included in the models. The interviewer and respondent interpersonal communication has three components: *Interviewer rapport behaviors*[68], *respondent cooperative behaviors*[69] and *respondent expression of difficulty and non-cooperation behaviors*[70].

According to Table 6.4, interviewers who attain more than average respondent cooperative behaviors are less likely to obtain item non-response. Respondents who provide higher levels of cooperative or difficulty and non-cooperative behaviors are more likely to provide higher levels of item non-response. Inconsistent with the hypothesis, respondents who are more vocal about their interviewing experiences (i.e., respondents who are providing higher levels of cooperative, as well as difficulty and non-cooperative behaviors) are also more comfortable with not resolving their "don't know" responses. On the contrary, interviewers who are more successful at obtaining respondent cooperative behaviors are more successful at decreasing item non-response. Moreover,

---

[68] **Interviewer Rapport behaviors include**: *Interviewer empathy, agreement, joking or sarcasm, direct apology, and laughter to a respondent joke or comment (*see Chapters III and IV for further discussion).
[69] **Respondent Cooperative behaviors include**: *Respondent's jokes or sarcastic comments, spontaneous clarifications, and corrections.*
[70] **Respondent Expressions of difficulty and non-cooperative behaviors include**: *Respondent expressions of difficulty answering a question, seeking clarification, guesses, estimates, and negative comments about the interview/task/questionnaire.*

Table 6.4

Item NR Models with Interpersonal Communication Verbal Behaviors[a] (1)

	w/ I'wer Rapport		+ w/ R Cooperation		+w/ R Difficulty & Non-Cooperation	
Parameters	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects						
Intercept	-3.841***	0.416	-3.819***	0.412	-3.796***	0.413
WI LN(I Expo)	0.275*	0.109	0.293**	0.108	0.247*	0.108
BI LN(I Expo)	0.122	0.167	0.106	0.165	0.082	0.165
BI IT = *Standardized*	0.408***	0.098	0.397***	0.097	0.421***	0.097
BI IExpe =*1 year or +*	1.130***	0.298	1.138***	0.296	1.032***	0.296
BIIExpe*WILN(I Expo)	-0.303*	0.123	-0.335**	0.122	-0.248*	0.122
WI I'wer Rapport VB	0.000	0.005	-0.007	0.006	-0.008	0.005
BI I'wer Rapport VB	0.006	0.008	0.015+	0.009	0.007	0.009
WI R Coop or Diff VB	--	--	0.002+	0.001	0.007***	0.002
BI R Coop or Diff VB	--	--	-0.005*	0.002	-0.001	0.004
Control Fixed Effects						
Respondent-level						
WI R Age *(grp-mean)*	0.015***	0.004	0.012**	0.004	0.013**	0.004
WI R Sex = *Women*	0.162+	0.096	0.145	0.095	0.133	0.094
WI R Race = *A-A*	0.008	0.123	0.008	0.121	-0.028	0.122
Interviewer-level						
BI R Age *(0= 62)*	-0.001	0.015	0.004	0.015	0.001	0.015
BI R Sex = *% Women*	-0.792**	0.267	-0.640*	0.273	-0.778**	0.266
BI R Race = *% A-A*	0.091	0.598	0.255	0.595	0.235	0.606
BI I Age *(0= 49)*	0.003	0.005	0.005	0.005	0.003	0.005
BI I Sex = *Women*	-0.176	0.158	-0.211	0.158	-0.178	0.157
BI I Race = *A-A*	0.284	0.199	0.208	0.198	0.292	0.197
Variance Components[b]						
Level 2 Intercept Var.	0.365	0.045	0.354	0.044	0.350	0.044
Level 3 Intercept Var.	0.000	0.086	0.000	0.064	0.000	0.087
Model Fit						
-2Log Likelihood	23219.70		23211.24		23205.97	
AIC	23257.70		23253.24		23247.97	
BIC	23280.85		23278.83		23273.57	
N (Sample Size)	76021		76021		76021	

[a] ***WI components:*** Group-mean centered *(grp-mean)*, ***BI components:*** Grand-mean centered

[b] The G-matrix (intercept variance) is estimated via Cholesky algorithm

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001; -- Not Applicable

according to Table 6.4, the interviewer and respondent communication behaviors do not mediate the relationship between interviewer experience, exposure and item non-response as this relationship [illustrated via BIIExpe*WILN (I Expo)] gets stronger when controlling for the interviewer rapport and respondent cooperative behaviors.

The scope of this section is to understand which types of interviewers provide higher interviewer rapport behaviors and obtain higher respondent communicative behaviors. Further analyses correspond to hypothesis 12 and 13 in Chapter IV and investigate whether interviewer and respondent interpersonal communication behaviors change the effect of experience and exposure on item non-response in calendar and standardized interviews. In other words, further analyses examine the role of within and between interviewer and respondent interpersonal communication dynamic behaviors (which are referred as VB throughout this section):

The moderator effect of interpersonal communication dynamics among interviewers:

- *(BIIT * BIIEXPE * BIVB)*[71] tests whether the effect of interviewer rapport behaviors and respondent communication behaviors on item non-response differs among interviewers with differential experience levels in calendar and standardized interviews, regardless of interviewers' exposure levels.

The moderator effect of interpersonal dynamics among interviews/respondents:

- *(BIIEXPE*WI [LN (IEXPO)]*WIVB)* tests whether the effect of the interviewer rapport behaviors and respondent communication behaviors on item non-response differs among interviewers with different experience levels and among

[71] Please note that the **initial** models that include the 3-way interactions illustrated above also include the relevant two-way interactions and main effects. Moreover, as the **final** models (illustrated in Tables 6.5) include only two-way interactions; these models also include the relevant main effects.

respondents who are interviewed via interviewers with differential exposure levels, regardless of the interviewing technique.

- *(BIIT * WI [LN (IEXPO)] * WIVB)* tests whether the effect of the interviewer rapport behaviors and respondent communication behaviors on item non-response differs across respondents who are interviewed by interviewers with differential exposure levels in calendar and standardized interviews, regardless of the interviewers' experience levels.

According to the analyses, the within-interviewer (WI) communicative behaviors (the interviewer rapport, respondent cooperation, difficulty, and non-cooperative verbal behaviors) moderate the relationship between interviewer experience and item non-response in calendar and standardized interviews. This moderation effect is not observed among the between-interviewer (BI) communicative behaviors. In other words, regardless of the interviewing technique and interviewer experience levels, the differences in interpersonal communication dynamics come into play in later interviews. This may indicate that when interviewers gain exposure, they are more likely to deviate from the training provided at the beginning of the study (see Table 6.5). Specifically, interviewers who provide lower levels of rapport than average are less likely to obtain unresolved "don't know" responses in earlier interviews than the interviewers who use higher levels of rapport behaviors. However, the item nonresponse gap diminishes at the end of the study and the interviewers providing low levels of rapport are more likely to obtain unresolved "don't know" responses in later interviews than the interviewers who provide higher rapport (see Figure 6.4). Consistent with these findings, interviewers who obtain lower levels of respondent cooperation and difficulty behaviors are more likely to

Table 6.5

Item NR Models with Interpersonal Communication Verbal Behaviors[a] (2)

	w/ I'wer Rapport by Exposure		w/ R Cooperation by Exposure		w/ R Diff & Non-Coop by Exposure	
Parameters	**Beta**	**SE**	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects						
Intercept	-3.922***	0.418	-3.967***	0.415	-3.857***	0.418
WI LN(I Expo)	0.242*	0.110	0.266*	0.109	0.219*	0.110
BI LN(I Expo)	0.120	0.167	0.144	0.165	0.089	0.166
BI IT = *Standardized*	0.431***	0.098	0.406***	0.096	0.427***	0.098
BI IExpe =*1 year or +*	1.110***	0.307	1.062***	0.295	1.004**	0.300
BIIExpe*WILN(I Expo)	-0.284*	0.125	-0.323**	0.121	-0.230+	0.124
WI I'wer Rapport VB	0.009	0.009	-0.008	0.006	-0.008	0.005
BI I'wer Rapport VB	0.035*	0.015	0.018+	0.009	0.006	0.009
WI R Coop or Diff VB	--	--	0.003	0.003	0.007+	0.004
BI R Coop or Diff VB	--	--	0.003	0.004	0.010	0.008
WI VB*WI LN(I Expo)	-0.006	0.004	0.000	0.001	0.000	0.002
BI VB*WI LN(I Expo)	-0.015*	0.006	-0.005**	0.002	-0.006+	0.004
Control Fixed Effects						
Respondent-level						
WI R Age *(grp-mean)*	0.015***	0.004	0.012**	0.004	0.012**	0.004
WI R Sex = *Women*	0.165+	0.096	0.141	0.094	0.140	0.095
WI R Race = *A-A*	-0.012	0.123	0.004	0.120	-0.024	0.122
Interviewer-level						
BI R Age *(0= 62)*	0.001	0.015	0.011	0.015	0.007	0.015
BI R Sex = *% Women*	-0.743**	0.271	-0.473+	0.281	-0.734**	0.270
BI R Race = *% A-A*	0.179	0.598	0.394	0.593	0.285	0.611
BI I Age *(0= 49)*	0.003	0.005	0.005	0.005	0.003	0.005
BI I Sex = *Women*	-0.180	0.158	-0.226	0.156	-0.184	0.157
BI I Race = *A-A*	0.266	0.201	0.145	0.199	0.282	0.199
Variance Components[b]						
Level 2 Intercept Var.	0.359	0.045	0.342	0.043	0.350	0.044
Level 3 Intercept Var.	0.000	0.096	0.000	0.069	0.000	0.130
Model Fit						
-2Log Likelihood	23212.17		23202.85		23202.57	
AIC	23254.17		23248.85		23248.57	
BIC	23279.77		23276.88		23276.60	
N (Sample Size)	76021		76021		76021	

[a] ***WI components:*** Group-mean centered *(grp-mean)*, ***BI components:*** Grand-mean centered

[b] The G-matrix (intercept variance) is estimated via Cholesky algorithm.

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001; -- Not Applicable

Figure 6.4 Interviewer Exposure, Interviewer Rapport, and Item Non-Response (Unresolved "DK" Responses) Relationship

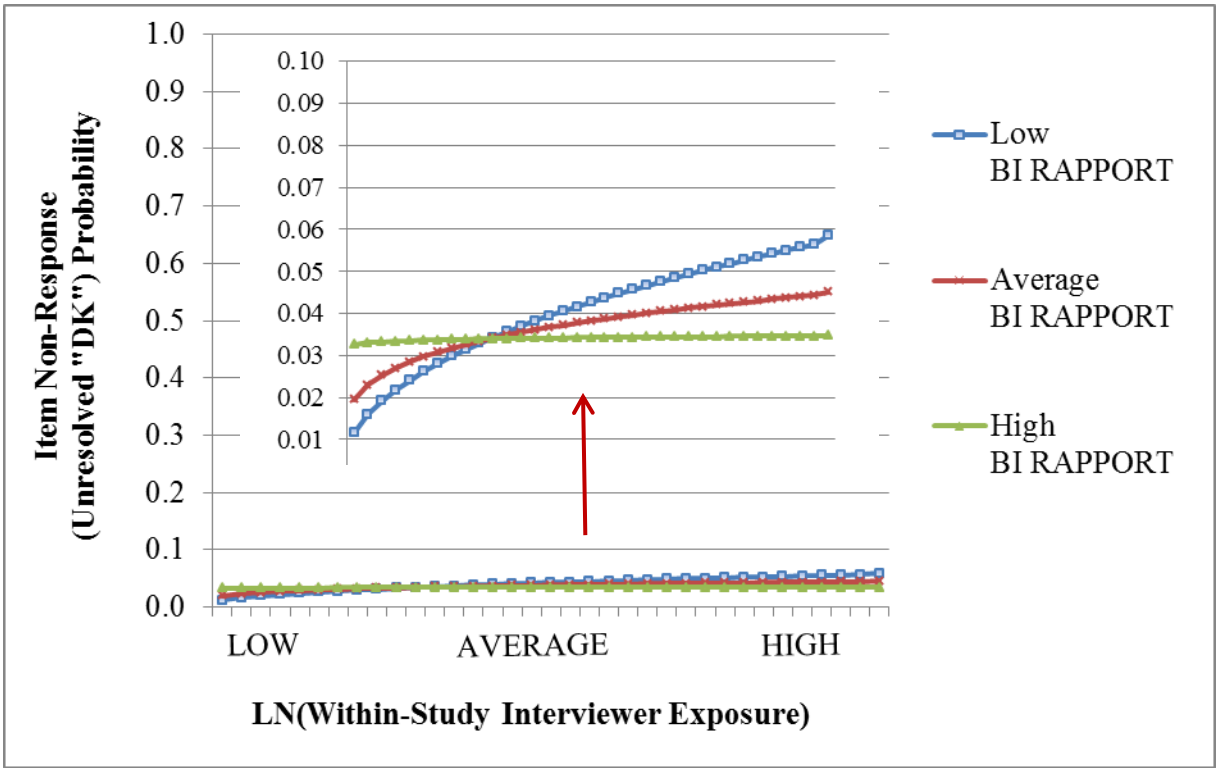


Figure 6.5 Interviewer Exposure, Respondent Cooperation, and Item Non-Response (Unresolved "DK" Responses) Relationship

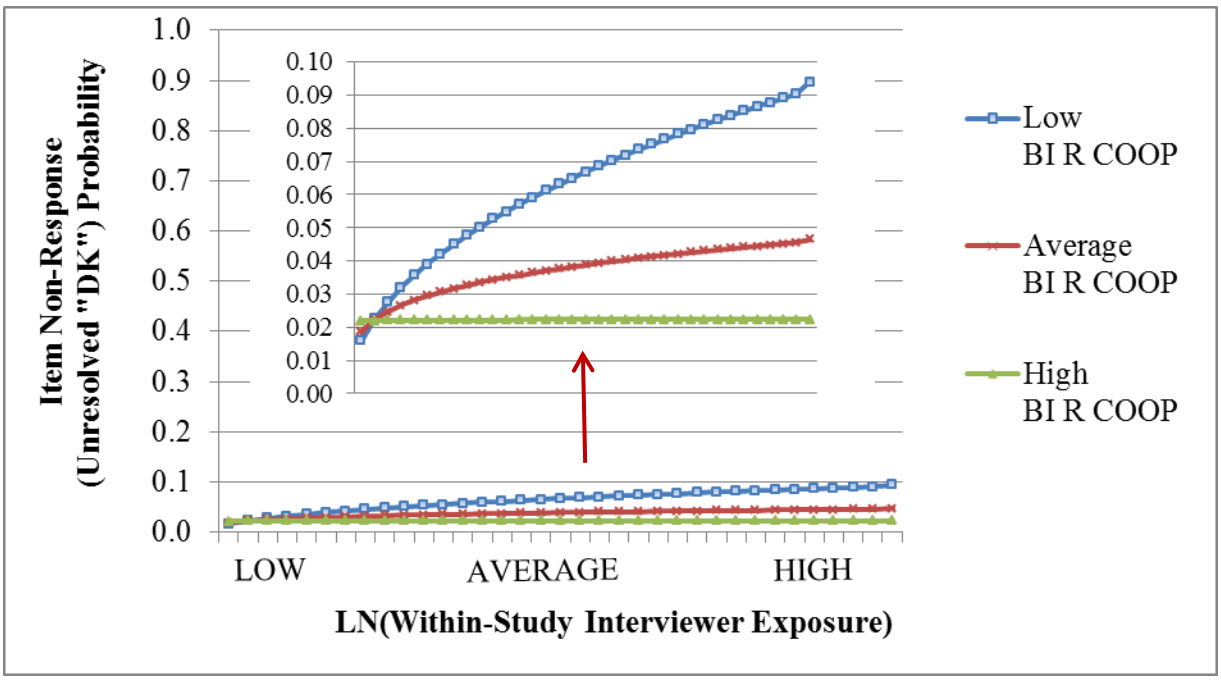
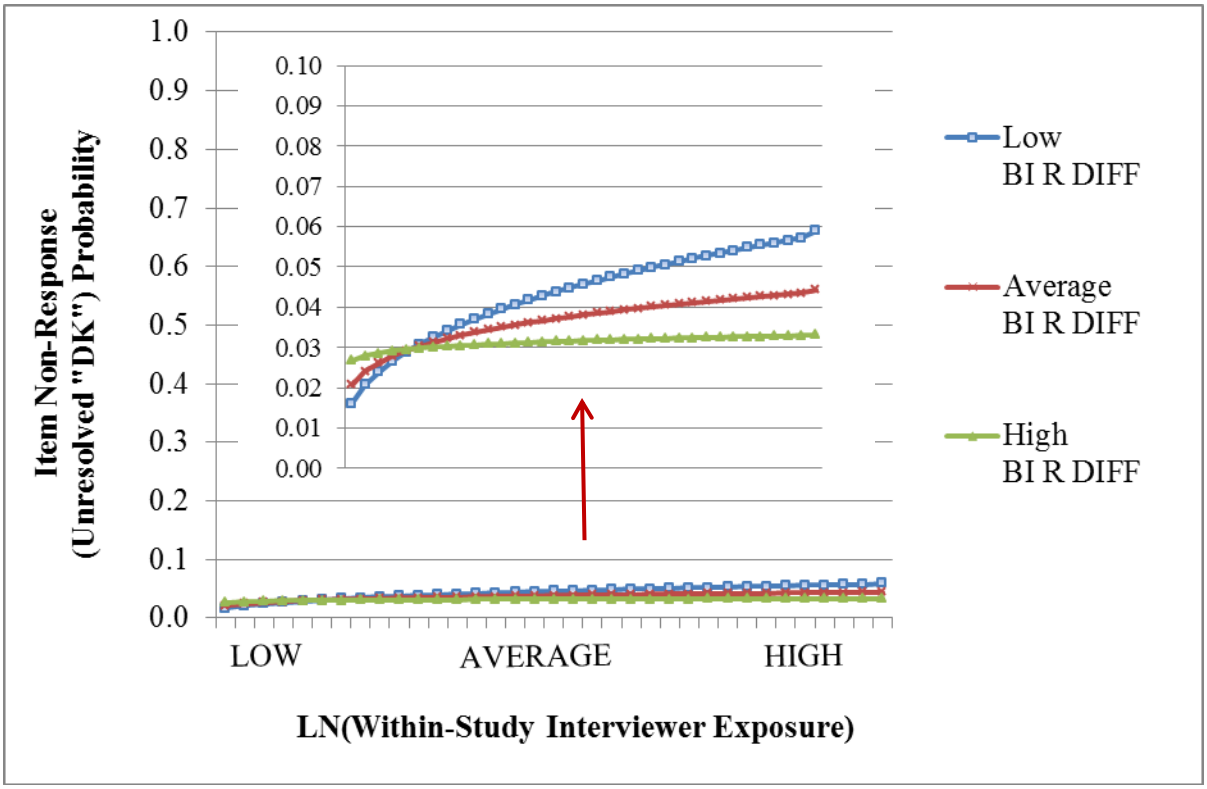


Figure 6.6 Interviewer Exposure, Respondent Difficulty/Non-Cooperation, and Item Non-Response (Unresolved "DK" Responses) Relationship



obtain unresolved "don't know" responses in later interviews in comparison to interviewers who obtain average or high levels of respondent cooperation and difficulty behaviors (see Figures 6.5 and 6.6).

The results are consistent with the theories in the literature regarding interviewer and respondent rapport (Goudy & Potter, 1975-76; Olson & Bilgen, 2011; Olson & Peytchev, 2007). Regardless of interviewers' experience levels and interviewing technique, interviewers who focus on establishing rapport throughout the study fielding period consistently obtain lower levels of item non-response, as they encourage respondents to think harder about their responses. However, interviewers who are less personable and reluctant to establish a harmonious and enjoyable interviewing

environment are more likely to accept respondents' unresolved "don't know" responses especially in later interviews throughout the study. To summarize, the low usage of rapport behaviors becomes detrimental to item non-response when they are used in later interviews (by interviewers who are exposed to the study); BUT, the low usage of rapport is beneficial in the earlier interviews. This indicates that interviewers may get better at providing rapport; as they learn from their earlier experiences in earlier interviews. In other words, they may have figured out how to interact with the questionnaire and the respondent simultaneously. Lastly, these findings may explain the inconclusive findings in the literature regarding the effect of rapport behaviors on item non-response.

RETRIEVAL STRATEGIES AND PROBES

Lastly, Table 6.6 illustrate the findings from the **final models** regarding the relationship between interviewer experience, exposure, and item non-response when the interviewer and respondent retrieval strategies[72] are included in the models. Inconsistent with the Hypothesis 15 in Chapter IV, interviews in which the respondents are provided higher than average interviewer retrieval behaviors are more likely to acquire unresolved "don't know" behaviors. However consistent with the hypothesis, interviewers who attain higher respondent retrieval behaviors are less likely to obtain unresolved "don't know" behaviors. This may indicate that interviewers who are able to encourage and coach respondents regarding how to use retrieval strategies effectively may be more successful at decreasing item non-response than the interviewers who just provide retrieval probes. Moreover, Table 6.6 demonstrates that the relationship between interviewer experience,

[72] Interviewer retrieval probes include: *Parallel, duration, and sequential probing.* Consistently, respondent retrieval strategies include: *Respondent parallel, duration response, and sequential response behaviors* (see Chapter III for further discussion).

exposure and item non-response [illustrated via BIIExpe*WILN (I Expo)] gets stronger when controlling for the interviewer rapport and respondent retrieval behaviors. This result indicates that while interviewer and respondent retrieval behaviors do not mediate but may potentially moderate this relationship. Further analyses correspond to hypothesis 18 and 17 in Chapter IV and examine the role of within and between interviewer and respondent retrieval behaviors.

First, *the moderation effect of different levels of interviewer and respondent retrieval strategies among interviewers* is tested[73]:

- *(BIIT * BIIEXPE * BIVB)* tests whether the effect of interviewer and respondent retrieval behaviors on item non-response differs among interviewers with differential experience levels in calendar and standardized interviews, regardless of interviewers' exposure levels (i.e. regardless of the usage of deviation behaviors in early versus later interviews during the study fielding period).

Then, *the moderation effect of different levels of interpersonal dynamics among interviews/respondents* is tested:

- *(BIIEXPE*WI [LN (IEXPO)]*WIVB)* tests whether the effect of the interviewer and respondent retrieval behaviors on item non-response differs among interviewers with different experience levels and among respondents who are interviewed by interviewers with differential exposure levels, regardless of the interviewing technique.

[73] Please note that the **initial** models that include the 3-way interactions below also include the relevant two-way interactions and main effects. Moreover, as the **final** models (illustrated in Table 6.7) include significant three-way interactions, these models also include the relevant two-way interactions and the main effects.

- *(BIIT * WI [LN (IEXPO)] * WIVB)* tests whether the effect of the interviewer and respondent retrieval behaviors on item non-response differs across respondents who are interviewed by interviewers with differential exposure levels in calendar and standardized interviews, regardless of the interviewers' experience levels.

The analyses indicate that systematic differences in within- and between- respondent retrieval behaviors do not significantly moderate the relationship between interviewer exposure, experience, and item non-response. As for interviewer behaviors, the systematic differences in the use of within-interviewer (WI) retrieval behaviors do moderate the interviewer experience, exposure, and item non-response relationship, whereas a moderation effect due to between-interviewer (BI) retrieval behaviors is not observed (see Table 6.7). According to the Figure 6.7, regardless of the interviewing technique, different mechanisms play a role among interviewers with different experience levels. Specifically, in both calendar and standardized interviews experienced interviewers *who provide higher levels of retrieval probing* are more likely to increase unresolved "don't know" responses in later interviews than in earlier interviews during the study fielding period. As experience is measured via *interviewers' experience with standardized interviews*, these results may indicate that experienced interviewers are less successful at using new probing techniques[74] proficiently than inexperienced interviewers especially in later interviews. This may indicate that experienced interviewers use their learned routine and more reluctant to use new behaviors in comparison to inexperienced interviews. In both calendar and standardized interviews, experienced interviewers *who provide lower levels of retrieval probing* are less likely to obtain unresolved "don't

[74] Such as retrieval strategies that are overall more commonly used in calendar interviews.

Table 6.6

Item NR Models with Interviewer and Respondent Retrieval Verbal Behaviors[a] (1)

	w/ I'wer Retrieval		+ w/ R Retrieval	
Parameters	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects				
Intercept	-3.841***	0.408	-3.918***	0.405
WI LN(I Expo)	0.289**	0.105	0.310**	0.104
BI LN(I Expo)	0.142	0.163	0.080	0.162
BI IT = *Standardized*	0.287+	0.157	0.462**	0.166
BI IExpe =*1 year or* +	1.144***	0.291	1.125***	0.287
BIIExpe*WILN(I Expo)	-0.336**	0.120	-0.364**	0.118
WI I'wer Retrieval VB	0.017***	0.004	0.014**	0.004
BI I'wer Retrieval VB	-0.006	0.010	0.012	0.011
WI R Retrieval VB	--	--	0.004	0.003
BI R Retrieval VB	--	--	-0.021**	0.007
Control Fixed Effects				
Respondent-level				
WI R Age *(grp-mean)*	0.014***	0.004	0.013**	0.004
WI R Sex = *Women*	0.104	0.092	0.093	0.092
WI R Race = *A-A*	0.014	0.119	0.028	0.117
Interviewer-level				
BI R Age *(0= 62)*	-0.005	0.016	0.019	0.017
BI R Sex = *% Women*	-0.701**	0.247	-0.499+	0.256
BI R Race = *% A-A*	-0.014	0.576	0.294	0.575
BI I Age *(0= 49)*	0.004	0.005	0.005	0.005
BI I Sex = *Women*	-0.103	0.139	-0.024	0.140
BI I Race = *A-A*	0.164	0.190	-0.090	0.204
Variance Components[b]				
Level 2 Intercept Variance	0.341	0.043	0.325	0.042
Level 3 Intercept Variance	0.000	0.098	0.000	0.053
Model Fit				
-2Log Likelihood	23201.82		23192.68	
AIC	23239.82		23234.68	
BIC	23262.97		23260.28	
N (Sample Size)	76021		76021	

[a] ***Within-interviewer (WI) components:*** Group-mean centered *(grp-mean)*, ***Between-interviewer (BI) components:*** Grand-mean centered

[b] The G-matrix (intercept variance) is estimated via Cholesky algorithm

+ p<0.10 * p<0.05 **p<0.01 ***p<0.001; -- Not Applicable

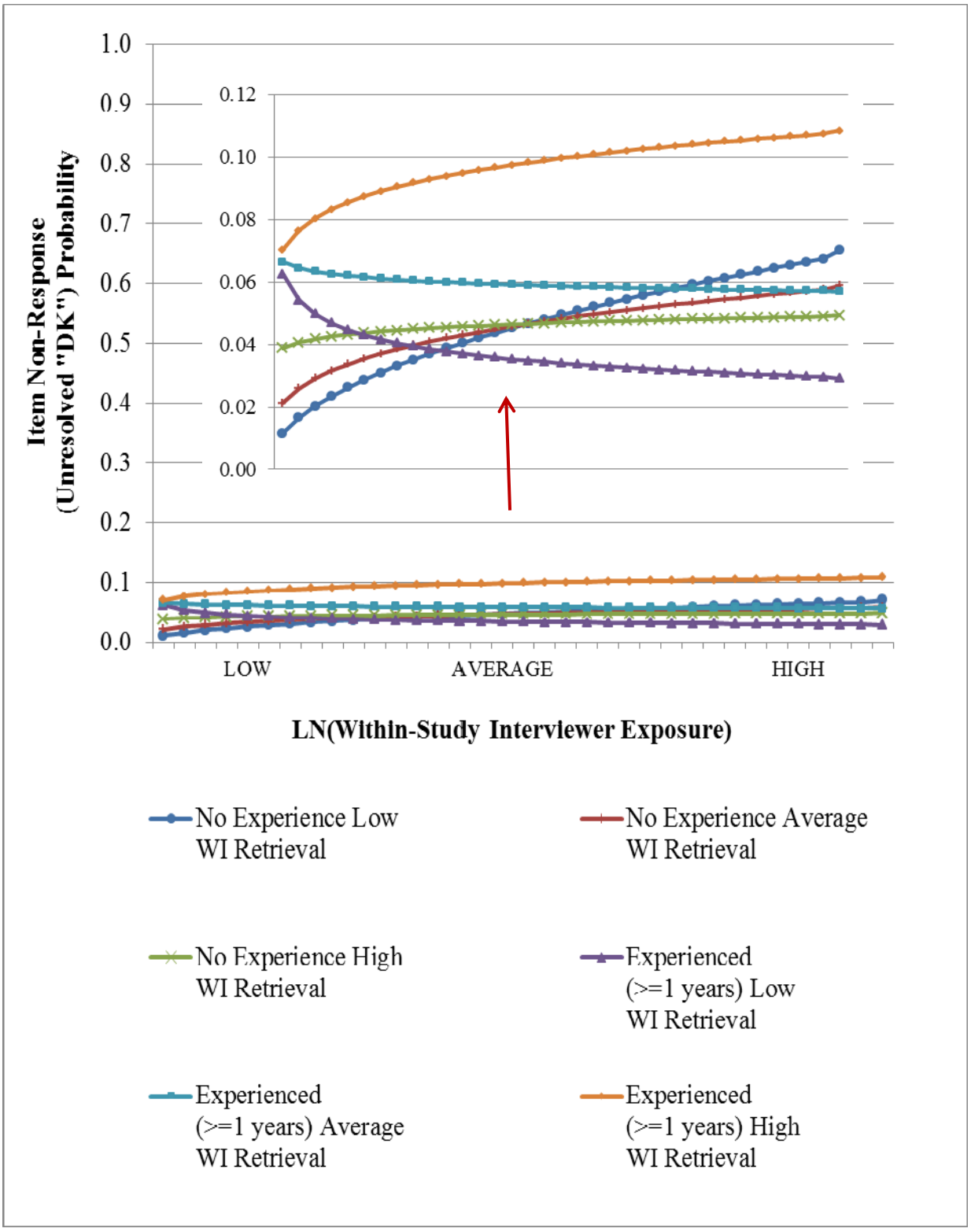
Table 6.7

Item NR Models with Interviewer Retrieval Verbal Behaviors[a]

	I'wer Retrieval by Exposure + Experience			
Parameters	**Beta**	**SE**	**Beta**	**SE**
Main Fixed Effects				
Intercept	-3.833***	0.407	-3.827***	0.439
WI LN(I Expo)	0.299**	0.105	0.291*	0.113
BI LN(I Expo)	0.137	0.162	0.145	0.170
BI IT = *Standardized*	0.276+	0.156	0.254	0.184
BI IExpe =*1 year or +*	1.183***	0.292	1.184***	0.310
BIIExpe*WILN(I Expo)	-0.332**	0.119	-0.340**	0.126
WI I'wer Retrieval (IR)	0.024	0.018	0.029	0.019
BI I'wer Retrieval (IR)	-0.007	0.010	-0.051	0.057
WI IR * WI LN(I Expo)	-0.008	0.008	-0.010	0.008
WI IR * BI IExpe	-0.022	0.021	-0.026	0.022
WI IR * BI IExpe * WI LN(I Expo)	0.017+	0.009	0.019+	0.010
BI IR * WI LN(I Expo)			0.014	0.023
BI IR * BI IExpe			0.050	0.056
BI IR * BI IExpe * WI LN(I Expo)			-0.017	0.023
Control Fixed Effects				
Respondent-level				
WI R Age *(grp-mean)*	0.015***	0.004	0.015***	0.004
WI R Sex = *Women*	0.087	0.092	0.097	0.093
WI R Race = *A-A*	0.011	0.118	0.016	0.118
Interviewer-level				
BI R Age *(0= 62)*	-0.005	0.015	-0.007	0.016
BI R Sex = *% Women*	-0.759**	0.247	-0.746**	0.250
BI R Race = *% A-A*	-0.102	0.574	-0.063	0.623
BI I Age *(0= 49)*	0.004	0.005	0.004	0.005
BI I Sex = *Women*	-0.109	0.138	-0.119	0.139
BI I Race = *A-A*	0.197	0.189	0.202	0.194
Variance Components[b]				
Level 2 Intercept Variance	0.334	0.043	0.334	0.043
Level 3 Intercept Variance	0.000	0.093	0.000	0.090
Model Fit				
-2Log Likelihood	23196.59		23195.40	
AIC	23240.59		23245.40	
BIC	23267.41		23275.87	
N (Sample Size)	76021		76021	

[a] ***WI components:*** Group-mean centered *(grp-mean)*, ***BI components:*** Grand-mean centered
[b] The G-matrix (intercept variance) is estimated via Cholesky algorithm
+ p<0.10 * p<0.05 **p<0.01 ***p<0.001

Figure 6.7 Interviewer Exposure, Interviewer Retrieval Probing, and Item Non-Response (Unresolved "DK" Responses) Relationship



know" responses in later interviews than in earlier interviews throughout the study fielding period. This result may be an indication of trade-offs among behaviors (Bilgen, Belli & Olson, 2009), such that experienced interviewers who choose not to utilize provided retrieval strategies may be using already learned strategies to decrease respondents' unresolved "don't know" behaviors. As mentioned earlier, experienced interviewers may be simply following what they have learned throughout their careers and use their already learned strategies. This finding is consistent with the earlier findings.

To conclude, these results indicate that each set of behavior moderates the relationship between interviewer experience, exposure, and item non-response in different ways. Nevertheless, one consistent finding is that the research question that examines *"which behaviors affect item non-response?"* is an oversimplification. Therefore, it may not always be very accurate to classify interviewer or respondent behaviors as "desirable" and "not desirable" because the association between behaviors and item non-response changes depending on when they are used (in early versus later interviews), who they are used by (experienced versus inexperienced interviewers), or which interviewing method are they used in (calendar versus standardized interviews). In addition, the effect of interviewer and respondent demographics on respondent's unresolved "don't know" responses is consistent among each model illustrated in this chapter (see Tables 6.1 through 6.7).

In all of the models, respondent age and sex significantly relates to respondent's unresolved "don't know" responses. Moreover, while older respondents consistently provide higher levels of unresolved "don't know" responses, interviewers who interview

more female respondents obtain lower numbers of unresolved "don't know" responses. The results related to age may be due to the relationship between question difficulty and respondent age. In these interviews, respondents are asked regarding their life-time histories; hence, respondents who are older than average are asked to remember events that happened longer ago than respondents who are younger than average. Moreover, as the long term memory declines as one gets older (Neisser & Hyman, 2000), the respondents (who are older than average=62) may have a harder time to remember events than the younger respondents (Belli, Stafford, & Alwin, 2009).

Lastly, as there are more female interviewers in the study (see Chapter III); the results related to sex may be due to the higher likelihood of demographic match between female interviewers and respondents. Therefore, the findings in this dissertation may not fully explain the role of interviewer and respondent characteristics on the use of interviewer and respondent behaviors. There may be several other meaningful interactive effects that could be explored. However, as mentioned earlier, it is not the scope of this paper to examine the role of interviewer and respondent characteristics on item non-response. The interviewer and respondent demographics are included for the purpose of controlling the confounding effects of interviewer and respondent characteristics due to the lack of interpenetration (see Chapter III for more discussion).

CHAPTER VII: CONCLUSIONS AND LIMITATIONS

1. Contribution of this Dissertation to the Scientific Community

The main problem of the literature that explores *the relationship among interviewer experience, exposure, and response errors* is that, many of the studies do not take into account (or do not have the data to take into account) the intervening role of interviewer and respondent behaviors in this relationship. Additionally, there are no studies that explore the differential effects of experience and exposure on item non-response between different interviewing techniques. This study is the first of its kind to explore several possible mechanisms involved with interviewer experience, exposure, and item non-response relationship in calendar and standardized interviews.

2. Conclusions and Discussions

Overall, interviewers with high experience and exposure levels have slightly increased item non-response in both standardized and calendar interviews. However, the mixed findings in the literature suggest that this relationship may be more multifaceted than it appears. The results in this dissertation indicate that there is a rather complex relationship among item non-response and survey context factors, which can be exposed via interviewer and respondent communication that is examined via interviewer and respondent behaviors. Hence, the main goal in this dissertation was to study verbal behaviors that occur in an interview as intervening factors to understand why interviewer experience and exposure is detrimental to item response rates. This goal is important because implications regarding how to improve interviewer training and monitoring in

both calendar and standardized interviews can be obtained from the examination of *the role of interviewer and respondent verbal behaviors* in interviewer experience, exposure, and item non-response relationship.

The main findings in this dissertation regarding the differential effect of interviewer experience and exposure on interviewer and respondent behaviors are mainly consistent with the literature (Cannell et al., 1977; Bradburn, Sudman, & Associates, 1979; Gfroerer et al., 2002; Olson & Peytchev, 2007; van der Zouwen, Dijkstra, & Smit, 1991). However, to my knowledge the current literature only examines standardized interviews, whereas this dissertation also focuses on calendar interviews. The findings reveal that in both calendar and standardized interviews ***inexperienced interviewers*** modify several of their behaviors (such as verbal behavior deviation and retrieval behaviors) in later interviews during the study fielding period. The findings also illustrate a modification among respondent behaviors (such as difficulty and non-cooperation) which are obtained by the ***inexperienced interviewers*** was observed in the later interviews. In contrast, in both calendar and standardized interviews, the increase in the use of behaviors among ***experienced interviewers*** is less pronounced—and not existent for most of the behaviors— throughout the study fielding period.

In sum, consistent with the expected findings, experienced interviewers do not modify or adapt their behaviors as much as inexperienced interviewers in the later interviews. One prominent theory in the literature that explains this finding is that experienced interviewers bring their own behavioral habits and traditions that were gained in earlier interviews and/or trainings throughout their interviewing careers (Cannell et al., 1977; Olson & Bilgen, 2011; van der Zouwen, Dijkstra, & Smit, 1991).

So experienced interviewers may be utilizing what has worked for them thus far throughout their interviewing careers; thus, they may be more reluctant to change their learned behaviors (or have too much learned knowledge that they cannot erase their learned and developed habits) even if they are assigned to use a different interviewing method than standardized interviewing (such as calendar interviewing).

One finding that is consistent with this theory but contradicts earlier findings (Bradburn et al., 1979; Gfroerer et al., 2002; Olson & Peytchev, 2007; van der Zouwen et al., 1991) is the finding regarding deviation from conventional ideals. Overall, ***experienced interviewers*** deviated less from conventional standardized interviewing ideals than inexperienced interviewers in both calendar and standardized interviews. As indicated earlier, general interviewer experience is specifically with standardized interviews; hence, the experienced interviewers may be simply following what they have been told to do throughout their careers (such as following a slower pace, avoiding directive probing, and not forgetting to ask a question/part of a question or probe to a "don't know" response). So then ***why is it noteworthy to examine whether interviewers with different levels of experience and exposure differ in their behaviors?*** This dissertation illustrates that the issue is not as simple as it seems.

One may argue that we may be able to predict which type of interviewers are "the good interviewers" via examining the behavioral differences among interviewers with diverse experience and exposure levels, given that some behaviors are detrimental and others are beneficial to data quality. However, Chapter VI findings indicate that some behaviors (such as deviation from conventional ideals) become detrimental when used by experienced interviewers BUT become beneficial when used by inexperienced

interviewers. Hence, the findings in this dissertation may suggest that the popular research question "which behaviors are detrimental or beneficial to data quality?" is likely a unidimensional view. Rather, we may need to take a multi-dimensional approach and ask ourselves ***"who uses which behaviors and how does this affect item non-response?"*** The findings in Chapter VI also illustrated that this question is more applicable in standardized interviews. This is because the item non-response rates in calendar interviews were less distressed by the differences in deviation behaviors among experienced and inexperienced interviewers. In other words, in calendar interviews differential use of deviation behaviors among interviewers with diverse experience levels did not relate to item non-response as much as in standardized interviews.

Chapter V findings reveal that in standardized interviews ***inexperienced interviewers*** obtained higher levels of respondent cooperation behaviors. Surprisingly, these interviewers also obtained higher levels of respondent difficulty and non-cooperation behaviors. One explanation for this finding is that in standardized interviews inexperienced interviewers may enable a more open communication; thus, may be getting higher levels of feedback from the respondents, regardless of the pleasantness, unpleasantness or difficulty of the interviewing situation. Thus, the "undesirable" respondent behaviors (such as respondent difficulty and non-cooperation) may be in fact desirable depending on the interviewing technique and interviewers' level of experience.

In addition, in calendar interviews inexperienced interviewers obtained lower levels of respondent difficulty and non-cooperation behaviors; however, they obtained higher levels of respondent cooperation behaviors than experienced interviewers. One explanation is that inexperienced calendar interviewers may put more emphasis on

respondent happiness and comfort during the interview (Belli, 2011 personal communication). Hence, the perceptional distinction between respondents' "good" (i.e., respondent cooperation) and "bad" (i.e., respondent difficulty and non-cooperation) behaviors may be more pronounced among inexperienced calendar interviewers than experienced calendar interviewers.

Nevertheless, findings regarding both calendar and standardized interviews may indicate that, trade-offs occur among behaviors when interviewers gain experience (Bilgen, Belli, & Olson, 2009). In a nutshell, while interviewers who have 1 year or more standardized interviewing experience put more emphasis on pace, reading questions as scripted, and avoidance of directive probing, interviewers with less than 1 year standardized interviewing experience overall put more emphases on interpersonal communication dynamics in both calendar and standardized interviews.

However, what are the implications of behavioral trade-offs on item nonresponse? Surprisingly, when it comes to interpersonal communication dynamics the trade-offs among interviewer behaviors do not play a role among interviewers with different experience levels or whether they are used in calendar or standardized interviews. Rather, interviewer exposure is what affects the relationship between interpersonal communication dynamics and item non-response. Interviewers who focus on establishing rapport throughout the study fielding period consistently obtain lower levels of item non-response, as they encourage respondents to think harder about their responses. However, interviewers who are less personable may be reluctant to establish a harmonious and enjoyable interviewing environment. Thus, they are more likely to accept respondents' unresolved "don't know" responses especially in later interviews throughout the study.

This indicates that the low usage of rapport behaviors becomes detrimental when used in later interviews (by interviewers who are exposed to the study) BUT becomes beneficial when used in the earlier interviews. This result suggests that interviewers may get better at providing rapport as they learn from their earlier experiences in earlier interviews. In other words, they may have determined how to interact with the questionnaire and the respondent simultaneously. Hence, these results may explain the equivocal findings in the literature regarding the effect of rapport behaviors on item non-response.

The findings in this dissertation indicate that each set of behaviors moderates the relationship between interviewer experience, exposure, and item non-response in different ways. The take-home message is that it may not be optimal to classify interviewer or respondent behaviors as "desirable" and "not desirable," because the association between behaviors and item non-response changes depending on when they are used (in early versus later interviews), who they are used by (experienced versus inexperienced interviewers), or which interviewing method are they used in (calendar versus standardized interviews).

Another possible explanation regarding higher levels of item non-response probabilities among experienced interviewers in comparison to inexperienced interviewers may be a differential interviewer attrition effect. There may be two different types of inexperienced interviewers. The first type consists of ambitious interviewers who tend to work very hard to obtain high item response via providing additional probes and responding to difficulties during the interview. The second type consists of efficient interviewers who do not necessarily put emphases on interview quality. Rather, they aim to get each interview done as quickly as possible. Hence, they have a faster pace, tend to

read questions as scripted, and avoid directive or additional probing when respondents are having trouble with the interview or when they provide "don't know" responses. The ambitious interviewers may eventually "burn out" quickly and discontinue their work as interviewers. Therefore, the experienced interviewers may largely consist of efficient interviewers. Another explanation as to why high interviewer experience relates to high respondent "don't knows" is that inexperienced interviewers may be working harder as they are still new and may feel the need to prove themselves to their supervisors and to the respondents. However, the more experienced interviewers may not feel as much pressure from their supervisors or the respondents, as they may think that they already have developed the skills, knowledge and practice the interviewing trade requires (McCutcheon, 2011 personal communication).

Lastly, another issue that needs to be addressed is how the item nonresponse measure used in this dissertation may potentially impact survey inferences and total survey error. In this dissertation, item non-response is measured via respondents' unresolved "don't know" responses. Then, the "so what?" question becomes whether unresolved DK responses are only an indication of item non-response rates, or an indication of both item non-response rates and item non-response bias. The unresolved "don't know" responses correspond to either missing spells (such as the year of health status change, the number of moves, whether there was a change in job status, number of changes in job status) or a missing element of a spell (such as a house number contained in an address). Hence, whether the incompleteness in the data set contributes to the error in survey estimates may be due to the type of items missing from the analyses (i.e. the information is missing regarding a spell or a specific element of a spell) and whether the

certain missing item is an important indicator of the reasons of missingness (i.e. whether the certain missing item is not missing at random).

In this dissertation, the item non-response measure captures all possible types of unresolved "don't know" behaviors. Therefore, it is not possible to disentangle the types of missingness and to draw conclusions regarding how the item nonresponse measure affects survey inferences. Hence, further research needs to explore which types of missingness (i.e. whether the information is missing regarding a spell or a specific element of a spell) captured in this dissertation only contribute to item non-response rates and which types of missingness contribute to both item non-response rates and bias. In other words, more research is necessary in order to understand which types of unresolved "don't know" behaviors contribute to error in survey estimates; hence, incorrect survey inferences.

What are the implications of these results for practice? First, in standardized interviews entirely separate training procedures may need to be developed for experienced and inexperienced interviewers, given that the mechanisms behind the verbal behavior and item non-response relationship differ among interviewers with different experience levels[75]. Moreover, experienced interviewers may not be reluctant to modify their already learned behaviors via standard basic interviewing training. Thus, several more focused training techniques may be developed through ***behavioral habit monitoring*** for experienced interviewers from the previous studies. This way, experienced interviewers with different behavioral habits (such as more personable interviewers versus more efficient interviewers) may be trained separately. However,

[75] For instance, the findings illustrated that the deviation behaviors are detrimental to item non-response when they are used by experienced interviewers BUT become beneficial when inexperienced interviewers are using these behaviors.

monitoring and collecting interviewer behavioral habits from the earlier studies can be costly; thus, researchers may need to analyze the cost-benefit relationship before these suggestions are applied to practice.

Second, in calendar interviews the results illustrate that the impact of some of the behavioral differences among experienced and inexperienced interviewers on item non-response are more similar[76] than in standardized interviews. Hence, developing separate training techniques for interviewers with different experience levels may not be necessary in calendar interviews especially for the behaviors that are a part of the standard basic training. However, the behaviors that are not a part of the basic training (such as, retrieval probing) may need to be taught and practiced separately among interviewers with different experience levels in both calendar and standardized interviews.

Third, in both calendar and standardized interviews, the effect of behavioral changes on item non-response throughout the study fielding period suggests constant monitoring throughout the study, given that the interviewer and respondent behavioral patterns change in later studies when interviewers are more exposed to the study and the questionnaire. Again, the ***behavioral habit monitoring*** may indicate that while some interviewers need re-training throughout the fielding period, others may not need such training according to the behavioral differences in later interviews. However, again as mentioned earlier, constant monitoring and training may be costly; hence, the cost-benefit relationship may need to be taken into consideration. Also, as item non-response and measurement error are likely to be correlated with each other, another concern is that

[76] For instance, the differences in the use of deviation behaviors among interviewers with diverse experience levels have a more similar impact on item non-response than in standardized interview. This is not the case for interviewer retrieval probing. The behavioral differences among experienced and inexperienced interviewers on item non-response systematically differ in both standardized and calendar interviews.

some of the changes that are detrimental to item non-response may be beneficial for reducing measurement error. In other words, both item non-response and measurement error needs to be taken into account when new training and monitoring techniques are developed and implicated. Hence, future research is necessary to explore the intervening role of interviewer and respondent behaviors on the relationships among survey context factors and measurement error in calendar and standardized interviews. The following section will discuss several other limitations regarding this dissertation and focus on future research goals, which may potentially overcome some of the limitations of this dissertation.

3. Limitations and Future Direction of the Dissertation

There are several limitations of this study. One of the major limitations is the age range of the respondents. In this study, the age of the respondents ranges between 46 and 98 due to the respondent selection criteria (as explained earlier in the data analysis plan). In addition, these are all panel respondents who have been PSID[77] respondents for a relatively long period of time; therefore, they may be more motivated and knowledgeable about surveys than sampled individuals in a general U.S. population study. Thus, the results cannot be generalized to the whole U.S. population. Further research is necessary to describe a more representative portion of the population.

Another limitation is that even though this is an experimental design, as the experiment was not geared towards exploring interviewer experience and exposure effects, there is a lack of interpenetration. The interpenetrated design method was developed by Mahalanobis (1946) and assigns households or respondents at random to

[77] PSID: Panel Study of Income Dynamics

interviewers (O'Muircheartaigh & Campanelli, 1999) in order to measure interviewer variance and separate the effects of the interviewer from the effects of other sources (such as regions). Due to the lack of interpenetration, neither general interviewer experience nor the within-study interviewer exposure is randomly assigned to respondents during the data collection process. Therefore, I controlled for the available interviewer and respondent characteristics such as age, sex, and race during the analyses stage of this dissertation to take into consideration the potential interviewer and respondent confounding effects. However, future research is necessary to replicate these findings with an interpenetrated study to disentangle other interviewer and respondent effects (such as interviewer and respondent education, socio-economic status, so on and so forth) from the effects of geographical distribution.

Also in this study, the general interviewer experience is identified as interviewers' lifetime experience with standardized interviews in both calendar and standardized interviews. As the interviewers have very little or no experience with calendar interviews, the interviewer experience with calendar interviews information cannot be used in this dissertation. Therefore, for both the standardized and calendar conditions, all the proposed hypotheses and theories are designed to explore general interviewer experience with standardized interviews. Future research is necessary to explore how interviewer and respondent behaviors differ among interviewers with differential general calendar experience and how these differential behaviors relate to respondents' behaviors and data quality. However, it is not the scope of my dissertation to explore these questions, given that I do not have sufficient data to do so.

Lastly, item nonresponse models, illustrated in Chapter VI, provided small but significant effects. These small effects may have occurred because item non-response measure is a turn-specific variable (rather than item-specific that occurs at a higher level). This may cause an under-estimation in the item non-response probabilities. As indicated earlier, further research is needed to examine the advantages and disadvantages to use turn as the unit of analyses.

REFERENCES

Axinn, W. G., Pearce, L. D., & Ghimire, D. (1999). Innovations in life history calendar applications. *Social Science Research, 28*(3), 243-264.

Bailar, B., Bailey, L., & Stevens, J. (1977). Measures of interviewer bias and variance. *Journal of Marketing Research, 14*(3, Special Issue: Recent Developments in Survey Research), 337-343.

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports, 19*(1), 3-11.

Beatty, P. (1995). Understanding the standardized/non-standardized interviewing controversy. *Journal of Official Statistics, 11*(2), 147-160.

Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (). New York: John Wiley & Sons.

Belli, R. F., & Lepkowski, J. M. (1996). Behavior of survey actors and the accuracy of response. *Health Survey Research Methods: Conference Proceedings,* (DHHS Publication No. (PHS) 96-1013) 69-74.

Belli, R. F. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in survey. *Memory, 6*, 383-406.

Belli, R. F. (2000). Computerized event history calendar methods: Facilitating autobiographical recall. *American Statistical Association Proceedings of the Section on Survey Research Methods,* Alexandria, VA. 471-475.

Belli, R. F., Alwin, D. F., & Stafford, F. P. (2009). Introduction: The application of calendar and time diary methods in the collection of life course data. In R. F. Belli, F. P. Stafford & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (, pp. 1-10). Thousands Oaks, CA: Sage.

Belli, R. F., & Callegaro, M. (2009). The emergence of calendar interviewing: A theoretical and empirical rationale. In R. F. Belli, F. P. Stafford & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (, pp. 31-52). Thousands Oaks, CA: Sage.

Belli, R. F., Lee, E. H., Stafford, F. P., & Chou, C. (2004). Calendar and question-list survey methods: Association between interviewer behaviors and data quality. *Journal of Official Statistics, 20*(2), 185-218.

Belli, R. F., Lepkowski, J. M., & Kabeto, M. U. (2001). The respective roles of cognitive processing difficulty and conversational rapport on the accuracy of retrospective reports of doctor's office visits. In M. L. Cynamon, & R. A. Kulka (Eds.), *Seventh conference on health survey research methods* (, pp. 197-203). Hyattsville, MD: U.S. Government Printing Office: DHHS Publication No. (PHS) 01-1013.

Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event history calendars and question list surveys. *Public Opinion Quarterly, 65*(1), 45-74.

Belli, R. F., Smith, L. M., Andreski, P. M., & Agrawal, S. (2007). Methodological comparisons between CATI event history calendar and standardized conventional questionnaire instruments. *Public Opinion Quarterly, 71*(4), 603-622.

Berk, M. L., & Bernstein, A. B. (1988). Interviewer characteristics and performance on a complex health survey. *Social Science Research, 17*(3), 239-251.

Biemer, P. P., & Lyberg, L. (2003). *Introduction to survey quality*. Hoboken, NJ: Wiley.

Bilgen, I., & Belli, R. F. (2010a). Impact of interviewer and respondent verbal behaviors on measurement error in different interviewing techniques. *65th Annual Conference of American Association for Public Opinion Research,* Chicago, IL.

Bilgen, I., & Belli, R. F. (2010b). Comparison of verbal behaviors between calendar and standardized conventional questionnaires. *Journal of Official Statistics, 26*(3), 481-505.

Bilgen, I., Belli, R. F., & Olson, K. (2009). Impact of within-study interviewer experience in different interviewing techniques. *64th Annual Conference of American Association for Public Opinion Research,* Hollywood, Florida.

Booker, H. S., & David, S. T. (1952). Differences in results obtained by experienced and inexperienced interviewers. *Journal of the Royal Statistical Society Series A (General), 115*(2), pp. 232-257.

Borland, D. M. (1975). An alternative model of the wheel theory. *The Family Coordinator, 24*(3), pp. 289-292.

Boyd, H. W. J., & Westfall, R. (1955). Interviewers as a source of error in surveys. *The Journal of Marketing, 19*(4), pp. 311-324.

Bradburn, N. M., Sudman, S., & Associates. (1979). *Improving interview method and questionnaire design* (1st ed.). San Francisco: Jossey-Bass.

Brenner, M. (1982). Response-effects of "role-restricted" characteristics of the interviewer. In W. Dijkstra, & J. van der Zouwen (Eds.), *Response behaviour in the survey-interview* (, pp. 131-165). London: Academic Press.

Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research, 5*(3), 215-238.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Callegaro, M., Belli, R. F., Serrano, E., & Palmer, D. (2007). Cultural variability in event history calendar and conventional questionnaire interviews: A verbal behavior analysis. *Proceedings of the Survey Research Methods Section, American Statistical Association,* 4058-4065.

Cannell, C. F., Fowler, F. J., & Marquis, K. H. (1968). The influence of interviewer and respondent psychological and behavioral variables in the reporting in household interviews. *Vital and Health Statistics, 26*(2), i-65.

Cannell, C. F., & Kahn, R. L. (1968). Interviewing. In G. L. Lindzey, & E. Aronson (Eds.), *The handbook of social psychology* (, pp. 526-595). Reading, MA: Addison Wesley.

Cannell, C. F., Marquis, K. H., & Laurent, A. (1977). A summary of studies of interviewing methodology: 1959-1970. *Vital and Health Statistics, 69*(2), i-78.

Cannell, C. F., & Oksenberg, L. (1988). Observation of behavior in telephone interviews. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls & J. Waksberg (Eds.), *Telephone survey methodology* (, pp. 475-495). New York: Wiley.

Cannell, C. F., Lawson, S. A. and Hausser, D. L. (1975). *A technique for evaluating interviewer performance: A manual for coding and analyzing interviewer behavior from tape recordings of household interviews.* [Video/DVD] Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan.

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 12*, pp. 389-437.

Chesnut, J. (2005). Item nonresponse error for the 100 percent data items on the census 2000 long form questionnaire. *Proceedings of the Survey Research Methods Section, American Statistical Association (2005),* Minneapolis, Minnesota. 2857-2864.

Chromy, J. R., Eyerman, J., Odom, D., McNeeley, M. E., & Hughes, A. (2005). *Association between interviewer experience and substance use prevalence rates in NSDUH.* Washington. DC: Substance Abuse and Mental Health Services.

Cleary, P. D., Mechanic, D., & Weiss, N. (1981). The effect of interviewer characteristics on responses to a mental health interview. *Journal of Health and Social Behavior, 22*(2), 183-193.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Collins, R., Lincoln, R., & Frank, M. G. (2002). *The effect of rapport in forensic interviewing* ePublications@bond.

Conrad, F. G., Schober, M. F., & Dijkstra, W. (2008). Cues of communication difficulty in telephone interviews. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (, pp. 212-230). New Jersey: Wiley.

Conrad, F. G., & Blair, J. (2004). Aspects of data quality in cognitive interviews: The case of verbal reports. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (, pp. 67-88). New York: Wiley.

Conrad, F. G., & Schober, M. F. (2000). Clarifying questions meaning in a household telephone survey. *Public Opinion Quarterly, 64*(1), 1-28.

Converse, J. M. (1977). Predicting no opinion in the polls. *Public Opinion Quarterly, 40*(4), 515.

Davis, D. W., & Silver, B. D. (2003). Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science, 47*(1), 33-45.

De Leeuw, E. D., Hox, J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics, 19*(2), 153-176.

Dijkstra, W. (1987). Interviewing style and respondent behavior. *Sociological Methods & Research, 16*(2), 309-334.

Dijkstra, W., Smit, J., & Ongena, Y. (2009). An evaluation study of the event history calendar. In R. F. Belli, F. P. Stafford & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (, pp. 257-275). Thousand Oaks, CA: Sage.

Dijkstra, W., & van der Zouwen, J. (1988). Types of inadequate interviewer behavior in survey interviews. In W. E. Saris, & I. N. Gallhofer (Eds.), *Data collection and scaling* (, pp. 24-35). New York: St. Martin's.

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

Drennan, J. (2003). Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing, 42*(1), 57-63. doi:10.1046/j.1365-2648.2003.02579.x

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly, 74*(1), 1-36.

Dykema, J., Lepkowski, J., & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. E. Lyberg, P. Biemer, M. Collins, E. D. De Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey measurement and process quality* (, pp. 311-330). New York: Wiley.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers.

Eyerman, J., Odom, D., Wu, S., & Butler, D. (2002). Nonresponse in the 1999 NHSDA. In J. Gfroerer, J. Eyerman & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (, pp. 23-51). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Ferber, R. (1966). Item nonresponse in a consumer survey. *The Public Opinion Quarterly, 30*(3), pp. 399-415.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378-382. doi:DOI: 10.1037/h0031619

Fowler, F. J. J. (1991). Reducing interviewer-related error through interviewer training, supervision, and other means. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz & S. Sudman (Eds.), *Measurement errors in surveys* (, pp. 259-278). New York: Wiley.

Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz, & S. S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (, pp. 15-36). San Francisco: Jossey-Bass.

Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Thousand Oaks: Sage Publications.

Fowler, F. J. J., & Mangione, T. W. (1986). *Reducing interviewer effects on health survey data*. Washington, DC: National Center for Health Services Research and Health Care Technology Assessment.

Fowler, F. J. (1992). How unclear terms affect survey data. *Public Opinion Quarterly, 56*(2), 218-231. doi: 10.1086/269312

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing; minimizing interviewer- related error*. Newbury Park, Calif.: Sage Publications.

Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988). The life history calendar: A technique for collecting retrospective data. In C. C. Clogg (Ed.), *Sociological methodology* (, pp. 37-68). San Francisco: Jossey-Bass.

Gales, K., & Kendall, M. G. (1957). An inquiry concerning interviewer variability. *Journal of the Royal Statistical Society Series A (General), 120*(2), pp. 121-147.

Gfroerer, J. C., Eyerman, J., & Chromy, J. R. (2002). *Redesigning an ongoing national household survey*. Rockville, MD: Dept. of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Goudy, W. J., & Potter, H. R. (1975-76). Interview rapport: Demise of a concept. *The Public Opinion Quarterly, 39*(4), 529-543.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J., M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, New Jersey: John Wiley & Sons Inc.

Groves, R. M., & Couper, M. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

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

Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *The Public Opinion Quarterly, 50*(2), 251-266.

Hedeker, D. (2005). Generalized linear mixed models. In B. Everitt, & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science*. New York: John Wiley & Sons.

Henson, R., Cannell, C. F., & Lawson, S. (1976). Effects of interviewer style on quality of reporting in a survey interview. *Journal of Psychology, 93*(2), 221.

Hill, R. J., & Hall, N. F. (1963). A note on rapport and the quality of interview data. *Southwestern Social Science Quarterly, 44*(3), 247-255.

Houtkoop-Steenstra, H. (1995). Meeting both ends: Between standardization and Recipient design in telephone survey interviews. In P. Ten Have, & G. Psathas (Eds.), *Situated order: Studies in the social organization of talk and embodied activities* (, pp. 91-106). Washington, DC: International Institute for Ethnomethodology and Conversation Analysis, and University Press of America.

Houtkoop-Steenstra, H. (1996). Probing behaviour of interviewers in the standardised semi-open research interview. *Quality and Quantity, 30*(2), 205-230.

Houtkoop-Steenstra, H. (2000). *Interaction and the standardized survey interview: The living questionnaire*. Cambridge; New York: Cambridge University Press.

Hox, J., & De Leeuw, E. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (, pp. 103-120). New York: Wiley.

Hox, J. J. (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological Methods & Research, 22*(3), 300-318. doi:10.1177/0049124194022003002

Hughes, A., Chromy, J., Giacoletti, K., & Odom, D. (2001). Impact of interviewer experience on drug use prevalence rates in the 1999 NHSDA. *Proceedings of the Survey Research Methods Section, American Statistical Association (2001),* Alexandria, VA.

Hughes, A., Chromy, J., & Katherine Giacoletti, and Dawn Odom. (2002). Impact of interviewer experience on respondent reports of substance use. In J. Gfroerer, J. Eyerman & J. Chromy (Eds.), *Redesigning an ongoing national household survey*. Washington, DC: Substance Abuse and Mental Health Services Administration.

Huisman, M., & Van der Zouwen, J. (1999). Item nonresponse in scale data from surveys: Types, determinants, and measures. In M. Huisman (Ed.), *Item nonresponse: Occurrence, causes, and imputation of missing answers to test items* (, pp. 63-90). Leiden: DSWO-press.

Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality & Quantity, 34*(4), 331.

Hurd, M., & Rohwedder, S. (2009). Time-use in the older population: Variation by socio-economic status and health. In R. F. Belli, F. P. Stafford & D. F. Alwin (Eds.), *Calendar and time diary: Methods in life course research* (, pp. 207-224). Los Angeles: Sage.

Hyman, H. H., Cobb, W. J., Feldman, J. J., Hart, C. W., & Stember, C. H. ([1954] 1975). *Interviewing in social research*. Chicago: University of Chicago Press.

Kahn, R. L., & Cannell, C. F. (1957). *The dynamics of interviewing; theory, technique, and cases*. New York: Wiley.

Kennickell, A. B. (2002). Interviewers and data quality: Evidence from the 2001 survey of consumer finances. *Proceedings of the Section on Survey Research Methods,* Annual Meetings of the American Statistical Association, New York, NY.

Kish, L. (1965). *Survey sampling*. New York: J. Wiley & Sons.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (Second Edition ed.). New York: Guilford Press.

Krosnick, J. A. (2002). The causes of no-opinion responses to attitude measures in surveys: They are rarely what they appear to be. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (). New York: Wiley.

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

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C., & Conaway, M. (2002). The impact of "no opinion" response options on data quality. *Public Opinion Quarterly, 66*(3), 371-403.

Lavin, D., & Maynard, D. W. (2001). Standardization vs. rapport: Respondent laughter and interviewer reaction during telephone surveys. *American Sociological Review, 66*(3), 453-479.

Lepkowski, J. M., Siu, V., & Fisher, J. (2000). Event history analysis of inter- viewer and respondent survey behavior. In *Developments in survey methodology* (Metodološki zvezki 18 ed., pp. 3-20). Ljubljana: FDV.

Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. New York: Wiley.

Littell, R. C. (2006). *SAS for mixed models* (2nd ed.). Cary, N.C.: SAS Institute.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Mahalanobis, P. C. (1946). Recent experiments in statistical sampling in the Indian statistical institute. *Journal of the Royal Statistical Society, 109*(4), 326-378.

Mangione, T., Fowler, J., & Louis, T. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics, 8*(3), 293-307.

Marker, D. A., Judkins, D. R., & Winglee, M. (2002). Large-scale imputation for complex surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (, pp. 329-341). New York: Wiley.

Marquis, K. H. (1971). Effects of race, residence, and selection of respondent on the conduct of the interview. In J. Lansing, S. Withey & A. Wolfe (Eds.), *Working papers on survey research in poverty areas*. Ann Arbor, MI: Institute for Social Research.

Mathiowetz, N., & Cannell, C. (1980). Coding interviewer behavior as a method of evaluating performance. *Proceedings of the Section on Survey Research Methods,* 525-528.

Maynard, D. W., & Schaeffer, N. C. (2002). Standardization and its contents. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (, pp. 3-45). New York: Wiley.

Means, B., Swan, G. E., Jobe, J. B., & Esposito, J. L. (1991). An alternative approach to obtaining personal history data. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz & S. Seymour (Eds.), *Measurement errors in surveys* (, pp. 167-183). New York: Wiley.

Means, B., & Loftus, E. F. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology, 5*(4), 297-318.

Merkle, D. M., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (, pp. 243-258). New York: Wiley.

Messmer, D. J., & Seymour, D. T. (1982). The effects of branching on item nonresponse. *The Public Opinion Quarterly, 46*(2), pp. 270-277.

Morton-Williams, J. (1979). The use of "verbal interaction coding" for evaluating a questionnaire. *Quality & Quantity, 13*(1), 59.

Morton-Williams, J., & Sykes, W. (1984). The use of interaction coding and follow-up interviews to investigate comprehension of survey questions. *Journal of the Market Research Society, 26*, 109-127.

Neisser, U., & Hyman, I. E. (2000). *Memory observed: Remembering in natural contexts* (2nd edition ed.). New York: Worth Publishers.

Oksenberg, L., Cannell, C. F., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics, 7*(3), 349-365.

Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly, 75*(1), 99-114.

Olson, K., & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly, 71*(2), 273-286. doi:10.1093/poq/nfm007

O'Muircheartaigh, C. A. (1977). Response errors. In C. A. O'Muircheartaigh, & C. Payne (Eds.), *The Analyses of Survey Data* (, pp. 193-235). New York: Wiley.

O'Muircheartaigh, C. (1997). Measurement error in surveys: A historical perspective. In L. Lyberg, P. P. Biemer, M. Collins, E. D. De Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey measurement and process quality* (, pp. 1-25). New York: Wiley.

O'Muircheartaigh, C., & Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society Series A (Statistics in Society), 162*(3), pp. 437-446.

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society Series A (Statistics in Society), 161*(1), 63-77.

Peeks, A. (2005). *"Can I change your depression?" Understanding the role of interviewer effects.* Unpublished manuscript.

Pickery, J., & Loosveldt, G. (1998). The impact of respondent and interviewer characteristics on the number of "no opinion" answers: A multilevel model for count data. *Quality & Quantity, 32*(1), 31.

Pickery, J., & Loosveldt, G. (2001). An exploration of question characteristics that mediate interviewer effects on item nonresponse. *Journal of Official Statistics, 17*(3), 337-350.

Pickery, J., Loosveldt, G., & Carton, A. (2001). The effects of interviewer and respondent characteristics on response behavior in panel surveys. *Sociological Methods & Research, 29*(4), 509.

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? *Sociological Methodology, 24*, pp. 73-104.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly, 68*, 109-130.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd edition ed.). Newbury Park, CA: Sage.

Roberts, J., & Mulvey, E. P. (2009). Reports of life events by individuals at high risk violence. In R. F. Belli, F. P. Stafford & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (, pp. 191-206). Thousands Oaks, CA: Sage.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91*(434), pp. 473-489.

Sanchez, M. E. (1992). Effects of questionnaire design on the quality of survey data. *Public Opinion Quarterly, 56*(2), 206-217.

Sayles, H., Belli, R. F., & Serrano, E. (2010). Interviewer variance between event history calendar and conventional questionnaire interviews. *Public Opinion Quarterly, 74*(1), 140-153.

Schaeffer, N. C. (1991). Conversation with a purpose or conversation? interaction in the standardized interview. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz & S. Sudman (Eds.), *Measurement errors in surveys* (, pp. 367-391). New York: Wiley.

Schaeffer, N. C., & Maynard, D. W. (1996). From paradigm to prototype and back again: Interactive aspects of cognitive processing in survey interviews. In N. Schwarz, & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (, pp. 65-88). San Francisco: Jossey-Bass.

Schaeffer, N. C., & Maynard, D. (2001). Standardization and interaction in the survey interview. In J. F. Gubrium, & J. A. Holstein (Eds.), *Handbook of interview research: Context & method* (, pp. 577-601). Thousand Oaks, Calif.: Sage Publications.

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61*(4), 576.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys* (New 1996 copyright by Sage Publications, Thousand Oaks, CA. ed.). New York: Academic Press.

Schwarz, N. (1996). *Cognition and communication :Judgmental biases, research methods, and the logic of conversation*. Mahwah, N.J.: L. Erlbaum Associates.

Shoemaker, P. J., Eichholz, M., & Skewes, E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research, 14*(2), 193-201.

Singer, E., Frankel, M. R., & Glassman, M. B. (1983). The effect of interviewer characteristics and expectations on response. *The Public Opinion Quarterly, 47*(1), 68-83.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, Calif.; London: Sage.

Stember, H., & Hyman, H. (1949). Interviewer effects in the classification of responses. *The Public Opinion Quarterly, 13*(4), pp. 669-682.

Suchman, L., & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85*(409), pp. 232-241.

Sudman, S., Bradburn, N. M., Blair, E., & Stocking, C. (1977). Modest expectations: The effects of interviewers' prior expectations on responses. *Sociological Methods & Research, 6*(2), 177.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers :The application of cognitive processes to survey methodology* (1st ed.). San Francisco: Jossey-Bass Publishers.

Su-Hao Tu, & Pei-Shan Liao. (2007). Social distance, respondent cooperation and item nonresponse in sex survey. *Quality & Quantity, 41*(2), 177-199.

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

Tourangeau, R., Shapiro, G., Kearney, A., & Ernst, L. (1997). Who lives here? Survey under-coverage and household roster questions. *Journal of Official Statistics, 13*(1), 1-18.

Turner, C. F., Lessler, J. T., & Devore, J. W. (1992). Effects of mode of administration and wording on reporting of drug use. In C. F. Turner, J. T. Lessler & J. C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (Department of Health and Human Services Publication no. (ADM) 92-1929 ed., pp. 177-220). Washington, DC: Government Printing Office.

Van der Zouwen, J. (1974). A conceptual model for the auxiliary hypotheses behind the interview. *Annals of Systems Research, 4*, 21-37.

Van der Zouwen, J. (2002). Why study interaction in the survey interview? response from a survey researcher. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (, pp. 47-65). New York: John Wiley.

Van der Zouwen, J. (2006). Interviewer and survey researcher: Mutual dependencies. *Bulletin De Méthodologie Sociologique, 89*, 49-64.

Van der Zouwen, J., & Dijkstra, W. (1995). Trivial and non-trivial question-answer sequences; types, determinants and effects on data quality. *Proceedings of the International Conference on Survey Measurement and Process Quality,* Bristol (UK). 81-86.

Van der Zouwen, J., Dijkstra, W., & Smit, J. H. (1991). Studying respondent-interviewer interaction: The relationship between interviewing style, interviewer behavior, and

response behavior. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz & S. Sudman (Eds.), *Measurement errors in surveys* (, pp. 419-438). New York: Wiley.

Weiss, C. H. (1968-69). Validity of welfare mothers' interview responses. *The Public Opinion Quarterly, 32*(4), 622-633.

Williams, J. A., Jr. (1968). Interviewer role performance: A further note on bias in the information interview. *The Public Opinion Quarterly, 32*(2), pp. 287-294.

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

Yoshihama, M. (2009). Application of the life history calendar approach: Understanding women's experiences of intimate partner violence over the life course In R. F. Belli, F. P. Stafford & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (, pp. 135-156). Thousands Oaks, CA: Sage.

Yoshihama, M., Gillespie, B., Hammock, A. C., Belli, R. F., & Tolman, R. M. (2005). Does the life history calendar method facilitate the recall of intimate partner violence? Comparison of two methods of data collection. *Social Work Research, 29*(3), 151-163.

APPENDICES

Appendix A: Coding Scheme that includes New Verbal Behavior Codes (Verbal Behaviors, Definitions, Examples and Coding Rules)

ITEM NON-RESPONSE COMPONENTS	
Resolved Don't Know Responses (RDK)* – *This is not used in the study, coders use this code in order to keep track of URDK*	Respondent indicates that he/she does not know the answer to the question asked by the interviewer. In this case, "resolved don't know" response is coded when the interviewer obtains an answer after probing for an answer or the respondent volunteers an answer later in the interview. *Example: I: And did you ever live in a different place after address4? R: Yes. In 1935 we moved to what they call city2. I: Uh, do you remember the month? R: I don't remember the month. (Okay.) It was 1935's the best I can do. It was in the fall, but – why – don't you put down September. I don't know what it… I: Okay… (Resolved DK, Guess)*
Unresolved DK Responses (URDK)	Respondent indicates that he/she does not know the answer to the question asked by the interviewer. In this case, an "unresolved don't know" response is coded when the interviewer accepts the answer and fails to probe, or fails to obtain an answer after probing for an answer. *Example: R: Uh… see I don't even remember the year when I broke all my ankle bones. Because then I got a metal plate and 2 screws in my hip. I: Wow, hmm, that sounds painful. R: All on the same side. I: M-pos. Yeah… R: <u>I don't remember the year</u> though. I: Let's just go ahead skip to the next one then. Did you ever smoke?*
RESPONDENT COOPERATION-RELATED BEHAVIORS:	
Respondent Cooperative Behaviors	Respondent provides a spontaneous study-related reaction, which indicates that he/she is interested in the interview. • Respondent joking, sarcasm (RJS). Respondent jokes or provides a sarcastic comment both regarding **a study-related or un-related** topic. *Example: I: Oops. Zipcode1. New sheet. Just one minute. My computer and I are not getting along right at this second. R: <u>Well, straighten it up.</u>* • Respondent empathy (RE): Respondent feels concern for the interviewer and tries to share and/or understand how the interviewer will deal with a study-related or task-related difficulty. This also includes implicit or explicit indications of understanding regarding interviewer's expressions of interviewing difficulties or reciprocations towards interviewers' non-study related personal experiences.

Example: R: So, that's, uh—how you're going to record that, I don't know.

- Attempt to resolve difficulty (ARD): Respondent attempts to resolve a study-related or task-related difficulty during the interview. Code this when respondent implicitly or explicitly offers help during the interview regarding resolving a technical or study-related difficulty such as respondents' mentions of an external validation (mentioning they need to ask their spouse; look up their external records such as tax refunds, diary, address book, etc…) or change their location to use a better working phone or go to a less noisy room. *Example: I: Well, we got a lot of static, don't we? R: There is a lot. Let me try another phone. (Okay.) I don't know if this phone, uh- just a minute. (Okay.) I'm going to pick up the other phone, then come back and put this one down. I think this will help me out.*
  Example: R: I'd have to go through--back through all my tax returns and stuff.
- Respondent offers/provides clarification (ROPC): Respondent **spontaneously** offers or provides clarification on any aspect of the study-related information she/he provided earlier. DO NOT code this if it is provided after clarification is asked by the interviewer. DO code this if respondent provides new information related to any aspect of the study without digressing. Also, DO code it when respondent is explaining why he/she cannot provide a response or having trouble remembering. ROPC can be also used when respondents use information from other domains in order to help them remember what is asked in a particular domain. *Example: R: I was working full time. Just to explain why I'm laughing, uh, these are, um, 18 to 20 hour days. I'm trying to operate, uh, tourist cabins.*
- Respondent correction (RC): Respondent corrects an earlier response. *Example: R: Well, this is crazy, but I worked for employer1 in country1. (Oh, okay.) 1939. September, 1939. I: I'm sorry. September, 1939? R: No, wait a minute. (Oh.) Uh, it was June of 1939.*
- Respondent corrects interviewer (RCI): Respondents corrects an interviewer study-related comment or assumption. *RCI is coded when respondent does not spontaneously provide the correct response but still indicates whether interviewer's comment or assumption was wrong by saying "No/ That's not correct/uh-neg/M-neg" OR when respondent provides a different response than what interviewer had assumed or provided. Example: I: 1, 2, 3, 4, 5, 6, 7, okay, so that's--you went to elementary, and then you*

went to middle school? R: No, high school.

- Respondent laughter to an interviewer joke, comment or feedback (RL).
- Positive regard for the interview or questionnaire (PRIQ): Respondent indicates that the interview /questionnaire is enjoyable or interesting. *Example: R: This is going to be an interesting interview.*
- Request for question repeat (RQR)*: Respondent requests that a question be repeated. *Example: Did you say—I—I couldn't quite hear the question.*
- Attempt to decrease pace (ADP)*: Respondent is expressing a need for a pace decrease during the interview. This code includes verbal cues such as "hold on a second, let me think about this" as well as implicit/unconscious attempts to decrease the pace. Implicit attempts include time elapses, reflective questions (i.e., respondent repeating the question asked by the interviewer) and respondents repetitions of two or more words in the same turn. *Example: I: Thinking back to your early childhood from birth until you reached the age of seven, would you say that you were very overweight, slightly overweight, just right, slightly underweight, or very under-weight? R: No, about—about--for my size--for my size, yeah--about, uh, regular size.*

Respondent Non-Cooperative Behaviors

Respondent provides a spontaneous study-related reaction, which indicates that he/she is not interested in the interview.

- Negative comment about the interview/task/ questionnaire (RNC): Respondent implicitly or explicitly indicates that the interview/question(/naire) is long, complicated, foolish, boring, repetitious, inappropriate etc... Also, these behaviors include respondents' expressions of anger, frustration, concern, stress, doubts or criticisms regarding the interview/question(/naire) content or their task. *Example: I: Okay, and just a minute. Let me get this in. 1942. R: Uh, I worked there for about a year, and then I went to a place called employer2. I: Oh, wait a minute. I just [Unintelligible]. R: I think this is foolish. I: Laugh-I. Well, I'm sorry. Example: I: Let's go to employment now. Um, they're asking for jobs that you have held since you were 14 years old. R: Okay. I mean, it would be a lot easier to go from present back, but we'll try.*
- Respondent uncomfortable laughter (RUL). Respondent is using laughter to deal with an uncomfortable situation such as expressions of the difficulty of the question/interview or during answering uncomfortable/sensitive questions or when they provide a comment, which can potentially increase the

tension between the actors. RUL is coded when respondent laughs during expressions of difficulty (RED) OR when they provide a negative comment (RNC) OR during talking about an uncomfortable or sensitive topic (such as weight, death of a relative, a topic of sorrow, etc…)
Example: I: Uh, any idea what the zip code was there? R: No, I wouldn't know that either. Laugh-R.

- Reluctant to provide information (RPI): Respondent is reluctant to provide study-related information. This includes respondent refusals to answer a specific question (e.g., "I won't answer this question") and respondents' indications of not being able to answer the question without thinking through a response (e.g., "I can't answer that"). *Example: Now—now, you—do you want street adder—addresses or cities? Because I'm not going to be able to give you many street addresses.*

Respondent Expressions of Difficulty

- Respondent seeks clarification (RSC): The respondent asks interviewer to clarify an aspect of the study or a question. DO NOT code this if respondent requests that a question be repeated. *Example: I: Oh, sure. Um, from February, 1952 until June, 1977, did you ever have a different main job than working for employer1? R: In other words, while I was working for employer1, did I get different jobs with the company? (Um--.) Is that--.*
- Respondent is expressing difficulty answering a question (RED): These include both explicit and implicit expressions of difficulty. In explicit REDs respondent verbalizes that he/she is having difficulty with the survey question(/s). Implicit RED includes expressions of frustration due to difficulty in remembering events (e.g., oh shoot, this occurred before my recollection; I was so young to remember these; that was so long ago). *Example: Uh, we're going to run into a little difficulty here. I hope you have experience with this because I don't know how to handle it. Example: Yes, I have—I haven't thought of these questions in 30 years.*
- Uncertainty: Respondent indicates uncertainty about his/her response, including indications that the answer is a guess or an estimate. Uncertainty codes are also coded if the respondent is uncertain in a clarifying comment. Please note that, uncertainty is not coded if a respondent appears uncertain about an off topic response. Uncertainty includes guess and approximation:

  Guess (G): Respondent is providing an answer expressing he/she does not have the sufficient information to be sure of

	being correct. If the respondent uses words such as **"maybe, I would say, I suppose, probably, I think, I guess, let's say, I believe, I believe so, I am pretty sure"** while providing a response then code this as guess. *Example: I: And did [your daughter] always live you – live with you until she was 18 or was there a period for 4 months or more when she didn't? R: She – she – she lived with us until she was married, and, uh, I think she was 18, yeah. I: Okay.*
	Approximation/Estimation (E): Respondent is providing an answer expressing his/her response is close to the actual, but not completely accurate or exact. If the respondent uses words such as **"about, around, between"** while providing a response then code this as approximation or estimate. *Example: I: Okay, and uh, you wouldn't have any idea of the year? R: Well, It – it must've been **about** 1925.*
INTERVIEWER DEVIATIONS FROM CONVENTIONAL IDEALS	
Interviewer failure to probe (IFP)	Interviewer fails to ask a question that is provided in the questionnaire (see provided scripts for the actual questions). This behavior DOES NOT include interviewer failure to probe after a "don't know" response, which results in an unresolved DK response. *Example: R: Well, I had an accident in 1951. I: Ok, Let me write than in. Would you say that before then would you say it was very good? R: Very good. I: Okay, then in 1951 things changed? R: Yes. R: I appreciate the Lord for letting me live this long, cause lots of people thought I'm gonna die lots of times, but the Lord spared my life. I: That's wonderful. Okay, have you ever smoked cigarettes?* (In this example, the interviewer failed to probe regarding respondent's health status change from 1951 until the year interview took place (2002). Within this 51-year period, respondent's health might have changed again).
Interviewer failure to probe after a DK response (IFPDK)	Interviewer fails to probe after a "don't know" response (i.e., accepts the "don't know" response), which results in an unresolved DK response. *Example: R: All I can tell you is it was on street1. I don't know the address of the house. I: Okay, that is fine. I am sorry, what city did you say that was in?*

	Failure to probe for the detailed address information after a DK response (FPDA): If the respondent provides a "don't know" response about specific piece of his/her address such as their ZIP code or apartment number; and the interviewer accepts the don't know response for the ZIP code or the apartment number, then do code this as "failure to probe for the detailed address information after a DK response." *Example: I- And, uh, thank you. Do you happen to know the zip code for that? R- No. I- No, okay. <Then, interviewer moves to the next question>*
INTERVIEWER COMMUNICATIVE BEHAVIORS	
Interviewer Rapport Behaviors	• Improper Positive Feedback (IPF): Interviewer provides a positive phrase following a response to a study-related probe. This behavior is also coded when interviewer provides feedback that carries a positive connation and is geared towards attempts to make respondents more engaged with the study. *Example: R: I'm not going to be able to give you many street addresses. I: Um, that's fine. Just do as good as you can, and uh, we'll—we'll work around it.* • Neutral Feedback (INF): Interviewer provides a neutral phrase following a response to a study- related probe that shows appreciation for receipt of the response. In these situations, you should code phrases like "okay," "fine," or "thank you" as neutral. • Interviewer Joking/Sarcasm (IJS). Interviewer jokes or provides a sarcastic comment both regarding a **study-related or un-related** topic. *Example: My computer and I are not getting along right at this second.* • Interviewer Empathy (IE): Interviewer feels concern for the respondent and tries to share or understand what the respondent is thinking or feeling about the interview OR about an experienced reported event. This includes implicit and explicit indications of understanding regarding respondents' past experiences or their interviewing experiences. *Example: I had to do this interview, and it was—it—it was like yours. There was a lot. Example: R: I'm trying to--I'd have to go through--back through all my--my tax returns and stuff, and I--. I: I understand.* • Interviewer Agreement (IA): Interviewer agrees with respondents' both study-related and non-study related comments. Code IA if the interviewer spontaneously provides an agreeing response such as "Yes/ M-pos/ Uh-pos/ Alright/ Alrighty/ I agree" to respondent's comments. DO NOT code IA if interviewer responds positively to an affirmative question.

	Example: R: This line is bad. I am having trouble. I: I can hear the static on it. Example: R: They didn't go downtown and buy it like they do now. I: Oh, no, uh-neg.
Interviewer Apologetic Behaviors	• Interviewer apology (regarding the interview) (IAP): Interviewer apologizes from the respondent regarding the interview/task/question(/naire)/computer program by specifically saying "I am sorry"/ "sorry"/ "I apologize". *Example: I am sorry that the interview takes too long. Example: R: I think this is foolish. I: Laugh-I. Well, I'm sorry.* Interviewer apology for his/her error (IAE). Interviewer apologizes from the respondent regarding his/her own error by specifically saying "I am sorry/ sorry/ I apologize". *Example: I'm sorry, I need to back up, I put a wrong date in here.* • Interviewer apologetic comment (IAC): Interviewer provides a comment to **indirectly** apologize from the respondent **without specifically saying** "I am sorry"/ "sorry"/ "I apologize" regarding the interview/ task/ question(/naire)/ coding program/ for his or her error. *Example: R: What did I say earlier? I: I don't, uh--.The screen goes on, and I can't see the answers.* • Interviewer negative comment about the interview/task/ questionnaire (INC)*. *Example: We'll go back and catch that after I get done with this. Oh, why don't we do that before I get started here, because there's lots of questions here.*
Interviewer Laughter	• Interviewer apologetic laughter (IAL)* Interviewer laughs before or after he/she indirectly or directly apologizes to the respondent. Hence, code this ONLY when interviewer laughs before or after IAE/ IAP/ IAC. *Example: Okay, and uh, is name2 your biological child, or is name2 adoptled—adopted? Sorry. Laugh-I* • Interviewer uncomfortable laughter (IUL)*. Interviewer is using laughter to deal with an uncomfortable situation such as after a respondent negative comment (RNC), which can potentially increase the tension between the actors or after respondent's expression of difficulties (RED) or during talking about an uncomfortable or sensitive topic. *Example: R: No, I think that is a foolish question. I: Okay, Laugh-I* • Interviewer laughter to a respondent joke or a sarcastic comment (IL). *Example: R: I'm watching The Godfather. I can do that without sound. (Laugh-I.)*

DEVIATION FROM THE INTERVIEW	
External Interruption (EI)*	Interview is interrupted by an external source. *Example: R: Um, I got some noise on the line. I don't know what's the matter.*

** Please note that these codes are not used in the study, they are included in the coding scheme and verbal behavior coding production process for future use.*

CODING RULES:

1) If the information does not exist and R indicates that they cannot provide the answer because it is not applicable then do not code this as DK response. ***Example:*** *R: But I can't tell you the number because back then, we didn't have numbers.*
2) Every time respondent provides a "don't know" response, code this as either resolved or unresolved DK. If respondent resolves their DK response in the same turn, still code this as resolved DK.
3) Every time interviewer accepts respondent's "don't know" response without probing after the DK response code this as IFPDK or FPDA.
4) When a DK response is resolved much later in the interview, we still consider this as an unresolved DK if the interviewer did not resolve the issue when they were talking about the topic.
5) If the interviewer probes before respondent provides an unresolved don't know due to anticipation of a "don't know" response; but does not probe again after the "don't know" response, then code it as "failure to probe after DK response". ***Example:*** *I-So in what year did your weight change from being just right? R- Sigh, well gosh… I-Or, you can give me an age, and we can do it that way. R. I don't know. I don't know what to tell you. R- Okay. So, you want to put a don't know in that ans--that question then?*
6) If R is spontaneously mentioning if he/she does not know study-related information, code this as reluctant to provide information. ***Example****: I: Okay, this is going to be a good one. And okay. The boys—okay, let's start with the oldest. R: All right, name2. Name2. I can't tell you their birthdays. Laugh-R.* → *So, if this was said right after I'wer asked question about their b-days then it would have been un/resolved DK BUT R is spontaneously telling I'wer that he/she does not know their b-days that is why this is reluctant to provide info.*
7) If one laughs right after the other then do not code the following laughter. ***Example:*** *R: No, I do not know ma'am. I: Okay, there. Laugh-I. (Laugh-R.) Now, I'd like to ask if--.*
8) R laughter that expresses difficulty answering a question needs to be coded as uncomfortable laughter. ***Example:*** *I: Okay uh, could you tell me the month and the year he was born? R: Laugh-R. No I—well, it was November the 22nd.*

9) If respondent spells out without interviewer asking for it/ for any clarification, code the turn as R provides clarification. ***Example:*** *I: What was the name of the place? R: Employer4. E-m-p-l-o-y-e-r4.*
10) If the respondent is repeatedly having difficulty answering a question, code this repeatedly at each turn. This rule would also apply to other codes except laughter.
11) If the respondent corrects an earlier response that she/he had provided in the same turn, code this as R-correction. ***Example:*** *I was on that ship until I got out of the Navy, which was February--no actually Mar--March, 1963.*
12) Every time interviewer says "I am sorry"/"sorry", code this as either IAP or IAE or IE. If interviewer says "I am sorry"/"sorry" to have respondents repeat themselves, then code it as IAE. If interviewer says "I am sorry"/"sorry" to sympathize with the reported experience then code this as IE.
13) The same utterance cannot be coded in two or more different ways in other words, the same phrase cannot be double coded. Therefore, specific behaviors have priority over more general behaviors.

Hierarchy Order: (The more specific the code is, the hierarchically prior it is)

a) Respondent uncomfortable laughter has priority over respondent laughter to an interviewer joke, comment or feedback.
b) Interviewer uncomfortable laughter has priority over interviewer laughter to a respondent joke or a sarcastic comment.
c) I'wer apology (IAP and IAE - direct apology such as "I am sorry") has priority over interviewer apologetic comment (IAC – indirect apology).
d) Interviewer failure to probe after a DK response (IFPDK) has priority over interviewer failure to probe (IFP – failure to ask a question that is provided in the questionnaire).
e) If interviewer agrees with the respondent due to emphasize with the situation or express sympathy then Interviewer Empathy (IE) has priority over Interviewer Agreement (IA).
f) Respondent corrects interviewer (RCI) has priority over respondent asks or provides clarification (ROPC). Example: *I: You stayed with your mom for a period of time on address3--.R: My mother and dad, yeah.* → *Respondent provides omitted information by the interviewer.*

Appendix B: Coding Scheme that includes Old Verbal Behavior Codes (Verbal Behaviors, Definitions, and Examples)**

RESPONDENT RETRIEVAL STRATEGIES	
Respondent Parallel	Respondent spontaneously refers to a contemporaneous state or event in an area different from the required elements of a domain. *Example*: *It was football season when it started up.*
Duration Response	Respondent spontaneously provides how long a spell occurred. *Example*: *I worked for a year.*
Sequential Response	Respondent spontaneously provides a data element for a spell that occurred earlier or later and has not yet been explicitly temporally defined in any way. *Example*: *So, if <u>three months</u> was a summer job, I guess it doesn't count.*
Timing Response	Respondent spontaneously provides the beginning or ending of a spell, or spontaneously indicates any specific date. *Example*: *Um, so that would have been <u>September of the year prior</u>.*
INTERVIEWER RETRIEVAL PROBES	
Parallel	Interviewer uses an event from the respondent's past as an anchor. This event is not part of the domain being administered. *Example: When you got married…that would be in May then of… (Used in residence domain).*
Duration	Interviewer is seeking how long a spell has occurred. *Example: How long did you work for them?*
Sequential	Interviewer is probing for data elements of a period of time that happened earlier or later and has not yet been explicitly temporally defined in any way. *Example: Okay, uh, sir, can you tell me where you lived before you moved to city1?*
Timing	Interviewer is seeking when a spell began or ended. *Example: And please tell me again when you were married?*
INTERVIEWER DEVIATION FROM CONVENTIONAL IDEALS	
Directive	Interviewer provides any probe that poses the risk of biasing the respondent's answer. *Example: Do you remember if it was winter, or…? (In this case, a nondirective way of asking the question can be "do you remember which season this incident occurred?")*

*** This coding scheme is a section of the coded scheme that has been used in Bilgen and Belli, 2010b, pages 27 through 32.*

Appendix C: Information regarding How to Use the Coding Program:

Figure C.1 Tooltip for Interviewer Codes

(Entering Interviewer Codes via Quick Entry (QE)/Mouse Click)

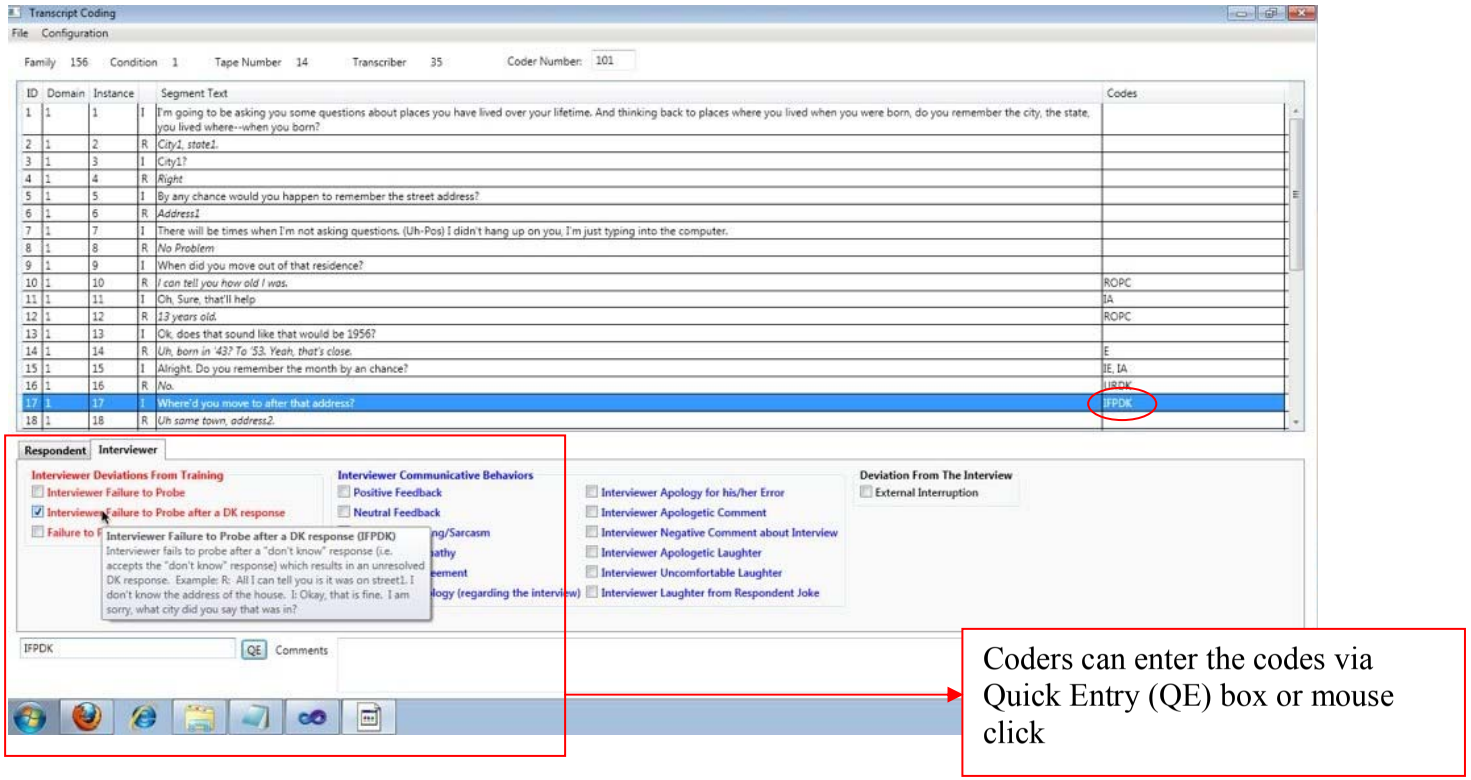


Figure C.2 Tooltip for Respondent Codes

(Entering Respondent Codes via QE/Mouse Click)

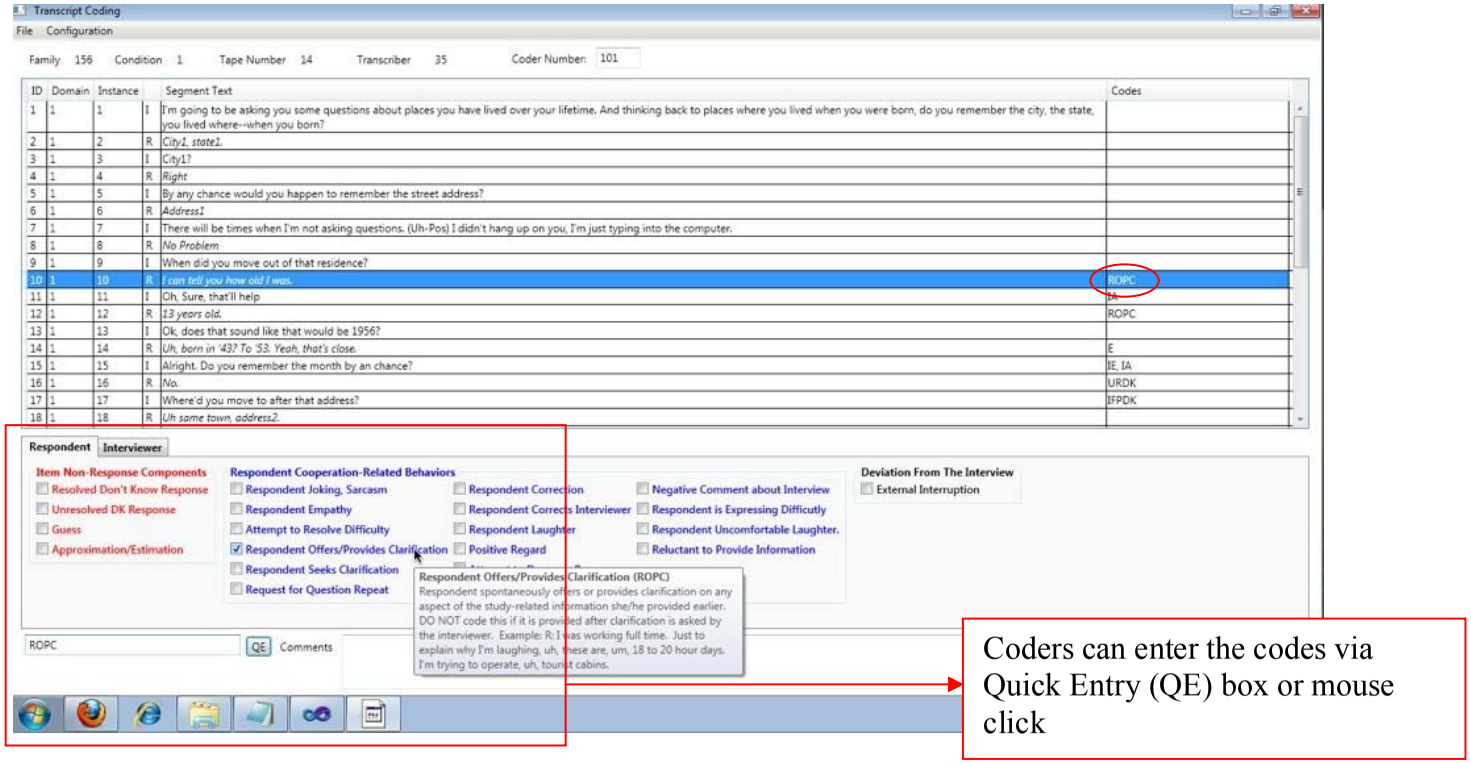


Figure C.3 Comment Box – Entering a Comment about a Turn

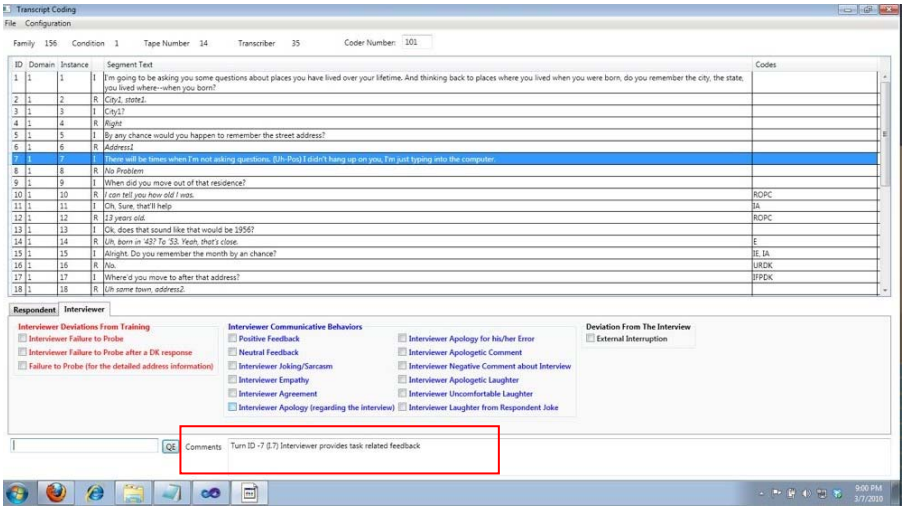


Figure C.4 Double Coding in One Turn

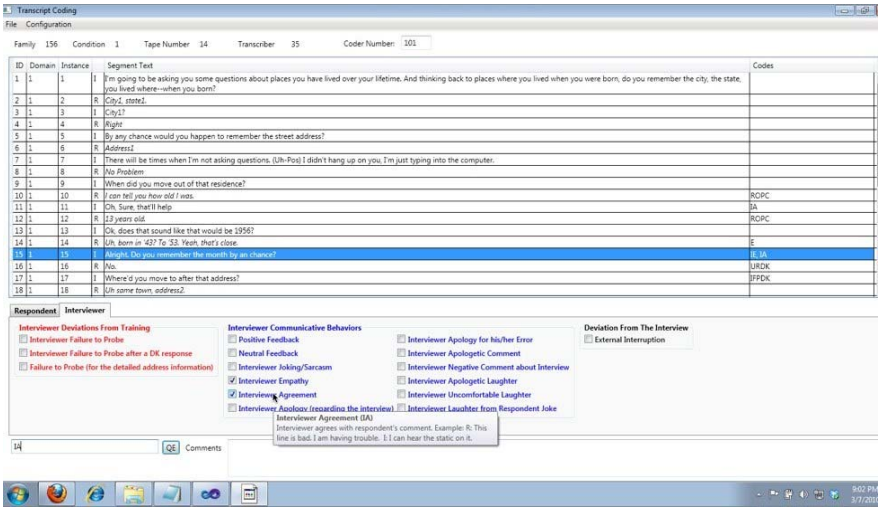


Figure C.5 File Menu – Importing and Exporting Files

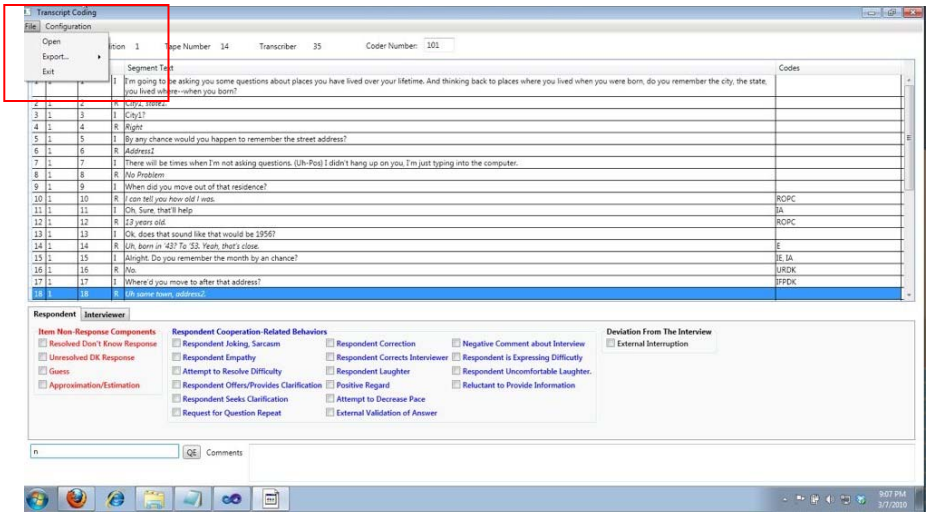
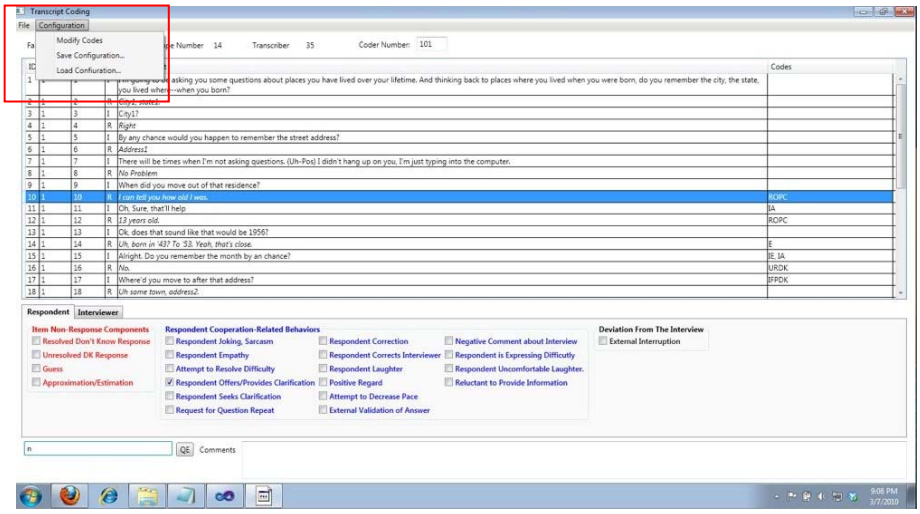


Figure C.6 Configuration Menu – Changing the Design Options



The coders can enter the codes either via clicking on the code or entering the acronyms into the Quick Entry (QE) box while they are coding each turn (see Figures A and B). The coding program allows coders to provide comments if they have any concerns or questions for me to look at (Figure C) and also allows them to code two verbal behaviors in one turn (Figure D). Using the *File Menu*, the coders will upload transcripts from a Word document into the coding program and when they are finished coding a transcript they will export the finished product in an Excel File format. The program also allows users to upload the coded Excel file into the program if the coders need to recode any completed transcripts or for me to monitor coding production quality (Figure E). Any design changes in the program can be made using the *Configuration Menu* (Figure F).

Appendix D: Definitions for Commonly used Terms in the Dissertation

Domain- The areas of life those are included in the calendar and standardized questionnaire in order to collect information from respondents on events during their entire lifetimes. In the transcripts each domain is numbered with a numbering system such as; *Domains: I, II, III…*

DOMAINS THAT ARE COLLECTED IN THE CURRENT STUDY:

- Respondent Residential History (Residence Domain)
- Respondent Partnership History (Marriage Domain)
- Respondent Parenting History (Children Domain)
- Respondent Education History (Education Domain)
- Respondent Labor History (Labor Domain)
- Respondent Health History (Health Domain)
    - Disability, Health Status, Smoking, and Weight

Probing- Probing occurs when interviewer asks the respondent for information. Probes can be divided into different kinds based on what kind of information they are seeking.

Response- An utterance that is given by the respondent in reaction to a probe by the interviewer.

Script- Scripts are lists of questions and/or statements written by the researcher, that are intended to be read by the interviewer to the respondent, to define the questions that will be asked and to present each area of data collection.

Spell: A spell is a continuous or ongoing type of activity. A spell refers to period of stability between two points of time. For example, an employment spell is the period between the beginning and end of a particular job.

String (Utterance) - A string (or an utterance) can be a single word or any combination of words that contains a meaning (i.e., some singular response other than "uh" or "uhm"). Occasionally, a string (or an utterance) may receive no code, or may receive more than one code (as in the case of directive probes) but typically receives only one code.

Turn- A turn is any combination of one or more strings. Turns are divided up between strings spoken by the respondent and the interviewer.

Appendix E: Calendar Questionnaire Script

Residence

Let's start with places that you have lived. Can you tell me the addresses of each of the places that you have lived over your entire lifetime and when you lived there? You can start since when you were growing up or you may want to think of where you are living now and work backwards in time.

Of course, I'd be interested in where you were living when you were born, and any moves that you made when you were very young. You may have been too little to remember some moves directly, but you may know about them from what other family members have told you.

Marriage

Now I would like to know whether you have ever been married. I would also like to know if you have ever lived with anyone as if married.

IF BOTH:
If you could, please tell me the first names or the initials of each person to whom you have legally married and to whom you have lived with as if married.
Now I am interested in those years in which you were living with NAME while legally married, and those years in which you were living with NAME while not married. I am also interested in any times in which you were not living with NAME for 4 months or more.

For spouse who stopped living with: Did you stop living with NAME because of separation, divorce, or because you were widowed?

IF MARRIED ONLY:
If you could, please tell me the first names or the initials of each person to whom you have legally married.
Now I am interested in those years in which you were living with NAME while legally married. I am also interested in any times in which you were not living with NAME for 4 months or more for any reason.

For spouse who stopped living with: Did you stop living with NAME because of separation, divorce, or because you were widowed?

IF UNMARRIED PARTNER ONLY:
If you could, please tell me the first names or the initials of each person to whom you have lived with as if married.

Now I am interested in those years in which you were living with NAME as if married. I am also interested in any times in which you were not living with NAME for 4 months or more for any reason.

Children:

So that I can keep track of each of your children as we talk about them, can you please tell me with the first name or provide initials of each child you (fathered/gave birth to) or formally adopted?

While NAME was growing up from (birth/time of adoption) until (turning 18 years of age/the time of death) was there ever a period of 4 months or more when NAME was living apart from you?

(If yes): Now I am interested in learning about those periods of time in which your children living apart from you, while they were growing up. In thinking about whether a child was living apart from you, please include any times in which your child was living elsewhere for 4 months or more.

Education

I am now interested in the formal education that you have had over your entire lifetime. Please tell me about those periods in which you were attending elementary school, middle or junior high school, and high school. If you did not graduate from high school, but took classes to earn a GED, I would like to know about this as well.

In addition, I would like to know if you attended college, and professional and graduate school. For these schools, I would like to know whether you were attending part- or full-time, based on the number of credits that you were taking. *(Please do not include trade or vocational schools such as beauty school, barber college, and so on).*

Labor

MAIN EMPLOYERS
Now I would like to talk with you about your work for pay. Have you ever worked for pay at the same job for 3 months or more? (If yes) Could you please tell me about your jobs for pay since you were 14 years of age, or when you began working, including any self-employment you may have had?

ANNUAL EMPLOYMENT STATUS
[IWER: Please note any blue years and verify that R did not work during those years]
So, where you were working for [employer name] for each year were you working the entire year or just part of the year? If you were working just part of the year for [employer name], did you have any other jobs that year so that you did have a job for the entire year?

FULL-TIME/PART-TIME
When you were working for [employer name] from [start date] to [end date], was that full-time, part-time or some of both? (If part-time) Did you work at any additional jobs so that all together it was full-time?

- Full-time = 30 hours a week or more
- Remember that in any given year, you could have been working both full- and part-time.

UNEMPLOYMENT
Finally, I am also interested in knowing about any periods of time of one week or longer since you were 14 years of age in which you did not have a job at all for pay, and you were looking for work at that time.

Health

DISABLING HEALTH CONDITIONS:
We would like to know about any instances in which, because of injury, illness or disability you missed attending school or work for one month or more. Please tell me when any of these periods happened, how long they occurred, and the type of injury or illness that it was. For periods during your very young childhood you may know about periods of injury or illness from what other family members have told you.

We are also interested in any periods of time during your lifetime in which, because of injury, illness, or disability, you were confined to a hospital or to a bed at home for one month or more.

HEALTH STATUS:
We would also like to know how your general health has been over your entire lifetime. For each year of your life, would you say that your health had been excellent, very good, good, fair, or poor. It may help to focus on those years in which there was a CHANGE in your general health from one state to another.

SMOKING:
Did you ever smoke cigarettes? (If yes) How old were you when you first started? We're interested in those times of your life when you did not smoke at all, when you smoked between 1-10 cigarettes a day, between 11-20 cigarettes a day, and 21 or more per day.

WEIGHT:
Finally, I'd like to ask you about your weight. Please think of five different categories that may have applied to you at different times during your life, you were very overweight, slightly overweight, just right, slightly underweight or very underweight. So, during any period of your life were you very overweight? Slightly overweight? Just right? Slightly underweight? Very underweight?

Appendix F: Standardized Questionnaire Script

I would like to start out by asking about places where you have lived. I would like you to tell me about the addresses for each place that you have lived in your entire lifetime. Please do not include any temporary visits with relatives or friends unless you considered this residence your main residence. Also, please tell me about any times that you were homeless.

Of course I'd be interested in where you were living when you were born, and any moves that you made when you were very young. You may have been too little to remember some moves directly, but you may know about them from what other family members have told you.

Let's start with your first address and then work forwards in time.

ADDRESS1

AA-1. What was your first address, what address did you live at when you were born?

Address1:_____
City:_____
State:____
Zipcode:_____ **Out of Country code= OC**

ALWAYS1

AA-2 Have you always lived at (address1)?

Yes.....1

No.....5

Flow Check AA-2: If yes go to BB-1

MOVE1

AA-3. In what month and year did you move from (ADDRESS1)?

__ __/__ __
Mo/Yr

{Interviewer should probe for month and year using age if necessary}

Flow Check AA-3: calculate age from month and year given:
If Age 0-17 go to AA-4 and AA-5
If Age 18+ skip AA-5

ADDRESS2

AA-4. What was the address of the place that you moved to ?

Street:_____
City:_____
State:____
Zipcode:_____

FINPROB1

AA-5. Did you and your family move to (ADDRESS2) because your family was experiencing financial problems?

Yes 1
No.....5

DIFPLAC1

AA-6. Did you ever live in a different place after you lived at (ADDRESS2)?

Yes 1
No.....5

CYCLE THROUGH AA-4 –AA-6 UNTIL AA-6=NO
Flow Check AA-6: if no go to BB-1.

TIMEMAR

BB-1. Now I would like to ask you a few questions about marriage. Altogether, how many times have you been married?

_____ times {max=20}

Flow Check BB-1: if 0 go to BB-8

CURMAR

BB-2a. Are you legally married at the present time?

Yes 1
No.....5

Flow Check BB-2a: If BB-1=1 go to BB-3~~a~~, if BB-1>1 go to BB-4.

MARBEG

BB-3. In what month and year did your marriage begin?

__ __ / __ __
Mo/Yr.

SEPARAT1

BB-3A. Was there ever a period of time when you and your spouse were living in separate places for four months or more while still legally married? Please consider any separation for any reason.

Yes 1
No.....5

Flow Check BB-3A : If no and BB-2a=no, go to BB-5. If no and BB-2a =yes, go to BB-8.

TIMESEP1

BB-3B. How many times were you separated from your spouse for four months or more while you were still married?

____ times {max=20}

Flow Check BB-3B: If times > 1, then use appropriate fill in BB-3C and BB-3D

SEPBEG1

BB-3C. (The (first,second,etc) time you were separated,) when did the separation begin?

__ __/__ __
.....Mo/Yr.

For the last cycle of separation, please insert:

BB-3C1. Are you still separated?

Yes 1
No.....5

Flow Check: If yes and BB-2a=yes, go to BB-8.
If yes and BB-2a=no, go to BB-5.

SEPEND1

BB-3D. (The (first,second,etc) time you were separated) when did the separation end?

__ __/__ __

.....Mo/Yr.

Flow Check BB-3D: continue to cycle between BB-3C – BB-3D until reach # times in BB-3B. After last cycle, if BB-2a=yes, go to BB-8. If BB-2a=no, go to BB-5.

MARRIED1
BB-4. In what month and year did your first marriage begin?

__ __ / __ __
Mo/Yr.

MAREND1
BB-5. Did the marriage end in divorce, or were you widowed?

Divorce.....1
Widowed.....2

YRSMAR1
BB-6. In what month and year did the marriage end?

__ __ / __ __
Mo/Yr.

Flow Check BB-6: if BB-1 = 1, go to BB-8

SEPARAT2

BB-6a. During this marriage, was there ever a period of time when you and your spouse were living in separate places for four months or more while still legally married? Please consider any separation for any reason.

Yes 1
No.....5

Flow Check BB-6a: if no go to BB-7

TIMESEP2
BB-6b. How many times were you separated from your spouse for four months or more while you were still married?

____ times {max=20}

Flow Check BB-6B: If times > 1, then use appropriate fill in BB-6C and BB-6D

SEPBEG2
BB-6c. (The (first,second,etc) time you were separated,) when did the separation begin?

__ __/__ __
Mo/Yr.

For the last cycle of separation, please insert:
BB-6C1. Are you still separated?

Yes 1
No.....5

Flow Check: If yes skip BB-6d.

SEPEND2
BB-6d. (The (first,second,etc) time you were separated) when did the separation end?

__ __/__ __
.....Mo./Yr.

Flow Check BB-6D: continue to cycle between BB-6C – BB-6D until reach # times in BB-6B.

MARRIED2
BB-7. In what month and year did your (second,third, etc. – from Q1) marriage begin?

__ __/__ __
Mo./Yr.

Flow Check BB-7: repeat BB-5 – BB-7 as necessary to reach number in BB-1. If BB-2a=yes, then the final cycle should ask BB-6A-BB-6D and then skip to BB-8.

LIVPART
BB-8. Have you ever lived with a partner as if married?

Yes 1
No.....5

Flow Check BB-8: If no go to CC-1

TIMESLIV
BB-9 How many times have you lived with a partner as if married?

_____ times {max=20}

Flow Check BB-9:
if BB-2A=yes and BB-3C1 ne yes and BB-9>1 then go to BB-12.
if BB-2A=yes and BB-3C1 ne yes and BB-9=1 then go to BB-11
else go to BB-10

CURPART
BB-10. Are you currently living with a partner as if married?

Yes 1
No.....5

Flow Check BB-10: if BB-9 > 1 go to BB-12

CURSTART
BB-11. In what month and year did you start living with this partner?

__ __ / __ __
Mo/Yr.

Flow check BB-11: if BB-10 = 1 or if BB-1 = 0, go to BB11-C

EVERMAR1
BB-11A Did you ever marry this partner?

Yes 1
No..... 5

Flow Check BB-11A: if no go to BB-11C

MARPART1
BB-11B In what month and year did you marry?

__ __ / __ __
Mo/Yr.

FLOWCHK1
BB-11C **Flow Check BB-11C:**
If BB-1 = 0, and BB-10 = 5, use second fill in BB-11D.
If BB-11A = 1, use first fill.
ELSE use no fill.

PARTSEP

BB-11D. Since the time when you began living with this partner (until you were married on (month, year)/until the time that you were no longer ever living with this partner), was there ever a period of time when you and your partner were living in separate places for four months or more?

Yes 1
No.....5

Flow Check BB-11D: if no and BB-10 = 1, go to CC-1; if no and BB-10 =5, go to BB-15

TIMEPAR

BB-11E. How many times were you separated from your partner for four months or more?

____ times {max=20}

Flow Check BB-11E: If times > 1, then use appropriate fill in BB-11F and BB-11G

PARBEG

BB-11F. (The (first,second,etc) time you were separated,) when did the separation begin?

__ __/__ __
Mo/Yr.

For the last cycle of separation, please insert:

BB-11F1. Are you still separated?

Yes 1
No.....5

Flow Check: If yes skip BB-11G.

PAREND

BB-11G. (The (first,second,etc) time you were separated)when did the separation end?

__ __/__ __
.....Mo./Yr.

Flow Check BB-11G: continue to cycle between BB-11F – BB-11G until reach # times in BB-11E. If BB-10 = 1, then go to CC-1, else go to BB-15

BEGPART1

BB-12. In what month and year did you start living with a partner as if married for the (first, second, third, etc) time?

__ __ / __ __

Flow check BB-12: if f BB-1 = 0, go to BB12-C

EVERMAR2

BB-12A Did you ever marry this partner?

Yes 1
No..... 5

Flow Check BB-12A: if no go to BB-12C

MARPART2

BB-12B In what month and year did you marry?

__ __ / __ __
Mo/Yr.

FLOWCHK2

BB-12C **Flow Check BB-12C:**
If BB-12A = 1, use first fill.
If last cycle on BB-9 and BB-10 = 1, use no fill.
ELSE use second fill.

PARTSEP2

BB-12D. Since the time when you began living with this partner (until you were married on (month, year)/until the time that you were no longer ever living with this partner as if married), was there ever a period of time when you and your partner were living in separate places for four months or more?

Yes 1

No.....5

Flow Check BB-12D: if no and on last BB-9 cycle and BB-10 = 1, go to CC-1; if no and not on last cycle go to BB-15.

TIMEPAR2
BB-12E. How many times were you separated from your partner for four months or more?

____ times {max=20}

Flow Check BB-11E: If times > 1, then use appropriate fill in BB-11F and BB-11G

PARBEG2
BB-12F. (The (first,second,etc) time you were separated,) when did the separation begin?

__ __/__ __
Mo/Yr.

PAREND2
BB-12G. (The (first,second,etc) time you were separated) when did the separation end?

__ __/__ __
.....Mo./Yr.

Flow Check BB-12G: continue to cycle between BB-12F – BB-12G until reach # times in BB-12E. If on last BB-9 cycle and BB-10 = 1 go to CC-1, else go to BB-15.

BB-15. Are you still living with this partner?

Yes 1
No.....5

Flow Check: If yes skip BB-15A.

DURPART1
BB-15A. In what month and year did you stop living with this partner?

__ __/ __ __
Mo/Yr.

Flow Check BB-15: If BB-9=1, go to CC-1. Else, cycle through BB-12-BB-15 until number in BB-9 met.

CHILDREN

CC-1. I am also interested in learning about your children. I will ask you questions about all of the births you have had during your life and about all children you have adopted. So, the next several questions refer only to children you have fathered/given birth to or whom you have adopted. I am not interested in miscarriages, stillbirths or abortions, or your step-children whom you have never adopted. How many children have you had? In this number include only children you have fathered/given birth to or you have formally adopted.

..... __ __ Children {max=12}
.....
..... **Flow Check CC-1: if 0 go to DD-1**

Now I would like to ask questions about each of your children starting with the first one you had.

NAME

CC-2. What is the name of your (first,second, third, etc) child?

_____

GENDER

CC-3 Is (name from CC-2) a boy or a girl?

Girl 1
Boy 2

DOB

CC-4 In what month and year was (name from CC-2) born?

__ __/__ __ __ __
Mo. Year

BIOCHILD

CC-5 Is (name) your biological child or is (name) adopted?

Biological.....1
Adopted.....2

Flow Check CC-5: if 1 go to CC-7

WHENADOPT
CC-6 In what month and year did you adopt (name)?

__ __ /__ __ __ __
Mo. Year

ALIVE
CC-7 Is (name) still alive?

Yes 1
No 5

Flow Check CC-7: if 1 go to CC-9

DIED
CC-8 In what month and year did (name) die?

__ __/__ __ __ __
Mo. Year

LIVEAWAY
CC-9 [Before (name) was 18 years old]/[Before (name) died]/[After (name) was adopted and before (name) was 18 years old], did he/she ever live away from you for 4 continuous months or more?

Yes 1
No 5
Always Lived Away ..6

Fill logic: **calculate age at death:**
If current age < 18 then use "Until now"
If current age >= 18 then use "Before (name) was 18 years old"
if CC-8 < 18 then use fill "Before (name) died",

If CC-5=adopted use "After (name) was adopted and before (name) was 18 years old" .
If adopted and died then use "After (name) was adopted and before (name) died.

Flow Check CC-9: if 5 Go to Next Child.
Cycle through 2-11 until reach number of children in CC-1.

TIMEDIF1
CC-9. How many different times did (name) live somewhere else for 4 continuous months or more before they were 18?

__ __ times {max=20}

BEGAWAY1

CC-10. In what month and year did (kid 1) start living away from you the (first, second, etc.) time before they were 18?

__ __/__ __
Mo/Yr.

DURAWAY1

CC-11. How long did (kid 1) live away from you for the (first, second, etc.) time before they were 18?

__ __ Months/Years

Flow Check CC-11: repeat CC-10 – CC-11 for each time in CC-9. After maximum times go to next child.

GRADHS

DD-1. Now I would like to ask a few questions about your education. Did you graduate from high school, pass a high school equivalency test, or neither?

No	1
Yes, High School	2
Yes, GED	3

Flow Check DD-1: If 1 go to DD-8, If 2 go to DD-4.

STRTGED

DD-2. In what year did you start taking classes for your a high school equivalency test?

Year _____

GEDYR

DD-3. In what year did you pass your high school equivalency test?

Year _____

Flow check DD-3: go to DD-5.

HSYR

DD-4. In what year did you receive your high school diploma?

Year _____

COLLEGE
DD-5. Did you ever attend a university or college? Do not include trade or vocational schools such as beauty school, barber college, and so on.

Yes 1
No 5

Flow Check DD-5: if no go to EE-1.

STRTCOL
DD-5A. In what year did you begin taking college classes?

__ __ __ __
year

COLGRAD
DD-6. Did you ever graduate from college?

Yes 1
No 5

Flow check DD-6: if yes go to DD-7

LASTCOL
DD-6A. In what year did you last attend college?

__ __ __ __ Year

Flow Check DD-6A: if year ≥ 1985, go to DD-7B, else go to EE-1.

DEGREE
DD-7. What was the highest degree that you received?

Associates 1
B.A. 2
B.S. 3
M.A. 4
M.P.H. 5
M.B.A. 6
Ph.D. 7
M.D. 8

J.D. or Law degree 9
Other (specify) 10

YRDEGREE
DD-7A. In what year did you receive your (fill) degree?

Year_____

Flow Check DD-7A: if year < 1985 go to EE-1.

COLLATT1
DD-7B Were you attending college during each of the years from when you started college in (year from DD-5A) until (year in DD-6A or year in DD-7A)?

Yes
No

Flow Check DD-7B: if yes, go to DD-7D

COLLATT2
DD-7C [Mark, please create year string for applicable years of possibly attending college from start and end years in DD-7B] In which of the years between (year from DD-5A) and (year in DD-6A or year in DD-7A) were you attending college?

FULLPART
DD-7D [Mark, please create year string for applicable years of attending college from start and end years in DD-7B if yes, or from years indicated from Dd-7C; then for each year, ask separately] I am interested, for each of the years that you were attending college, whether you were attending full- or part-time. During (year 1, 2, 3, etc.), did you attend college full-time, part-time, or some of each?

Full time 1
Part time 2
Some of each 3

GO TO EE-1

HIGRADE
DD-8. What was the highest grade or year of school you have completed?

1.....1
2.....2

3.....3
4.....4
5.....5
6.....6
7.....7
8.....8
9.....9
10.....10
11.....11

GRADYR
DD-9. In what year did you complete this year of school?

__ __ __ __ year

EVRWRK1
EE-1. Now I would like to talk with you about your work for pay. I am going to ask you questions about your main employers from the time you were 14 years old until now. In thinking about your work for pay, include all types of jobs. This includes any jobs in which you worked for someone else, any self-employment, and both full and part-time work. Have you ever worked for pay at the same job for 3 months or more?

Yes 1
No.....5

Flow Check EE-1: If no go to EE-3

CURRWRK
EE-2. Are you currently working for pay?

Yes 1
No.....5

Flow Check EE-2: If yes go to EE-5

CURRLOOK
EE-3. Are you currently looking for a job?

Yes 1
No.....5

Flow Check EE-3: If no and EE-1=1 go EE-19A,
if no and EE-1=5 go to EE-34.

STRTLOOK
EE-4. In what month and year did you start looking for work?

__ __/__ __
Mo/Yr

Flow Check EE-4: If EE-1=1 go to EE-19A, if EE-1=5 go to EE-34

SLFOTHR1
EE-5. For the next few questions, we would like you to think of your main job or main employer. In your current main job, are you working for someone else or are you self-employed?

Someone Else.....1
Self-Employed...2
Flow Check EE-5: If Self-Employed go to EE-7

EMPNAME1
EE-6. What is the name of your employer, who do you work for?
(IF R DOES NOT WANT TO PROVIDE EMPLOYER NAME, ASK FOR SOMETHING THAT WILL HELP IDENTIFY THIS JOB, I.E. POSITION, TYPE OF EMPLOYER, ETC…)

Employer Name_____

STRTJOB1
EE-7. In what month and year did you start (with EMPNAME1/your current business) as your main job?

__ __/__ __
Mo/Yr

STOPJOB1
EE-7b. Since (EE-7) until now, did you ever stop working (for employer name/ at your current business) entirely for any period of time so that during this time you did not consider ([employer name] as your employer/ your current business as a source of income)]?

YES 1
NO 5

Flow Check EE-7b.: If yes go to EE-7c

DIFFJOB1

EE-7a. Since (EE-7) until now, did you ever have a different main job than working (for employer name/ at your current business)?

YES	**1**
NO	5

Flow Check EE-7a: If no go to EE-8

RECSTRT1

EE-7c. When was the most recent start date in which you had been working continuously (for employer name/ at your current business) as your main job? Please provide month and year.

__ __ / __ __
Mo Yr

CONSISTENCY CHECK: Check start dates in EE-7 and EE-7c. If they are the same, I'wer needs to clarify with respondent that continuous work means no intervening other main job or no intervening work stoppage.

FULLPRT1

EE-8. (READ SLOWLY). We would like you to tell us whether you have been working full or part-time while working (for employer name/ at your current business) since (start date in EE-7c or EE-7 if not asked EE-7c) until now as your main employer. (In your answer consider full-time work as an average of 30 or more hours per week). Thinking back to when you first started working (for EMPNAME1/your current business) as your main employer in (more recent of STRTJOB1 or RECSTRT1), were you working full or part time?

Full time.....1
Part time.....2

FLPTCNT1

EE-9. Have you been working continuously (full/part) time since (more recent of STRTJOB1 or RECSTRT1) until now?

Yes 1
No.....5

**Flow Check EE-9: If yes and EE-8=1 go to EE-18,
if yes and EE-8=2 go to EE-11A**

FLPTCHG1
EE-10. In what year did you change to working (full/part) time?

__ __ __
Yr

FLPTCNT2
EE-11. Since starting to work (full/part) time in (FLPTCHG1), have you continuously worked (full/part) time until now?

Yes 1
No.....5

Flow Check EE-11: If no cycle through EE-10 - EE-11 until EE-11=yes.

NEWVAR1
EE-11A. **Flow check EE-11A. Determine number of part-time periods and whether the start end of each period is the same year, or different years. If there is more than one part-time period during the same year, only use one. If different years, go to EE-15.**

OTHRJOB1
EE-12. From what you have just told me, you were working part-time (for EMPNAME1/at your current business) during (year). Did you have any other jobs during (year) while working part-time (for EMPNAME1/at your current business)?

Yes 1
No.....5

Flow check EE-12: If no and number of part time periods from EE-11A>1 repeat EE-11A. If no and number of part time periods is exhausted go to EE-18.

OTHRFUL1
EE-13. When you were working part-time for (EMPNAME1/your current business) during (year) and working for the other job or jobs during the same weeks, would you say you were ever working more than 30 hours per week on average considering the time spent on all jobs?

Yes 1
No.....5

Flow check EE-13: If no and number of part time periods from EE-11A>1 repeat EE-11A. If no and number of part time periods is exhausted go to EE-18.

NEWVAR2

EE-14. During (year), was there ever a period of time when you were working, on average, less than 30 hours per week considering the time spent on all jobs?

Yes 1
No.....5

Flow check EE-14: If number of part time periods from EE-11A = 1 or EE-11A>1 and number of part time periods is exhausted go to EE-18; else repeat EE-11A.

OTHRJOB?

EE-15. From what you have just told me, you were working part-time (for EMPNAME1/at your current business) from (part time start year) until (part time end year/until now). Did you have any other jobs during this period of time?

Yes 1
No.....5

Flow check EE-15: If no and number of part time periods from EE-11A>1 repeat EE-11A for each part time period. If no and number of part time periods is exhausted go to EE-18.

OTHRFUL?

EE-16. When you were working part-time for (EMPNAME1/your current business) from (part time start year) until (part time end year/until now), and working for the other job or jobs during the same weeks, would you say you were ever working more than 30 hours per week on average considering the time spent on all jobs?

Yes 1
No.....5

Flow check EE-16: If no and number of part time periods from EE-11A>1 repeat EE-11A. If no and number of part time periods is exhausted go to EE-18.

NEWVAR?
EE-17. When you were working part-time for (EMPNAME1/your current business) from (part time start year) until (part time end year/until now), and working for the other job or jobs during the same weeks, in what years were you ever working more than 30 hours per week on average considering the time spent on all jobs?

Include year string that encompasses each year from the start until the end years inclusive. For each year indicated, ask EE-17A until all years are exhausted. Then go to EE-11A or EE-18, depending on whether all part-time periods are exhausted.

NEWVAR?
EE-17A. During (year), was there ever a period of time when you were working, on average, less than 30 hours per week considering the time spent on all jobs?

Yes 1
No.....5

EVRWRK2
EE-18A. We would like to know what kind of work you were doing before working (for yourself / at EMPNAME1) from (RECSTRT1) until (ENDJOB2 or now if no ENDJOB2). Before this time, have you ever worked at a job for 3 months or more?

Yes 1
No.....5

Flow Check EE-18A: If no go to EE-34.

SLFOTHR2
EE-19. For the next few questions, we would like you to think of your most recent main job or main employer prior to your working on the main job we just discussed (from (more recent of STRTJOB2 or RECSTRT2 or more recent of STRTJOB1 or RECSTRT1 if no STRTJOB2 or RECSTRT2) until (ENDJOB2 or now if no ENDJOB2)). In this most recent main job, were you working for someone else or were you self-employed?

Someone else.....1
Self-employed...2

Flow Check EE-19: If Self-Employed go to EE-21, all others Go to EE-20.

EE-19A For the next few questions, we would like you to think of your most recent job or main employer. In this most recent job, were you working for someone else or were you self-employed

Someone else.....1
Self-employed...2

Flow Check EE-19A: If Self-Employed go to EE-21

EMPNAME2
EE-20. What was the name of your employer, who did you work for?

Employer Name_____

STRTJOB2
EE-21. In what month and year did you start (with EMPNAME2/that business) as your main job?

__ __/__ __
Mo/Yr

Check: If no EE-19 then begin question EE-22 with "In what month…"

ENDJOB2
EE-22. Prior to (use same RECSTRT1, RECSTRT2, STRTJOB1, STRTJOB2 date as used in EE-19), in what month and year did you stop working (EMPNAME2/ at that business) as your main job?

__ __/__ __
Mo/Yr

STOPJOB2
EE-22b. Since (STRTJOB2) until (ENDJOB2), did you ever stop working (for EMPNAME2/ at that business) entirely for any period of time so that during this time you did not consider ([EMPNAME2] as your employer/ that business as a source of income)?

YES 1
NO 5

Flow Check EE-22b: If yes go to EE22c

DIFFJOB2
EE-22a. Since (STRTJOB2) until (ENDJOB2), did you ever have a different main job than working (for EMPNAME2/ at that business)?

YES 1
NO 5

Flow Check EE-22a: If no go to EE-23

RECSTRT2
EE-22c. Prior to ENDJOB2, when was the most recent start date in which you had been working continuously (for EMPNAME2/ at that business) as your main job until ENDJOB2? Please provide month and year.

__ __/ __ __
Mo Yr

CONSISTENCY CHECK: Check start dates in EE-21and EE-22c. If they are the same, I'wer needs to clarify with respondent that continuous work means no intervening other main job or no intervening work stoppage.

FULLPRT2
EE-23. *(READ SLOWLY). We would like you to tell us whether you have been working full or part-time while working (for employer name/ at your current business) since (start date in EE-22c or EE-21 if not asked EE-22c) until **(ENDJOB2)** as your main employer. Thinking back to when you first started working (for EMPNAME2/your current business) as your main employer in (more recent of STRTJOB2 or RECSTRT2), were you working full or part time? (In your answer consider full-time work as an average of 30 or more hours per week).*

Full time 1
Part time 2

FLPTCNT3
EE-24. Did you work continuously (full/part) time from (*start date in EE-22c or EE-21 if not asked EE-22c*) until (ENDJOB2)?

Yes 1
No.....5

Flow Check: If yes and EE-23=1 go to EE-33,
if yes and EE-23=2 go to EE-26A

FLPTCHG2
EE-25. In what year did you change to working (full/part) time?

__ __ __
Year

FLPTCNT2
EE-26. Since starting to work (full/part) time in (FLPTCHG2), have you continuously worked (full/part) time until (ENDJOB2)?

Yes 1
No.....5

Flow Check EE-26: If no cycle through EE-25 - EE-26 until EE-26=yes.

NEWVAR?
EE-26A. **Flow check EE-26A. Determine number of part-time periods and whether the start end of each period is the same year, or different years. If there is more than one part-time period during the same year, only use one. If different years, go to EE-30.**

OTHRJOB?
EE-27. From what you have just told me, you were working part-time (for EMPNAME2/at that business) during (year). Did you have any other jobs during (year) while working part-time (for EMPNAME2/at that business)?

Yes 1
No.....5

Flow check EE-27: If no and number of part time periods from EE-26>1 repeat EE-26A. If no and number of part time periods is exhausted go to EE-33.

OTHRFUL?
EE-28. When you were working part-time for (EMPNAME2/that business) during (year) and working for the other job or jobs during the same weeks, would you say you were ever working more than 30 hours per week on average considering the time spent on all jobs?

Yes 1
No.....5

Flow check EE-28: If no and number of part time periods from EE-26>1 repeat EE-26A. If no and number of part time periods is exhausted go to EE-33.

NEWVAR?
EE-29. During (year), was there ever a period of time when you were working, on average, less than 30 hours per week considering the time spent on all jobs?

Yes 1
No.....5

Flow check EE-29: If number of part time periods from EE-26 = 1 or EE-26>1 and number of part time periods is exhausted go to EE-33; else repeat EE-26A.

OTHRJOB?
EE-30. From what you have just told me, you were working part-time (for EMPNAME2/at that business) from (part time start year) until (part time end year). Did you have any other jobs during this period of time?

Yes 1
No.....5

Flow check EE-30: If no and number of part time periods from EE-26>1 repeat EE-26A for each part time period. If no and number of part time periods is exhausted go to EE-33.

OTHRFUL?
EE-31. When you were working part-time for (EMPNAME1/that business) from (part time start year) until (part time end year), and working for the other job or jobs during the same weeks, would you say you were ever working more than 30 hours per week on average considering the time spent on all jobs?

Yes 1
No.....5

Flow check EE-31: If no and number of part time periods from EE-26>1 repeat EE-26A. If no and number of part time periods is exhausted go to EE-33.

NEWVAR?
EE-32. When you were working part-time for (EMPNAME1/that business) from (part time start year) until (part time end year), and working for the other job or jobs during the same weeks, in what years were you ever working

more than 30 hours per week on average considering the time spent on all jobs?

Include year string that encompasses each year from the start until the end years inclusive. For each year indicated, ask EE-32A until all years are exhausted. Then go to EE-26A or EE-33, depending on whether all part-time periods are exhausted.

NEWVAR?
EE-32A. During (year), was there ever a period of time when you were working, on average, less than 30 hours per week considering the time spent on all jobs?

Yes 1
No.....5

EVRWRK3
EE-33A. We would like to know what kind of work you were doing before working (for yourself / at EMPNAME2) (from (RECSTRT2) until (ENDJOB2 or now if no ENDJOB2)). Before this time, have you ever worked at a job for 3 months or more?

Yes 1
No.....5

Flow Check EE-33A: If yes cycle through EE-19 - EE-33A until EE-33A=no. If no go to EE-34.

EVERSRCH
EE-34 Now I would like to ask you about times when you have been out of work and actively looking for a job but unable to find one right away. Have there ever been any times when you were not working for pay at all and actively looking for work?

Yes 1
No.....5

Flow Check EE-34: If no go to FF-1.

YEARSRCH
EE-35 During which years were there times when you were not working for pay at all and actively looking for work?

Provide applicable year string based on age 14 to present

EVERHOSP

HH-1. Thinking over your entire life, from birth to the present, have you ever been hospitalized for one month or more?

Yes 1
No.....5

Flow Check HH-1: If no go to HH-12

NUMHOSP
HH-2. How many different times have you been hospitalized for one month or more?

_____times

WHYHOSP
HH-3. The first (second, third, etc.) time that you were hospitalized for one month or more, was it due to an injury, an illness, a pregnancy-related complication, or something else?

Injury.....1
Illness 2
Pregnancy-related……..3
Other (specify)_____4

CONDHOSP
HH-4. What was the specific condition that caused this first (second, third, etc.) hospitalization?

Condition_____

YEARHOSP
HH-5. The first (second, third, etc.) time you were hospitalized for one month or more, in what year did this begin?

__ __ __ __
Year

DURRHOSP
HH-7. How long were you hospitalized for the first (second, third, etc.) time?

____years _____month ____weeks ____days

BEDHOSP
HH-8. The first (second, third, etc.) time you were hospitalized for one month or more, did you spend any additional time confined to a bed at home due to this health condition, after leaving the hospital?

Yes 1
No.....5

Flow Check HH-8: If no go to HH-10

DURBDHSP
HH-9. How long were you were you confined to a bed at home after leaving the hospital the first (second, third, etc.) time, not including the time spent in the hospital?

____years _____month ____weeks ____days

WORKHOSP
HH-10. The first (second, third, etc.) time you were hospitalized for one month of more, did you miss additional school or work or did you become disabled due to this health condition, not including the time you spent in the hospital or confined to a bed?

Yes 1
No.....5

Flow Check HH-10: If no go to HH-12

DURWKHSP
HH-11. How long were you away from school or unable to work after the first (second, third, etc.) hospitalization, not including the time you spent in the hospital or confined to a bed?

____years _____month ____weeks ____days

Flow Check HH-11: If HH-2>1, cycle through HH-3 - HH-11 as many times as needed to reach number in HH-2.
For HH-12: If HH-1=Yes use intro fill

EVERBED
HH-12. (Besides the periods of time that you have told me about), have you ever been confined to a bed for one month or more outside of the hospital due to a health condition?

Yes 1
No.....5

Flow Check HH-12: If no go to HH-21

NUMBED
HH-13. (Besides the periods of time that you have told me about), how many different times have you been confined to a bed for one month or more due to a health condition?

_____times

WHYBED
HH-14. The first (second, third, etc.) time that you were confined to a bed for one month or more due to a health condition, was it due to an injury, an illness, a pregnancy-related complication, or something else?

Injury.....1
Illness 2
Pregnancy-related……..3
Other (specify)_____4

CONDBED
HH-15. What was the specific condition that caused you to be confined to a bed for one month or more the first (second, third, etc.) time?

Condition_____

YEARBED
HH-16. The first (second, third, etc.) time you were confined to a bed for one month or more, in what year did this begin?

__ __ __ __
Year

DURRBED
HH-18. How long were you confined to bed the first (second, third, etc.) time?

____years _____month ____ weeks ____days

WRKBED
HH-19. The first (second, third, etc.) time you were confined to a bed due to a health condition, did you miss additional school or-work or did you

become disabled due to this health condition, not including the time you spent confined to a bed?

Yes 1
No 5

Flow Check HH-19: If no go to HH-21

DURWKBED

HH-20. How long were you away from school or unable to work after the first (second, third, etc.) bed confinement, not including the time you spent confined to a bed?

____years _____month ___ weeks ____days

Flow Check HH-11: If HH-13>1, cycle through HH-14 - HH-20 as many times as needed to reach number in HH-13.
For HH-21: If HH-1=1 or HH-13=1 use intro fill

EVERWORK

HH-21. (Besides the periods of time that you have told me about), have you ever missed school or work or have you ever become disabled for one month or more due to a health condition?

Yes 1
No.....5

Flow Check HH-21: If no go to HH-28

NUMWORK

HH-22. (Besides the periods of time that you have told me about), how many different times have you missed school or work, or become disabled, for one month or more due to a health condition?

_____times

WHYWORK

HH-23. The first (second, third, etc.) time that you missed school or became unable to work for one month or more due to a health condition, was it due to an injury, an illness, a pregnancy-related complication, or something else?

Injury.....1
Illness 2
Pregnancy-related complication..3

Other (specify)_____3

CONDWORK
HH-24. What was the specific condition that caused you to miss school or become unable to work for one month or more the first (second, third, etc.) time?

Condition_____

YEARWORK
HH-25. The first (second, third, etc.) time you missed school or became unable to work for one month or more, in what year did this begin?

__ __ __ __
Year

DURRWORK
HH-27. How long were you out of school or unable to work the first (second, third, etc.) time?

____years _____month ___ weeks ____days

GENHLTH1
HH-28. Now I'd like to ask about your general health. Thinking back to your early childhood, from birth until you reached the age of 7, would you say that your health was excellent, very good, good, fair, or poor?

excellent 1
very good 2
good.....3
fair 4
poor 5

HLTHCHG1
HH-29. You've indicated that during early childhood, your health was (fill HH-28). Since early childhood, has your health consistently stayed at this level?

Yes 1
No.....5

Flow Check HH-29: If yes go to HH-33.

HTHCHGYR
HH-30. In what year did your health change?

__ __ __ __
Year

GENHLTH2

HH-31. When your health changed in (fill HH-30), would you say it became excellent, very good, good, fair, or poor (eliminate the response option given in HH-28)?

excellent 1
very good 2
good.....3
fair 4
poor……5

HLTHCHG2

HH-32. Since your health changed in (fill HH-30), has your health consistently stayed at this level?

Yes 1
No.....5

Flow Check HH-32: If no cycle through HH-30 -HH-32 until HH-32=yes

WEIGHT1

HH-33. Now I'd like to ask about your weight. Thinking back to your early childhood, from birth until you reached the age of 7, would you say that you were very overweight, slightly overweight, just right for your size, slightly underweight, or very underweight?

very overweight.....1
slightly overweight.....2
just right 3
slightly underweight.....4
very underweight.....5

WTCHG1

HH-34. You've indicated that during early childhood, your weight was (fill HH-33). Since early childhood, has your weight consistently stayed as being (fill HH-33)?

Yes 1
No.....5

Flow Check HH-34: If yes go to HH-38.

WTCHGYR

HH-35. In what year did your weight change from being (fill HH-33)?

__ __ __ __
Year

WEIGHT2

HH-36. When your weight changed in (fill HH-35), would you say you became very overweight, slightly overweight, just right for your size, slightly underweight, or very underweight (eliminate the response option given in HH-33)?

very overweight.....1
slightly overweight.....2
just right 3
slightly underweight.....4
very underweight.....5

WTCHG2

HH-37. Since your weight changed in (fill HH-35), has your weight consistently stayed as being (fill HH-33)?

Yes 1
No.....5

Flow Check HH-37: If no cycle through HH-35 - HH-37 until HH-37=yes

CURWT

HH-37A What is your current weight?

_____pounds

CURHTFT

HH-37B1 What is your current height?

_____feet

CURHTIN

HH-37B2 _____inches (ROUND TO NEAREST INCH)

EVERSMK

HH-38. Have you ever smoked cigarettes?

Yes 1
No.....5

Flow Check HH-38: If no go to END

STRTSMK
HH-39. In what year did you first start smoking cigarettes?

__ __ __ __
Year

NUMSMK
HH-40. At the time when you first started smoking, about how many cigarettes per day did you smoke? 10 or fewer, between 11 and 20, or more than 20?

10 or fewer 1
10-20 2
More than 20 ..3

SMKCHG1
HH-41. From when you started smoking until now, have you consistently smoked about (fill HH-40) cigarettes per day?

Yes 1
No.....5

Flow Check HH-41: If yes go to END

SMKCHGYR
HH-42. In what year did your number of cigarettes per day change?

__ __ __ __
Year

NUMSMK2
HH-43. In (fill HH-42), when your number of cigarettes per day changed, about how many cigarettes per day did you smoke? Did you quit, smoke 10 or fewer, smoke between 11 and 20, or smoke more than 20 cigarettes (eliminate response option given in HH-40)?

Quit……0
10 or fewer 1
10-20 2
More than 20 ..3

Flow Check HH-43: If 1, 2, or 3 go to HH-44

HH-43A Did you ever begin to smoke cigarettes again after you quit?

Yes……1
No…….5

Flow Check HH-43A: If yes cycle through HH-42 - HH-43A until HH-43A=no
If no go to END

SMKCHG2
HH-44. Since your cigarettes per day changed in (year from 42), until now, have you consistently smoked about (fill from 43) per day?

Yes 1
No.....5

Flow Check HH-44: If no cycle through HH-42 - HH-44 until HH-44=yes

END